LUGOVA Alexandra

Application des algorithmes de l'échantillonnage de Thompson et Epsilon-Greedy pour la comparaison de l'efficacité de la distribution des messages publics à grande échelle à l'aide des médias et des réseaux sociaux selon le format, le contenu et le canal de distribution

Mémoire de Master 1
Mathématiques et informatique appliquées aux sciences humaines et sociales
Parcours "Business et Data Analyst"
Année universitaire 2021-2022

Tuteur universitaire : Michal Urdanivia
Tuteur d'entreprise : Stéphanie Gauttier

Faculté d'Economie de Grenoble
1241, rue de résidences – Université Grenoble Alpes CS 40700 38058
GRENOBLE CEDEX9

SOMMAIRE

INTRODUCTION

La communication des messages d'intérêt public est un enjeu central pour les décideurs publics en ce qui concerne la sensibilisation des citoyens, l'influence sur leur attitudes, opinions et comportements face aux problèmes actuels. L'interaction efficace entre les décideurs publics et les citoyens est aussi crucial en matière de l'adoption et l'acceptation des mesures prises par le gouvernement. En particulier, la nécessité de la mise en œuvre d'un système efficace de communication avec le grand public a été mise au-devant de la scène par la crise actuelle du COVID-19 car l'acceptation de la vaccination, le respect de « gestes-barrières », et l'approbation des mesures sanitaires prises par le gouvernement sont essentiel pour faire face à la pandémie au niveau aussi bien national qu'international.

Aujourd'hui, l'emploi massif du numérique au quotidien par les citoyens (ex., internet au sens large, réseaux sociaux, messageries mobiles, etc.) fait que les outils numériques sont des vecteurs particulièrement intéressants pour les décideurs publics. Ainsi, les médias et les réseaux sociaux sont devenus des intermédiaires majeurs pour la diffusion des messages public entre les citoyens et leurs représentants.

Cependant, la réalisation des campagnes publicitaires à grande échelle sur les médias et les réseaux sociaux est très coûteux et demandent, par conséquent, le choix des paramètres des messages bien réfléchi et justifié pour pouvoir assurer l'efficacité. En particulier, la détermination du format, du contenu et des canaux les plus efficaces pour la diffusion des messages est une des questions primordiales lors de la réalisation de telles campagnes.

Un des méthodes de la comparaison des différents formes, contenus et canaux de distribution des messages public les plus populaires est l'application des A/B tests. Quand même, efficace pour les petites expériences, cette méthode devient peu performante, à la fois en termes de perte de temps et de perte de finances, en cas des campagnes à grande échelle. Il serait donc judicieux d'investir dans le développement des méthodes de la sélection des paramètres les plus optimaux pour la diffusion des messages publics dans un contexte spécifique à la fois fiables et économes en ressources financières et en temps.

Dans cet étude le problème de choix des paramètres optimaux des messages publics est considéré comme un problème du bandit manchot. L'objectif de ce papier est de considérer l'application des algorithmes de l'échantillonnage de Thompson et Epsilon-Greedy pour la comparaison plus efficace des résultats de la distribution des messages publics à grande échelle à l'aide des médias et des réseaux sociaux selon le format, le contenu et le canal de distribution et les comparer avec l'application des A/B tests.

REVUE DE LA LITTERATURE

Le problème du bandit manchot fait l'objet de décennies d'études intenses dans les domaines des statistiques, de la recherche opérationnelle, de l'électrotechnique, de l'informatique et de l'économie. Il existe de nombreux papiers contemporains visant l'évaluation de l'efficacité de l'utilisation des algorithmes bayésiens pour la solution de ce problème. Par instance, l'étude de Daniel Russo (2016) [10, p.17] compare l'efficacité des algorithmes bayésiens de l'échantillonnage probabiliste, y compris l'échantillonnage de Thompson, pour allouer de manière adaptative l'effort de mesure. Carlos Alberto Gomez-Uribe (2016) [9, p.17] étudie l'utilisation d'algorithmes exponentiels multivariés en combinaison avec l'échantillonnage de Thompson en termes de scénario de bandit multi-armé. Agrawal, S., V. Avadhanula, V. Goyal, and A. Zeevi (2017) [7, p.17] proposent également l'algorithme de l'échantillonnage de Thompson comme une solution efficiente du problème du bandit manchot. Plusieurs auteurs, par exemple, Hariharan, N., Paavai, A. G. (2022) [4, p.17], se concentrent également sur les possibilités de l'apprentissage par renforcement profond avec l'algorithme Epsilon-Greedy.

Aujourd'hui, l'une des applications les plus populaires des algorithmes bayésiens est le marketing en ligne. Il existe de nombreuses propositions de l'application des solutions bayésiens du problème du bandit manchot pour l'optimisation des campagnes publicitaires, des stratégies de prix, des algorithmes de recommandations personnalisées etc. Ainsi, D. N. Hill, H. Nassif, Y. Liu, A. Iyer, et S. V. N. Vishwanathan (2017) [8, p.17] ont introduit une approche des bandits multivariés pour l'optimisation de la mise en page des web sites. Agarwal, D., B. Long, J. Traupman, D. Xin, and L. Zhang (2014) [6, p.17] proposent une solution holistique d'apprentissage automatique de bout en bout, incluant l'algorithme de l'échantillonnage de Thompson, pour déployer des modèles de prédiction de réponse basés sur la régression logistique vers un grand système de publicité en ligne. Agarwal, D (2013) [5, p.17] démontre également les possibilités de l'utilisation des algorithmes de la publicité informatique sur un exemple de Linkedin. L'étude expérimentale de Björn Brodén, Mikael Hammar, Bengt J. Nilsson, and Dimitris Paraschakis (2018) [2, p.17] présente une extension de la solution du problème du bandit manchot avec l'algorithme de l'échantillonnage de Thompson pour orchestrer la collection d'algorithmes de recommandation de base pour le commerce électronique. Enfin, Toshihiro Kamishima, Shotaro Akaho (2011) [3, p.17] proposent un système de recommandation de prix personnalisé se basant sur l'algorithme multi-étape Epsilon-Greedy.

Les algorithmes de l'échantillonnage de Thompson et Epsilon-Greedy constituant les deux approches le plus populaires pour remplacer l'approche classique des A/B tests dans le marketing en ligne, il serait utile de comparer leur efficacité à résoudre les tâches du type du bandit manchot répandus dans ce domaine d'activité. Pour ce but, Izzatul Umami and Lailia Rahmawati (2021) [1, p.17] ont réalisé une étude comparant des modèles de l'échantillonnage de Thompson et Epsilon-Greedy, ainsi que l'UCB-1, pour le problème du bandit manchot dans le secteur marketing. En résultat, ils sont arrivés à la conclusion que les algorithmes de l'échantillonnage de Thompson et l'UCB-1 a montré des résultats plus efficaces, sans perdre l'expérimentation et les variations statistiques pour maximiser les paiements totaux ce qui est généralement conforme aux conclusions des études pareilles antérieures.

Ce papier a pour le but de comparer l'efficacité de l'application de ces deux algorithmes, ainsi que des A/B tests pour la sélection du format, du contenu et du canal de distribution optimaux pour la distribution des messages publics à grande échelle ce qui est une nouvelle application du problème du bandit manchot, qui n'a pas encore de couverture solide dans la littérature.

PROBLEME DU BANDIT MANCHOT ET SOLUTIONS

Le problème du bandit manchot est l'un des problèmes les plus fondamentaux de la science de la décision et, en particulier, la théorie des probabilités. C'est le problème de l'allocation optimale des ressources dans le cas où il y a l'incertitude sur le rendement de divers investissements. Le nom même de « bandit manchot » vient des anciennes machines à sous, qui étaient contrôlées avec des stylos. Ces machines étaient surnommées « bandits » parce qu'après avoir interagi avec elles, la plupart des gens perdait beaucoup d'argent et se sentaient volés. Supposons qu'il y a plusieurs machines de ce type et que chacune d'entre elles a une probabilité différente de gagner. Le problème qui se pose pour maximiser les gains est celui de la détermination de la machine avec la plus grande probabilité de gagner au moindre coût possible. Typiquement, la politique de l'utilisateur oscille entre exploitation (utiliser la machine dont il a appris qu'elle récompense beaucoup) et exploration (tester une autre machine pour espérer gagner plus). Ce dilemme, appelé le dilemme « exploration-exploitation », consiste à choisir l'action qui maximise la récompense attendue par rapport à une croyance tirée au hasard.

C'est un problème auquel les entreprises et les organisations publics sont confrontées très souvent dans leurs pratiques, en particulier, par exemple, dans le domaine de marketing. Imaginons qu'une organisation dispose de plusieurs options pour les messages d'intérêt public, par exemple, des messages incitant au port du masque dans les lieux publics, qu'elle planifie à

diffuser à un grand public à l'aide des réseaux sociaux. Dans ce cas, il est important pour l'organisation de choisir pour de telles messages le format (vidéo, photo, texte, etc.), le canal (Facebook, Instagram, Twitter, etc.) et le contenu les plus optimaux qui vont au mieux attirer l'attention des gens et maximiser la conversion. La conversion est calculée comme un pourcentage du nombre total de visiteurs ayant effectué une certaine action. Pour notre cas, la conversion peut consister à « liker » la publication, à s'inscrire à la page de l'organisation, à s'abonner aux actualités, à cliquer sur un lien et bien d'autres actions.

La façon typique et la plus répondue de résoudre ce problème consiste à exécuter un A/B test. Généralement, un A/B test est effectué comme suit : on prend toutes les variantes de messages et on teste chacun pendant la même durée. Par exemple, il y a 2 types de messages à comparer (supposons, un message en forme d'une photo et l'autre en forme d'un vidéo). Ils sont diffusés dans des conditions identiques sur le même canal choisi et pendant la même période. À la fin de l'expérience, il est possible de calculer et comparer directement leurs conversions, pat exemple, en divisant le nombre de « likes » par le nombre d'impressions, et choisir le meilleur format.

Cette méthode peut être vraiment efficace lorsqu'il y a peu d'options à tester. Quand même, lorsqu'il y a beaucoup de décisions à prendre concernant à la fois le choix du format, du contenu et du canal, cette approche devient inefficace en termes de perte de temps et de perte de ressources financières. Cela arrive parce que pour atteindre la significativité statistique des résultats obtenus il faut assurer un A/B test de duration assez important. En même temps, l'expérimentateur ne sait pas laquelle option est la meilleure tant que le test n'est pas terminé et il perd, par conséquent, beaucoup de ressources sur des options de message inefficaces lors du déroulement du test. Il serait donc très profitable de pouvoir rapidement éliminer les mauvaises options, et investir les ressources dans celle qui est la meilleure.

Une des solutions possibles pour résoudre ce problème est d'utiliser l'algorithme Epsilon-Greedy. « Greedy » signifie que cet algorithme est « gourmand » dans la mesure où il considère toujours le résultat immédiat sans tenir compte des résultats à long terme. L'application de cet algorithme commence par plusieurs impressions de chaque option de message. Ainsi, il devient possible de calculer des estimations initiales des conversions. Après cela, une option avec la meilleure conversion est choisie pour être la première à tester. Elle est montrée à un utilisateur et sa conversion est recalculée selon le résultat de cette impression (conversion ou pas de conversion). Les conversions de toutes les options sont ensuite comparées à nouveau et l'option avec la meilleure conversion est choisi pour être la suivante à tester. Cette

procédure est répétée jusqu'à ce qu'une option devienne significativement plus efficace que les autres (sa conversion étant la plus importante) et l'algorithme ne choisit plus pour un test que cette option.

Plus formellement, à chaque instant t, l'algorithme Epsilon-Greedy ajuste un modèle à des paires de données historiques $\mathbb{H}_{t-1} = ((x_1, y_1), \dots, (x_{t-1}, y_{t-1}))$, générant une estimation $\hat{\theta}$ des paramètres du modèle. Le modèle résultant peut ensuite être utilisé pour prédire la récompense $r_t = r(y_t)$ de l'application de l'action x_t. Ici, y_t est un résultat observé, tandis que r est une fonction connue qui représente les préférences de l'agent. Étant donné les paramètres de modèle estimés $\hat{\theta}$, un algorithme d'optimisation sélectionne l'action x_t qui maximise la récompense attendue, en supposant que $\theta = \hat{\theta}$. Cette action est ensuite appliquée au système exogène et un résultat y_t est observé.

Il s'agit d'un algorithme assez efficace qui assure généralement une meilleure performance que l'application des A/B tests. Cependant, l'algorithme Epsilon-Greedy a un point faible très important : il existe un risque qu'il reste bloqué sur une option sous-optimale et sauter une autre option plus efficace sans lui donner une chance de prouver l'efficacité. Par conséquent, même s'il permet de résoudre le problème d'exploration du dilemme « exploration-exploitation », la qualité de l'exploration de l'espace des options reste limitée.

Pour faire face à ce défaut, cet algorithme dispose d'un argument « epsilon » qui définit la probabilité avec laquelle l'algorithme va choisir une option à tester au hasard et pas en appuyant sur la comparaison des conversions. En variant cette probabilité, il est possible, dans la plupart des cas, d'atteindre un bon niveau d'exploration de l'espace des options et assurer que toutes les options ont leur chance de prouver l'efficacité et être choisi comme au final. L'introduction de cet argument rend l'algorithme Epsilon-Greedy très efficace et fiable pour résoudre le problème du choix des paramètres des messages publics en minimisant les coûts de temps et de finances.

Un autre algorithme qui peut rivaliser avec l'Epsilon-Greedy en efficacité est l'algorithme heuristique de l'échantillonnage de Thompson. Cet algorithme permet d'explorer vraiment très bien toutes les options de messages. Au lieu de simplement calculer et comparer à chaque étape les conversions des différentes options de messages, il construit un modèle probabiliste avec des distributions bêta de la probabilité de gagner (avoir la conversion) pour chaque message.

La distribution bêta prend deux paramètres, « α » et « β » qui peuvent être considérés, dans notre cas, comme le nombre de réussites (conversion) et d'échecs (pas de conversion) respectivement. De plus, une distribution bêta a une valeur moyenne donnée par :

$$\frac{a}{a+\beta} = \frac{nombre\ de\ réussites}{nombre\ total\ d'éssais}$$

Initialement, l'algorithme définit les paramètres « α » et « β » égaux à 1 pour chaque option, ce qui produit une ligne plate de la distribution uniforme. Cette estimation initiale de la probabilité qu'un message produise une conversion est appelée « la probabilité a priori ». Il s'agit de la probabilité que l'événement spécifique se produise avant qu'il y ait des preuves réelles.

Une option est choisie aléatoirement pour être la première à tester. Elle est ensuite montrée à un utilisateur et sa probabilité de gagner est recalculée selon le résultat de cette impression en mettant à jour les paramètres « α » et « β » ce qui modifie aussi sa fonction de densité de distribution. Cette nouvelle probabilité, après que certaines preuves ont été recueillies, est connue sous le nom de « la probabilité postérieure ». Puis, un point est choisi au hasard pour chaque distribution bêta et le type de message dont la point représente la meilleure conversion devient le suivant à tester. Au fur et à mesure que davantage de données sont collectées, la distribution bêta passe d'une ligne plate à un modèle de plus en plus précis de la probabilité de gains (figure 1). L'algorithme donc fonctionne jusqu'à ce qu'une option devienne significativement plus efficace que les autres (sa probabilité de gagner étant la plus importante) et l'algorithme ne choisit plus pour un test que cette option.

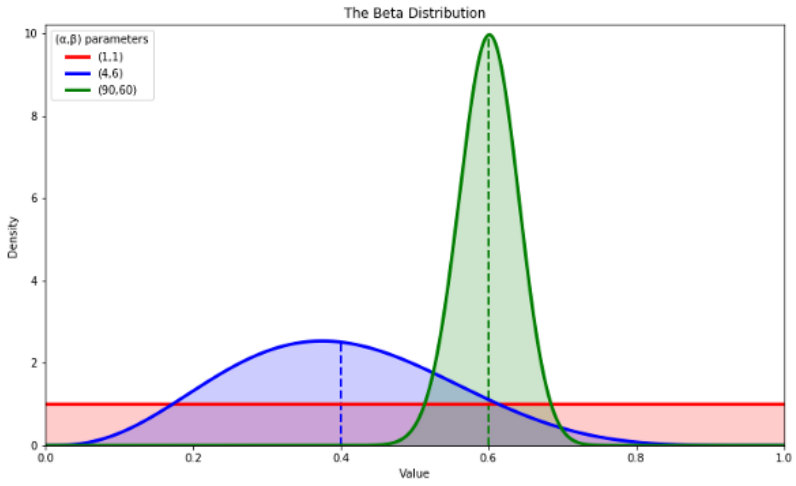

Figure 1 - Comparaison de la distribution bêta pour différentes valeurs d'alpha et de bêta.

Contrairement à l'algorithme Epsilon-Greedy, qui à chaque itération sélectionne le message avec l'estimation de la conversion la plus élevée, même si la confiance dans cette estimation est faible, l'échantillonnage de Thompson échantillonne à partir de la distribution bêta des probabilités d'avoir la conversion de chaque option et choisit celle avec la probabilité la plus élevée. Étant donné que les options qui ont été testées rarement ont de larges distributions (la courbe bleue sur le graphique ci-dessus), elles ont une plus grande plage de valeurs possibles. De cette manière, un message qui a actuellement une probabilité à gagner estimée faible, mais qui a été testée moins de fois qu'une bannière avec une probabilité plus élevée, peut renvoyer une valeur d'échantillon plus grande et a une chance d'être sélectionnée à cette itération.

Par exemple, dans le graphique ci-dessus, la courbe bleue a une probabilité à gagner estimée inférieure à la courbe verte. Par conséquent, en cas d'utilisation de l'algorithme Epsilon Greedy, le message correspondant à la courbe verte serait choisi et la bleue ne serait jamais sélectionnée. En revanche, l'échantillonnage de Thompson considère effectivement toute la largeur de la courbe, qui, pour le message bleu, peut être considérée comme s'étendant au-delà de celle de la verte. Dans ce cas, le bleu peut être sélectionné de préférence au vert. Cela signifie que l'algorithme de l'échantillonnage de Thompson permet l'exploration de l'espace des options de plus haute qualité que l'Epsilon Greedy.

Quand même, au fur et à mesure que le nombre d'essais d'un message augmente, la confiance dans sa probabilité à gagner augmente, la distribution de probabilité devient plus étroite et la valeur échantillonnée est alors tirée d'une plage de valeurs plus proches de la vraie conversion (la courbe verte sur le graphique ci-dessus). En conséquence, l'exploration diminue et l'exploitation augmente.

D'autre part, les messages avec la probabilité estimée faible commencent à être sélectionnées moins fréquemment et ont tendance à être abandonnées plus tôt dans le processus de sélection. Étant donné que nous sommes uniquement intéressés à trouver la meilleure option le plus rapidement possible, il n'y a pas d'intérêt dans les bannières peu performantes, donc ce n'est pas vraiment un problème.

Maintenant on va approcher la description du fonctionnement de l'algorithme de l'échantillonnage de Thompson et sa comparaison avec l'algorithme Epsilon-Greedy plus formellement. Supposons que l'agent dispose d'une séquence d'actions $x_1, x_2, \dots, x_n$ pour un système, en sélectionnant chacun de l'ensemble X. Cet ensemble d'actions peut être fini ou infini. Après avoir appliqué l'action x_t, l'agent observe un résultat y_t, que le système génère aléatoirement selon une mesure de probabilité conditionnelle $q_\theta(\cdot \mid \mathrm{x_t})$. L'agent bénéficie d'une

récompense $r_t = r(y_t)$, où r est une fonction connue. L'agent est initialement incertain de la valeur de θ et représente son incertitude à l'aide d'une distribution a priori p.

Les deux algorithmes ont une manière différente pour générer les paramètres du modèle $\hat{\theta}$. L'algorithme Epsilon-Greedy prend $\hat{\theta}$ comme l'espérance de θ par rapport à la distribution p, tandis que l'échantillonnage de Thompson tire un échantillon aléatoire de p. Les deux algorithmes appliquent ensuite des actions qui maximisent la récompense attendue pour leurs modèles respectifs. S'il existe un ensemble fini d'observations possibles y_t, cette espérance est donnée par :

$$\mathbb{E}_{q\hat{\theta}}[r(y_t)|x_t = x] = \sum_o q_{\hat{\theta}}(o|x)r(o).$$

La distribution p est mise à jour en conditionnant sur l'observation réalisée $\widehat{y_t}$. Si θ est limité aux valeurs d'un ensemble fini, cette distribution conditionnelle peut être écrite par la règle de Bayes comme :

$$\mathbb{P}_{p,q}(\theta = u|\mathrm{x_t}, y_t) = \frac{p(u)q_u(y_t|\mathrm{x_t})}{\sum_v p(v)q_v(y_t|\mathrm{x_t})}.$$

Le problème du bandit manchot avec une distribution a priori bêta est un cas particulier de cette formulation plus générale. Dans ce cas particulier, l'ensemble des actions est $X = \{1, \dots, K\}$ et seules les récompenses sont observées, donc $y_t = r_t$. Les observations et les récompenses sont modélisées par les probabilités conditionnelles $q_\theta(1|\mathrm{k}) = \theta_k$ et $q_\theta(0|\mathrm{k}) = 1$. La distribution a priori est codée par les vecteurs α et β, avec une fonction de densité de probabilité donnée par :

$$p(\theta) = \prod_{k=1}^{K} \frac{\Gamma(a+\beta)}{\Gamma(a_k)\Gamma(\beta_k)} \theta_k^{a_k - 1}(1-\theta_k)^{\beta_k - 1},$$

où Γ désigne la fonction gamma. En d'autres termes, sous la distribution a priori, les composantes de θ sont indépendantes et bêta-distribuées, avec les paramètres α et β.

Pour ce problème, les deux algorithmes commencent chaque t-ième itération avec des paramètres postérieurs a_k, β_k pour $\mathrm{k} \in \{1, \dots, \mathrm{K}\}$. L'algorithme Epsilon-Greedy fixe $\widehat{\theta_k}$ à la valeur attendue $E_p[\theta_k] = \frac{a_k}{a_k + \beta_k}$, alors que l'échantillonnage de Thompson tire aléatoirement $\widehat{\theta_k}$ à partir d'une distribution bêta de paramètres a_k, β_k. Chaque algorithme sélectionne alors l'action x_t qui maximise $E_{q_{\hat{\theta}}}[r(y_t)|\mathrm{x_t} = x] = \widehat{\theta_x}$. Après avoir appliqué l'action sélectionnée, une récompense $r_t = y_t$ est observée et les paramètres de distribution des croyances sont mis à jour en fonction de :

$$(a, \beta) \leftarrow \left(a + r_t 1_{x_t}, \beta + (1 - r_t) 1_{x_t}\right),$$

où 1_{x_t} est un vecteur dont la composante x_t est égale à 1 et toutes les autres composantes sont égales à 0.

Pour illustrer maintenant l'application de cet algorithme sur un exemple concret, supposons qu'il y a 3 formats de message différents à tester – texte, image, vidéo – pour un message d'intérêt public, disons encourageant le port du masque dans les transports en commun. Supposons aussi que les probabilités réelles d'obtenir une conversion sont déjà connues pour chaque type de message (cela est nécessaire pour pouvoir mesurer l'efficacité réelle des algorithmes) : un message en forme d'image assure la probabilité de conversion la plus importante de 0.8 (figure 2, distribution bleue), un message vidéo est presque aussi efficace avec la probabilité de 0.7 (figure 2, distribution rouge) et la probabilité de conversion assurée par un message textuel n'est que 0.3 (figure 2, distribution verte).

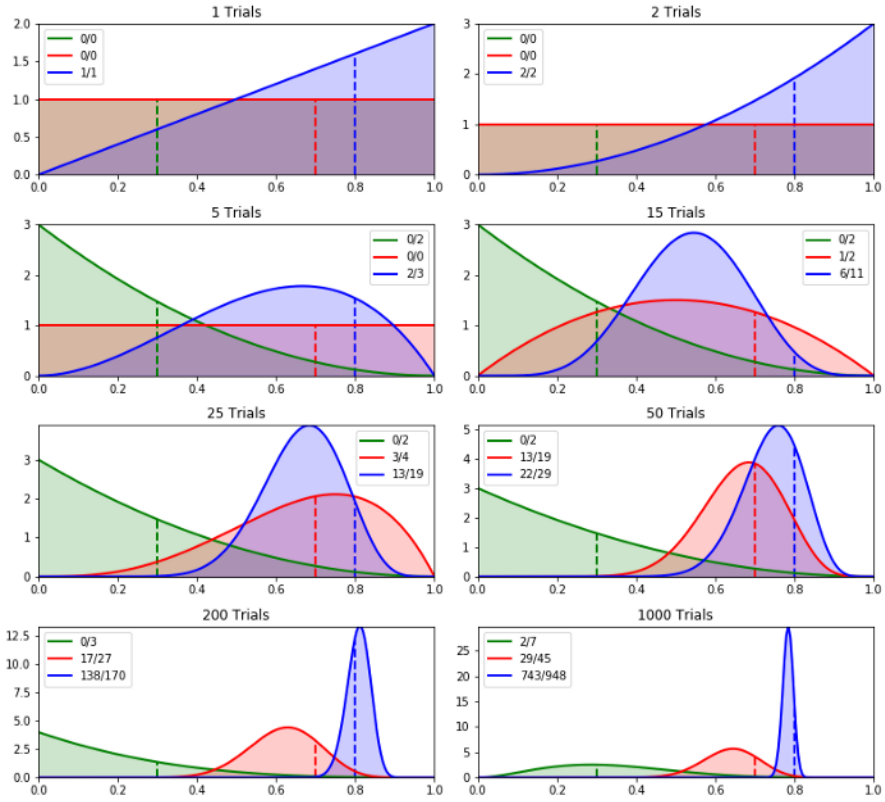

Figure 2 - Le processus de l'échantillonnage de Thompson pour comparaison des trois canaux pour la diffusion d'un message d'intérêt public

Le processus de l'échantillonnage de Thompson pour comparaison des trois types de messages d'intérêt public, avec des probabilités réelles de 0,3, 0,7 et 0,8 est illustré dans la figure ci-dessus. Les vraies moyennes des distributions (probabilités d'avoir la conversion) sont représentées par les lignes pointillées. La légende affiche le nombre d'essais pour chaque

type de message et le nombre d'impressions ayant entrainé une conversion qui ont résulté de ces essais.

Avant le commencement des essais, toutes les distributions bêta constituent une distribution uniforme plate. Étant donné que tous les messages ont une distribution égale, le message en forme d'image est sélectionné au hasard pour être le premier à tester. L'impression du message à un utilisateur résulte dans une conversion, donc la valeur α de la distribution bleue est incrémentée de 1 et sa courbe de densité de probabilité se déplace vers la droite. Au deuxième essai, l'image est à nouveau sélectionnée car la valeur choisie au hasard de sa distribution est plus importante que celles des autres et elle renvoie encore une fois une conversion. La courbe bleue se resserre donc légèrement plus vers la droite.

Au cinquième essai, l'image a été sélectionnée une fois de plus, mais cette fois elle n'a pas donné de conversion. En conséquence, sa probabilité de gagner diminue. D'autre part, le message textuel a maintenant été testé deux fois et n'a pas encore renvoyé de conversion, par conséquent sa courbe de densité de probabilité est décalée vers la gauche avec la valeur de la probabilité égale à 0. A ce moment, on peut dire qu'il y a une chance que celle-ci ne renvoie jamais une conversion.

À 15 essais, le message vidéo a été essayée deux fois. Puisqu'il a renvoyé une conversion une fois, il a une probabilité de conversion moyenne estimée à 0,5. À ce stade, le message en forme d'image a été essayée déjà 11 fois et a renvoyé une conversion 6 fois, ce qui lui donne une probabilité de conversion estimée légèrement plus élevée de 0,54. Dans le cas de l'utilisation de l'algorithme Epsilon-Greedy, l'image serait choisie pour un essai suivant, mais comme le message vidéo a été essayée moins de fois que le bleu, on peut voir qu'il a une courbe de densité de probabilité beaucoup plus large, ce qui lui donne de bonnes chances d'être sélectionnée de préférence à l'image dans le cas d'application de l'échantillonnage de Thompson.

Plus un message est testé, plus on est confiant dans l'estimation de son efficacité et plus sa courbe de densité de probabilité devient étroite. La meilleure option est alors testée plus souvent et les tests des options sous-optimaux s'épuisent. Ce comportement peut être constaté à la fin de notre test, lorsque le message en forme d'image est essayé beaucoup plus souvent que les autres et, les distributions ne changeant plus significativement quel que soit le nombre d'essais additionnels, il devient évident que c'est l'option la plus efficace.

On peut constater que le choix de l'algorithme correspond à la réalité, le message en format d'image ayant la probabilité réelle d'avoir la conversion la plus importante. Ainsi, l'algorithme de l'échantillonnage de Thompson se prouve d'être efficace pour définir le format

de message optimal pour assurer la conversion la plus important. Cet algorithme peut être appliqué de même façon pour définir le canal de diffusion des messages le plus optimal ou, par exemple, comparer les messages de différent contenu (formulation des phrases, personnalité dans le cadre, spectre de couleurs, etc.).

APPLICATION DES ALGORITHMES SUR LES DONNEES

Supposons maintenant qu'il y a 5 messages du contenu différent à tester et choisir le meilleur pour la diffusion à un grand public. En utilisant Python, on peut simuler l'application des trois algorithmes discutés ci-dessus pour la sélection du message optimal avec la conversion la plus élevée. Pour pouvoir comparer leurs performances, on va attribuer au hasard une probabilité d'obtenir une conversion réelle aléatoire à chaque type de message (tableau 1).

Message	0	1	2	3	4
Probabilité de conversion	0.180995	0.351419	0.304502	0.189610	0.495015

Tableau 1 - Probabilités de conversion définies au hasard pour chaque type de message

Selon les probabilités générées aléatoirement, le message le plus optimal est 4. En appliquant les algorithmes à l'étude pour choisir le meilleur message et simulant le processus de tests (impression des messages aux utilisateurs et modification graduelle des estimations des conversions), il est possible à prouver que tous les trois algorithmes sont efficaces et parviennent à choisir le message 4 comme le meilleur.

En réalisant l'algorithme de l'échantillonnage de Thompson avec le nombre de tests égal à 300, on obtient les résultats suivants :

Message	**Nombre de gains**	**Nombre total de tests**	**Conversion estimée**	**Probabilité de conversion réelle**
0	3	18	0.167	0.180995
1	13	43	0.302	0.351419
2	8	31	0.258	0.304502
3	7	28	0.250	0.189610
4	85	181	0.470	0.495015

Tableau 2 - Résultats de l'application de l'échantillonnage de Thompson (300 itérations)

On peut constater que l'algorithme a bien défini le message avec la probabilité de conversion la plus élevée et son estimation de conversion est assez proche à la valeur réelle.

Comme on peut voir sur le graphique (figure 3) contenant des distributions de probabilités de conversion de chaque message, la distribution de message 4 se situe plus à droite et est beaucoup plus étroite que les autres ce qui signifie que l'estimation est déjà assez précise à cette étape et il n'y a pas besoin de continuer l'expérience car il est très peu probable qu'un autre message dépasse le message 4.

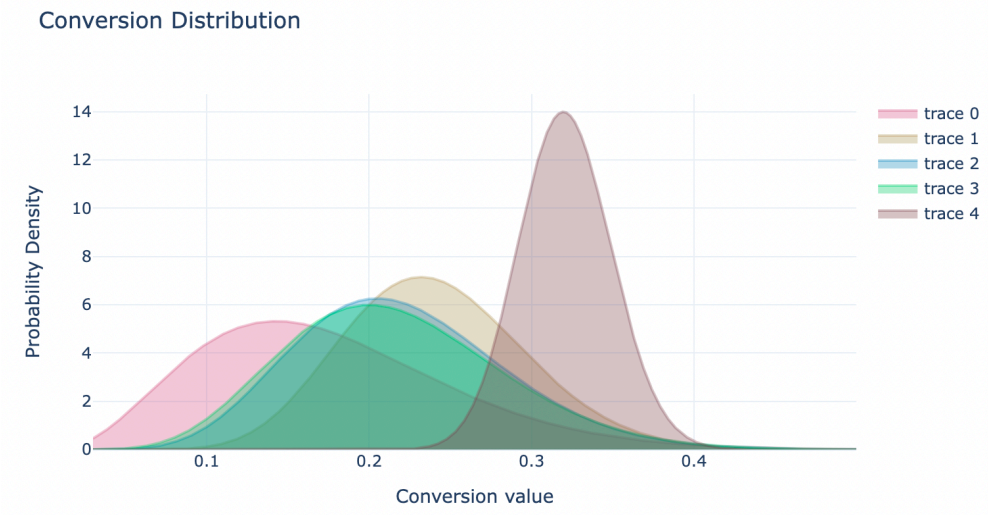

Figure 3 - Distributions de probabilités de conversions après 300-ème test

L'algorithme Epsilon-Greedy avec l'argument epsilon définissant le pourcentage de choix aléatoire d'un meilleur message après un test égal à 0.15 permet aussi atteindre le résultat désiré. Au moment de 300 tests, les résultats de l'application de cet algorithme sont les suivants :

Message	**Nombre de gains**	**Nombre total de tests**	**Conversion estimée**	**Probabilité de conversion réelle**
0	3	11	0.270	0.180995
1	5	15	0.330	0.351419
2	3	8	0.375	0.304502
3	3	15	0.200	0.189610
4	130	251	0.518	0.495015

Tableau 3 - Résultats de l'application de l'algorithme l'Epsilon-Greedy (300 itérations)

Pourtant, on peut constater que la qualité de l'exploration de l'espace des options est un peu pire pour l'Epsilon-Greedy car l'algorithme s'est fixé plus tôt sur le message 4 et a donné moins de chances aux autres. De plus, si on répète cet algorithme plusieurs fois, on constate qu'il n'est pas si stable que l'échantillonnage de Thompson et nécessite parfois un nombre de tests plus élevé pour assurer un bon résultat. Il est mieux donc de réaliser un nombre de tests un peu plus élevé dans le cas de l'application de l'algorithme Epsilon-Greedy.

Pour un A/B test classique avec la réalisation de 300 tests peut être pas suffisant parce que, dans ce cas, on va simplement tester chaque message 60 fois et calculer ensuite leurs conversions ce qui peut donner des résultats pas significatifs. Comme il est possible de constater en appuyant sur le tableau des résultats d'un A/B test de 300 tests, le message 4 démontre tout à fait la meilleure conversion (tableau 4). Quand même, il a un nombre de gains très proche à celui du message 1 ce qui ne permet pas de garantir que le dernier ne va pas dépasser le leadeur

actuel avec un nombre de tests plus important. Il est mieux donc de réaliser un nombre de tests beaucoup plus élevé dans le cas de l'application d'un A/B test.

Message	Nombre de gains	Nombre total de tests	Conversion estimée	Probabilité de conversion réelle
0	12	60	0.200	0.180995
1	23	60	0.383	0.351419
2	13	60	0.216	0.304502
3	11	60	0.183	0.189610
4	27	60	0.450	0.495015

Tableau 4 - Résultats de l'application d'un A/B test (300 tests)

Quand même, lorsque en réalisant les algorithmes de l'échantillonnage de Thompson et l'Epsilon-Greedy la grande majorité des tests est fait sur le message optimal, le format de l'A/B test oblige à repartir les tests équitablement entre les messages ce qui signifie qu'une grande partie des ressources est dépassée sur les messages sous-optimaux.

En total, il est possible de dire que tous les trois algorithmes testés ci-dessus sont plus ou moins efficace pour la comparaison des messages de différents types et la sélection de la meilleure en termes de conversion. Pourtant, dans le cas des campagnes de diffusion des messages public à grande échelle, on a parfois beaucoup plus de paramètres de message et canaux de diffusion à tester et comparer. Supposons, il y a 10/30/50 d'options différents. Dans ce cas, il est raisonnable de choisir un nombre de tests beaucoup plus important (par exemple, 10000/35000/50000 respectivement). Les A/B tests devient donc très couteaux car on est obligé à répartir équitablement le nombre de tests sur tous les messages et, par conséquent, même si on parvient à choisir un message optimal, on va dépasser 4/5 de nos ressources sur la diffusion des messages sous-optimaux. Dans ce cas, il est donc judicieux de choisir entre les algorithmes de l'échantillonnage de Thompson et l'Epsilon-Greedy.

Pour comparer l'efficacité des trois algorithmes mentionnés ci-dessus, on les lance en parallèle trois fois (10000 tests pour comparer 10 options, 35000 - pour 30 options et 50000 - pour 50 options ce qui constitue en total 95000 tests). Les résultats de cette expérience sont présentés dans le tableau 5 ci-dessous.

Algorithme	Nombre de gains total	Pourcentage de gains
Échantillonnage de Thompson	44376	46,71%
Epsilon-Greedy	42861	45,12%
A/B test	26623	28,02%

Tableau 5 - Comparaison des résultats des tests pour les trois algorithmes

Comme on peut voir, l'algorithme de l'échantillonnage de Thompson est la plus efficace avec 46,71% de tests ayant résulté en une conversion. L'Epsilon-Greedy a presque la même efficacité en assurant 45,12% des conversions. Au contraire, l'A/B test ne permet d'atteindre que 28,02% de tests réussis. Il est donc environ 40% moins efficace que les deux premiers. De plus, si on considère les trois essais (pour 10/30/50 options) séparément, on voit que, l'algorithme de l'échantillonnage de Thompson est le meilleur dans tous les cas.

CONCLUSION

En résume, il est possible d'optimiser le budget et les ressources temporelles dépensés par les organisations publiques dans le cadre de réalisation des campagnes de communication publique à grande échelle grâce à l'utilisation des algorithmes de l'échantillonnage de Thompson ou Epsilon-Greedy au lieu des A/B tests pour la sélection des paramètres d'une campagne optimaux.

Comme il a été prouvé par l'expérience sur les données simulées, tous les trois algorithmes sont efficaces pour choisir une option de message optimal en termes de récompense maximale. Quand même, lorsque les A/B tests deviennent très couteux avec le nombre d'options comparés élevé, les algorithmes de l'échantillonnage de Thompson et Epsilon-Greedy permettent de diminuer considérablement les coûts en assurant la distribution des ressources optimale pour la maximisation de la récompense totale.

De plus, l'application de l'algorithme Epsilon-Greedy parfois entraîne un risque de blocage sur une option de message sous-optimale. L'algorithme de l'échantillonnage de Thompson, au contraire, permet l'exploration de l'espace des options de haute qualité. En outre, cet algorithme assure une meilleure performance en termes de l'économisation des ressources financières et, par conséquent, la maximisation des gains avec le nombre d'options comparés particulièrement élevé.

Les limites de l'application des algorithmes proposées peuvent consister du mécanisme de son action et des formulation mathématiques difficiles à comprendre, ainsi que de la mise en œuvre du côté technique de l'algorithme, notamment l'assurance de son fonctionnement correct en mode en ligne et son intégration dans les systèmes informatiques existants.

LITTERATURE

1. Umami, Izzatul, Lailia Rahmawati (2021). Comparing Epsilon Greedy and Thompson Sampling model for Multi-Armed Bandit algorithm on Marketing Dataset. *Journal of Applied Data Sciences*, 2.2.
2. Björn Brodén, Mikael Hammar, Bengt J. Nilsson, and Dimitris Paraschakis (2018). Ensemble Recommendations via Thompson Sampling: an Experimental Study within e-Commerce. *Proceedings of the 23rd International Conference on Intelligent User Interfaces*. Association for Computing Machinery, New York, NY, USA, 19–29.
3. Toshihiro Kamishima, Shotaro Akaho (2011). Personalized pricing recommender system: multi-stage epsilon-greedy approach. *Proceedings of the 2nd International Workshop on Information Heterogeneity and Fusion in Recommender Systems*. Association for Computing Machinery, New York, NY, USA, 57–64.
4. Hariharan, N., Paavai, A. G. (2022). A Brief Study of Deep Reinforcement Learning with Epsilon-Greedy Exploration. *International Journal of Computing and Digital Systems*, 11(1), 541.
5. Agarwal, D. (2013). Computational advertising: the LinkedIn way. *Proceedings of the 22nd ACM International Conference on Information & Knowledge Management*. ACM. 1585–1586.
6. Agarwal, D., B. Long, J. Traupman, D. Xin, and L. Zhang (2014). Laser : a scalable response prediction platform for online advertising. *Proceedings of the 7th ACM international conference on Web search and data mining.* ACM. 173–182.
7. Agrawal, S., V. Avadhanula, V. Goyal, and A. Zeevi (2017). Thompson sampling for the MNL-bandit. *Proceedings of the 30th Annual Conference on Learning Theory*. 76–78.
8. Hill, D. N., H. Nassif, Y. Liu, A. Iyer, and S. V. N. Vishwanathan (2017). An efficient bandit algorithm for realtime multivariate optimization. *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 1813–1821.
9. Gómez-Uribe, C. A. (2016). *Online algorithms for parameter mean and variance estimation in dynamic regression*.
10. Russo, D. (2016). Simple bayesian algorithms for best arm identification. *Proceedings of the Conference on Learning Theory*. 1417–1418.