

CIRJE-J-311

長期的顧客指標のためのAdaptive Minimax Regret (AMR)
による確率型A/Bテスト

東京大学大学院経済学研究科
阿部 誠

2024年8月

CIRJE ディスカッションペーパーの多くは
以下のサイトから無料で入手可能です。
http://www.cirje.e.u-tokyo.ac.jp/research/03research02dp_j.html

このディスカッションペーパーは、内部での討論に資するための未定稿の段階にある論文草稿である。著者の承諾なしに引用・複写することは差し控えられたい。

長期的顧客指標のための Adaptive Minimax Regret (AMR) による確率型 A/B テスト

阿部誠

東京大学大学院経済学研究科

Email: abe@e.u-tokyo.ac.jp

2024 年 8 月 20 日

長期的顧客指標のための Adaptive Minimax Regret (AMR) による確率型 A/B テスト

概要

経済学では不確実性を、確率で評価できる「リスク」と確率が未知である「曖昧性」の 2 種類に区別する。リスク下の意思決定では、事象がもたらす効用をその確率で重み付けた期待効用が最大になるような選択をすることが合理的な行動である。一方、確率が未知な曖昧性下では、どのように意思決定をするべきなのだろうか。

本論文では、まず、このような曖昧性下において、確率は未知だが区間のみ分かっている状況での意思決定基準の 1 つである最大リグレット最小化 (Minimax Regret) 基準を紹介する。具体例として、既存と新規 2 台スロットマシンで、前者の当選確率は既知であるが、後者の当選確率は未知で確率の区間のみ分かっているとしよう。この場合、最大リグレット最小化基準による最適な行動は、この 2 台のスロットをそれぞれ一定の確率でランダムに引くことになる。

次に、効用が長期的な指標の場合、この指標の区間不確実性は時間が経つにつれて減少していく。そこで、その時点で得られた情報に基づいて、最大リグレット最小化基準による確率を更新しながら効用を最大化する Adaptive Minimax Regret を導入する。上述の、既存と新規 2 台のスロットマシンのケースでシミュレーション検証を行った結果、導入手法はバンディット・アルゴリズムに匹敵する高いパフォーマンスをもたらした。この AMR を使ったマーケティングへの応用として、ターゲットとなる顧客の平均生存期間 (効用) が一番長くなる逐次的なキャンペーン施策、確率型 A/B テストを提案した。

Keywords: ミニマックス・リグレット、バンディット・アルゴリズム、確率型 A/B テスト

Probability-based A/B testing with Adaptive Minimax Regret (AMR) criterion for long-term customer metrics

Evaluating long-term metrics like customer lifetime requires an extended observation period. We propose Adaptive Minimax Regret, which permits companies to make an optimal decision under minimax criteria based on the current information at hand and update it sequentially, enabling them to respond before knowing the complete results of long-term metrics.

1. はじめに

生涯価値やリピート率のような顧客の長期的指標の重要性が増す中、企業はそれらの結果が全て観測される前に、早急な意思決定を迫られている。

まずは、以下の状況を考えてみよう。

『実績が4年ある既存キャンペーン E において、獲得した顧客の平均生存期間は2.7年であった。今回、生存期間を増やすべく新規キャンペーン N を企画している。N を評価期間（例えば4年間）実施、観測して、E か N、顧客の平均生存期間が長いキャンペーンを採択するのでは、結果が出るまで4年も待たなければならない。もし事前の市場調査から N の平均生存期間の確率分布が分かっていたら（例えば確率0.4で2年、確率0.6で4年）、期待値が大きい方の N (3.2年 vs. E の2.7年) を選択するのが妥当であろう。しかし事前情報が無いなど極度の不確実性から、確率分布は未知で区間のみ分かっている場合はどうすべきであろう。』

この状況は、不確実性を確率で評価できる「リスク」と、確率が未知である「曖昧性」の2種類に区別した状況における意思決定 (Savage 1951) と関連している。「リスク」下の意思決定では、結果を確率で重み付けした期待効用が高い方を選択するというベイズ的な判断が合理的だ。一方、確率が未知で区間のみ分かっている「曖昧性」下における意思決定の基準としては、マキシミン (効用の最小値を最大化)、ミニマックス・リグレット (最大リグレット最小化)、マクシマックス (効用の最大値を最大化) などが提案されている。

本研究では、キャンペーン E と N の選択に、各時点までに得られた E と N の結果に基づいてミニマックス・リグレット基準を当てはめ、それを逐次的に更新するアプローチ、アダプティブ・ミニマックス・リグレット (AMR) を提案する (Manski 2011)。

2. 確率的意思決定：ミニマックス・リグレット

ミニマックス・リグレットは、効用の最大損失を最小化する意思決定基準である。一方、マキシミン基準はワーストシナリオの最大化を目指しているため意思決定志向が過度に保守的になる傾向がある。他方で、マクシマックス基準ではベストシナリオの最大化を目指しているため過度に楽観的になる傾向がある。ミニマックス・リグレット基準の一番の特徴は、意思決定志向にバランスがとれていることである (Manski 2007)。

2つ目の特徴は、意思決定が確率的に選択されることである。先ほどのキャンペーンの例であれば、表のように E と N をそれぞれ確率 $1-d$ と d でランダムに実施することで、最大リグレット最小化が達成できる。

campaign	utility $u(\cdot)$	implementation probability
E: existing	a fixed value, known	$1-d$
N: new	$b \in [b_L, b_U]$ pdf is unknown	d

リグレット(後悔) R は、最善な行動から得られる効用 ($\max(a, b)$) と実際に選択した行動から得られる効用 ($u(d, b)$) の差で、(1) 式のように d と未知の値 b の関数で定義される。

$$(1) \quad R(d, b) = \max(a, b) - u(d, b) \quad \text{where } d = \text{probability of choosing N}$$

最大リグレット最小化では、(2) 式のように最大のリグレット(ワーストシナリオ)が最小になるように d を決める。

$$(2) \quad \min_d \max_b \{R(d, b)\} = \min_d \max_b \{\max(a, b) - u(d, b)\}$$

その最適解 d^* は(3) 式で表される (Appendix 1)。

$$(3) \quad d^* = \frac{b_U - a}{b_U - b_L}$$

キャンペーン E の効用 a も未知で、その区間しか分からない状況 ($a \in [a_L, a_U]$) に拡張すると、最大リグレット最小化の d^* は(4) 式で表される (Appendix 2)。

$$(4) \quad d^* = \frac{b_U - a_L}{a_U - a_L + b_U - b_L} \quad \text{where } b_L < a_L < a_H < b_H$$

3. 手法の概要

3.1. アダプティブ・ミニマックス・リグレット (AMR)

時間が経つにつれて新たなデータが入ってくるため、長期的指標(N の効用)の不確実性 (b_L と b_U で規定される)は時間と共に減少する。したがって、その時点までに得られた情報に基づいて、最大リグレット最小化基準の最適確率 d^* を随時、更新するアプローチが AMR である。本研究のキャンペーン事例の文脈では、 t 期までに観測された顧客の生存状況から $b_U(t)$ と $b_L(t)$ を更新して、その時点における最適確率 d_t^* を(3) 式に基づいて算出し、E と N を確率的に選択、実施する。

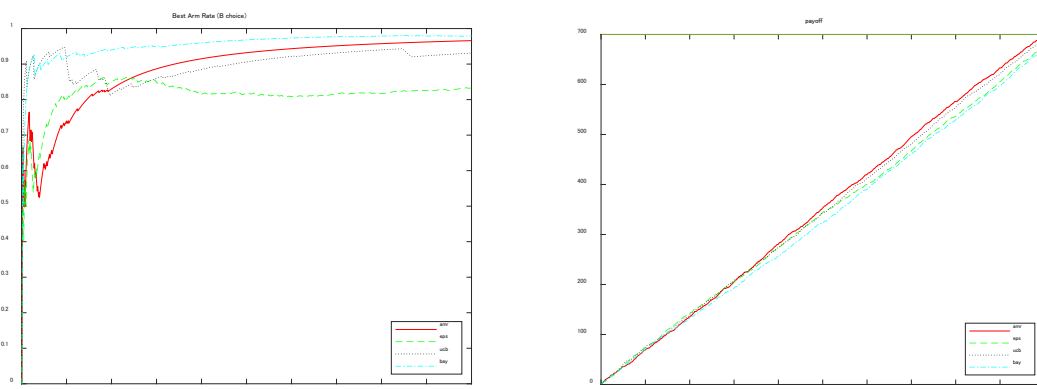
3.2. バンディット・アルゴリズムとの比較

今回の事例は、E と N どちらのアームのリターンが高いか不明な状況で、なるべく当選確率の高い方のアームを引き続けながら累積リターンを最大化するというバンディット・アルゴリズムのフレームワークと共通点がある。バンディット・アルゴリズムでは、リターンの高いアームを引きたい「活用」と各アームのリターンを把握したい「探索」とのバランスをどう執るかがカギになるため、さまざまな手法が提案されている (Lattimore and Szepesvári 2020)。そこで、まずは AMR の有効性を、バンディット・アルゴリズムでよく使われる3手法 (ϵ -greedy, Upper Confidence Bound (UCB), Thompson Sampling (Bayesian)) と比較検証する。

4. シミュレーション実験

実験では当選確率の異なる A と B 2つのアーム ($a=0.5$, $b=0.7$) を用意する。ここでの目的は、当選確率が未知である2つのアームを1000回引き続けながら、累積当選回数を最大化することである。どちらかのアームを引くたびに当たり／はずれの結果が観測されるため、AMR では、各アームの当選確率の区間 (不確実性) をその時点までに得られた情報から推定した 95%信頼区域によって更新する。

左図は、アームを引く回数を X 軸に、リターンが高いアーム B の方を引く確率を Y 軸にプロットしたものである。最初は試行錯誤による探索が行われているが、最終的には B を引く確率が 1.0 に近づくことが分かる。AMR はベイズ法に基づいた Thompson sampling の次にパフォーマンスが良かった。右図は、アームを引く回数を X 軸に、累積リターンを Y 軸にプロットしている。リターンが高いアーム B のみを前回引き続けると、傾きが 0.7 の直線になる。ここでは AMR のパフォーマンスが一番優れていた。



このシミュレーション実験では、AMR がバンディット・アルゴリズムでよく使われる既存の手法と同等かそれ以上のパフォーマンスを生み出すことが分かった。

5. マーケティングへの活用

試行回数の増加によって、アームの当選確率という静的な母数の推定精度が向上することを賢く利用するのがバンディット・アルゴリズムである。一方、AMR の真のアドバンテージは、バンディット・アルゴリズムでは対応できない、動的な母数がセンサーされているような状況において活用できることである。このことを、冒頭で紹介した顧客生存期間に関するキャンペーン事例で見てみよう。

既存キャンペーン E では、全 t 期の顧客維持率が既知で、平均生存期間 a は 2.7 年 ($=0.8+0.7+0.6+0.6$) であることが、過去の実績から分かっている。新たなキャンペーン N が候補として挙がっており、E と N、最初の 4 年間で平均生存期間(効用)が長くなる施策を活用したい(5 年目以降は考慮しない)。

新規キャンペーン N の顧客維持率データは存在しないため、平均生存期間は実際に 4 年間運用されるまで未知である。しかしながら、運用開始後は各 t 期の顧客維持率 (0.9, 0.8, 0.7, 0.7) を観測することで、平均生存期間の不確実性 ($b_U(t)$ と $b_L(t)$) が減少する。そこで、時点 t までに得られた情報に基づいて、AMR によってキャンペーン N の最適割合 d_t^* を更新する。この結果をまとめたものが下の表である。

year	survival probability		lifetime bound		proportion of N	ave. lifetime
t	E	N	b_L	b_H	d_t^*	
0	1	1	0	4	0.325	2.83
1	0.8	0.9	0.9	3.6	0.333	2.83
2	0.7	0.8	1.7	3.3	0.375	2.85
3	0.6	0.7	2.4	3.1	0.571	2.93
4	0.6	0.7	3.1	3.1	1.000	3.1
ave. lifetime	2.7	3.1				3.1

例えば $t = 2$ の行では、生存期間 0 年(1 年未満)の確率が 0.1 ($=1-0.9$)、1 年(1 年以上 2 年未満)の確率が 0.1 ($=0.9-0.8$)なので、残りの確率 0.8 で最短 2 年、最長 4 年なので、 $b_L=1.7$ ($=0.1 \times 0 + 0.1 \times 1 + 0.8 \times 2$)、 $b_H=3.3$ ($=0.1 \times 0 + 0.1 \times 1 + 0.8 \times 4$) となる。

キャンペーン N の平均生存期間は 3.1 年だが、それは4年後まで分からない。時間 t が進むにつれて、平均生存期間の不確実性が減少している。それに伴って、ミニマックス基準で最適なキャンペーン N の割合 d_t^* が増加して、最終的には 1.0 になる(左図)。E と N を確率的に d_t^* で混合したキャンペーンを実施することによって、平均生存期間は 2.83 年から 3.1 年に伸びている(右図)。



6. 結論

本研究では、キャンペーン E と N の選択に、各時点までに得られた結果に基づいてミニマックス・リグレット基準を当てはめ、それを逐次的に更新するアプローチ、AMR を提案した。AMR では、ミニマックス・リグレット基準を達成するために、E と N、2つのキャンペーンを最適な確率で混合して実施する。

顧客生存期間のような長期的指標を評価するためには、長い観測期間を要する。ビッグデータで統計誤差が無くても把握はできない。現時点での情報に基づいて最適な混合確率を算出し、それを逐次的に更新することで、企業は長期的指標の結果が全て分かる前に対応することが可能になる。マーケティングにおける長期的な指標としては、顧客生存期間以外にも、製品・サブスクリプション・サービスのリピート率などにも応用が可能である。

参考文献

- Lattimore, T. and C. Szepesvári (2020), *Bandit Algorithms*, Cambridge University Press.
- Manski, C. (2007), *Identification for Prediction and Decision*, Cambridge, Mass.: Harvard University Press.
- Manski, C. (2011), “Choosing Treatment Policies Under Ambiguity,” *Annual Review of Economics*, 2011.3: 25–49.
- Savage, L. J. (1954), *The foundations of statistics*, John Wiley & Sons.

Appendix 1: Derivation of d^* for Minimax Regret Criterion when b is unknown

Recall the payoff of the two campaigns E and N.

campaign	utility $u(\cdot)$	implementation probability
E: existing	a fixed value, known	$1-d$
N: new	$b \in [b_L, b_U]$ pdf is unknown	d

Without loss of generality, assume that the support (uncertainty interval) of b is such that $b_L < a < b_H$.

Utility $u(\cdot)$ is express as a function of d and the unknown payoff b as (A1).

$$(A1) \quad u(d, b) = (1 - d)a + d \cdot b = a + (b - a)d$$

Since the regret is defined as follows,

$$R(d, b) = \max(a, b) - u(d, b) \quad \text{where } b \text{ is unknown}$$

decision d under the Minimax regret criterion is shown in (A2)

$$(A2) \quad \min_d \max_b \{R(d, b)\} = \min_d \max_b \{\max(a, b) - u(d, b)\}$$

Maximum regret $\max_b \{R(d, b)\}$ is expressed as (A3).

$$\begin{aligned}
 \max_b \{\max(a, b) - u(d, b)\} &= \max_b \{l(b > a)[b - \{a + (b - a)d\}] + l(a > b)[a - \{a + (b - a)d\}]\} \\
 &= \max_b \{l(b > a)(b - a)(1 - d) + l(a > b)(a - b)d\} \\
 &= \max\{(b_U - a)(1 - d), (a - b_L)d\}
 \end{aligned}$$

(A3)

Since decision d^* that minimizes (A3) satisfies $(b_U - a)(1 - d^*) = (a - b_L)d^*$,

$$(3) \quad d^* = \frac{b_U - a}{b_U - b_L}$$

Figure A1 depicts the maximum regret as a function of d and the optimum decision d^* .

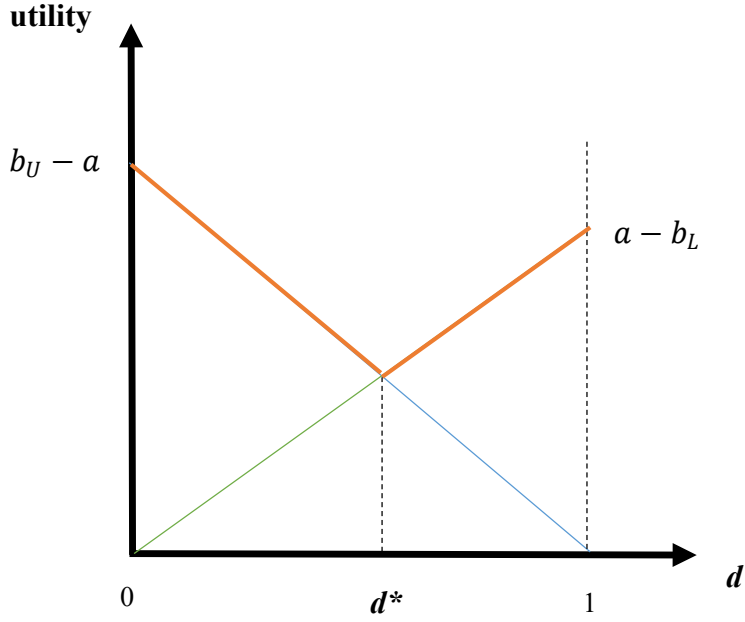


Figure A1. d^* for minimax regret criterion when b is unknown.

Appendix 2: Derivation of d^* when both a and b are unknown

Without loss of generality, assume that the supports (uncertainty intervals) of a and b is such that $b_L < a_L < a_H < b_H$ (b is more uncertain).

Utility $u(\cdot)$ is express as a function of d and the unknown payoffs a and b as (A4).

$$(A4) \quad u(d, b, a) = (1 - d)a + d \cdot b = a + (b - a)d$$

Maximum regret $\max_b \{R(d, b, a)\}$ is expressed as (A5).

$$\begin{aligned}
 & \max_{b,a} \{\max(a, b) - u(d, b, a)\} \\
 &= \max_{b,a} \{l(b > a)[b - \{a + (b - a)d\}] + l(a > b)[a - \{a + (b - a)d\}]\} \\
 &= \max_{b,a} \{l(b > a)(b - a)(1 - d) + l(a > b)(a - b)d\} \\
 &= \max\{(b_U - a_L)(1 - d), (a_U - b_L)d\}
 \end{aligned}$$

(A5)

Since decision d^* that miminizes (A3) satisfies $(b_U - a_L)(1 - d^*) = (a_U - b_L)d^*$

$$(4) \quad d^* = \frac{b_U - a_L}{a_U - a_L + b_U - b_L}$$

Figure A2 depicts the maximum regret as a function of d and the optimum decision d^* .

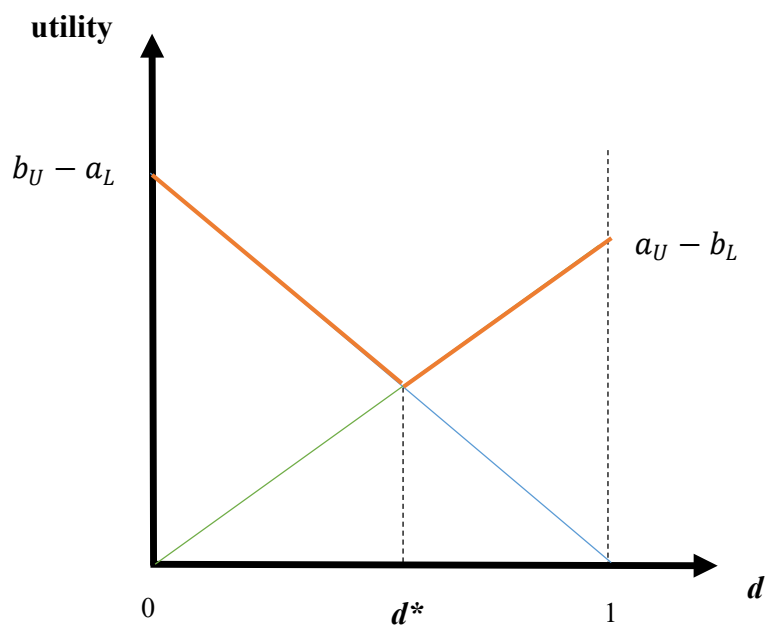


Figure A2. d^* for minimax regret criterion when both a and b is unknown.