

探し方の優劣は、ある時点の損ではなく、損の伸び方で決まる

— 十倍ごとの損の伸びは、UCB では **5.982350**・**3.151112**・**1.828572** 倍と縮み、決まった割合で探す方法では **8.360495** 倍まで上がる — 分離子は、ある回数での損を比べるか、十倍ごとの伸びを比べるか — [論文873]

山岸 勇輝 (Yuuki YAMAGISHI)

Yamagishi Complexity Torus Lab, Japan | 山岸勇輝ラボ

2026年9月

YOI: <https://www.yamagishiyuuki.com/yoi/873>

— *From thinking to syncing.* —

要旨

どの探し方が良いかは、同じ回数だけ引いて損を比べればわかる、と思いがち。本稿は、比べる回数を十倍ずつ伸ばすと勝ち負けが入れ替わることを、三本の腕で数える。新しい定理も法則も主張しない。第一に、一万回では、決まった割合で探す方法が勝つ。損は割合 **0.1** で探す方法が **122.396667**、試した回数の少ない腕に上乗せする UCB が **145.970000** である (第2節)。第二に、これが本稿の芯である。十万回では入れ替わる。損は **1023.296667** と **266.916667** で、決まった割合で探す方法が UCB の **3.833768** 倍になる (第2節)。第三に、入れ替わりを決めるのは損の伸び方である。十倍ごとの伸びは、UCB では **5.982350**・**3.151112**・**1.828572** 倍と縮み、決まった割合で探す方法では **4.345181**・**5.090670**・**8.360495** 倍と上がる (第3節)。第四に、探すのをやめると、損は時間に比例して積もる。最初に一回ずつ引いて後は平均の高い腕だけを引くと、**30** 回中 **10** 回は最良でない腕に居着き、十倍ごとの伸びは **9.914480** 倍、十万回の損は **5004.796667** である (第3節)。第五に、決まった割合で探し続けても、損は時間に比例して積もる。最良の腕を見つけたあとも **0.1** の割合で他の腕を引き続けるので、伸びは時間に比例する **10** 倍へ上がっていく (第3節)。第六に、分離子は、ある回数での損を比べるか、十倍ごとの伸びを比べるかである。ある回数での比較は、その回数を選んだ時点で勝者を決めてしまう (第4節)。キーワード：多腕バンディット、UCB、 ϵ 貪欲、損、探索と活用。

本稿の射程（射程注記）：新しい定理も法則も主張しない。多腕バンディット、決まった割合で探す方法（ ϵ 貪欲）、試した回数の少ない腕に上乗せする方法（UCB1）、よい方法では損が対数でしか増えないことは、いずれも既知である。木に入らない。モンテカルロ木探索のように、この上乗せを木の各節で使う方法には入らない。実在の応用を扱わない。広告の出し分けや治験の割り付けは主題ではない。上界を証明しない。損の理論上の上界や下界は扱わない。三本の腕で 30 回ずつ平均した数だけを見る。既刊との関係：論文387 は、いつ出るかはその場所の中では決まらないことを示し、射程注記で探索と活用などへの転用を主張しなかった。本稿はその席に座る。ただし採餌のモデルを転用するのではなく、多腕バンディットという別のモデルで、探すことと稼ぐことの配分を数える。加えたのは、三つの方法の損を百回から十万回まで十倍ごとに並べたこと、一万回と十万回で勝ち負けが入れ替わると測ったこと、十倍ごとの伸びが縮む方法と上がる方法を分けたこと、分離子を「ある回数の損か、伸び方か」に置いたことである。

1. 記法と問い

設定：当たりの確率 $\mu_1 = 0.6 \cdot \mu_2 = 0.5 \cdot \mu_3 = 0.4$ の三本の腕を、毎回一本ずつ引く。 T 回までの損を、引いた腕と最良の腕の確率の差の合計とする。

$$\text{損}(T) = \sum_{t=1}^T (\mu_1 - \mu_{a_t}), \quad \text{UCB の選び方: } a_t = \arg \max_i \left(\hat{\mu}_i + \sqrt{\frac{2 \ln t}{n_i}} \right)$$

$\hat{\mu}_i$ はそれまでの平均、 n_i は引いた回数である。三つの方法はどれも最初に一回ずつ引き、そのあと、探すのをやめる方法は平均の最も高い腕だけを、決まった割合で探す方法は 0.1 の割合で無作為の腕を、残りは平均の最も高い腕を引く。十万回を 30 回ずつくり返して平均した。

問い：同じ回数だけ引いて損を比べれば、どの探し方が良いかわかるのか。

2. 回数ごとの損

計算 2.1：三つの方法の損を、百回から十万回まで十倍ごとに出した。

方法	100 回	1000 回	10000 回	100000 回
探すのをやめる	6.480000	54.736667	504.796667	5004.796667
割合 0.1 で探す	5.533333	24.043333	122.396667	1023.296667
UCB	7.743333	46.323333	145.970000	266.916667

観察 2.2

これが本稿の芯である。一万回と十万回で勝ち負けが入れ替わる。一万回までは、割合 0.1 で探す方法の損が UCB より少ない。十万回では UCB が下回り、割合 0.1 で探す方法の損は UCB の 3.833768 倍になる。UCB が下回った区切りは十万回だけだった。

3. 十倍ごとの伸び

計算 3.1：回数を十倍にしたとき、損が何倍になるかを出した。

方法	100 → 1000	1000 → 10000	10000 → 100000
探すのをやめる	8.447016	9.222276	9.914480
割合 0.1 で探す	4.345181	5.090670	8.360495
UCB	5.982350	3.151112	1.828572

観察 3.2

入れ替わりを決めるのは伸び方である。UCB の伸びは 5.982350 から 1.828572 倍へ縮み、割合 0.1 で探す方法の伸びは 8.360495 倍へ上がる。UCB は引いた回数の少ない腕ほど大きく上乘せするので、悪い腕は次第に引かれなくなる。割合 0.1 で探す方法は、最良の腕を見つけたあとも同じ割合で悪い腕を引き続けるので、伸びは時間に比例する 10 倍へ上がっていく。

観察 3.3

探すのをやめると、はじめの運で居着く。最初に一回ずつ引いて後は平均の高い腕だけを引くと、30 回中 10 回 (0.333333) は最良でない腕をいちばん多く引いていた。十倍ごとの伸びは 9.914480 倍で、十万回の損は 5004.796667、UCB の 18.750409 倍である。

4. 分離子

観察 4.1

分離子は、ある回数での損を比べるか、十倍ごとの伸びを比べるかである。一万回で比べれば割合 **0.1** で探す方法が、十万回で比べれば **UCB** が勝つ。ある回数での比較は、その回数を選んだ時点で勝者を決めてしまう。伸びを比べれば、どちらがいずれ勝つかは回数を選ばずにわかる。

観察 4.2

論文**387**の席に座る。**387**は、いつ出るかがその場所の中では決まらないことを示し、探索と活用への転用を主張しなかった。ここでは採餌の模型ではなく多腕バンディットで、探すことと稼ぐことの配分を数えた。どちらの問いでも、今いる腕（場所）の中だけを見ていては決まらない。

観察 4.3

言わないこと。木の探索も、実在の広告や治験も、損の理論上の上界も扱わない。示したのは、探し方の優劣が、比べる回数ではなく損の伸び方で決まるということである。

5. 点呼と結び

1. 損は最良の腕との差の合計（第1節）：三本の腕。
2. 一万回で **122.396667** 対 **145.970000**（計算 2.1）：割合で探すほうが少ない。
3. 十万回で **1023.296667** 対 **266.916667**（計算 2.1）：入れ替わる。
4. 比 **3.833768**（観察 2.2）：本稿の芯。
5. **UCB** の伸び **1.828572** 倍（計算 3.1）：縮んでいく。
6. 割合で探す伸び **8.360495** 倍（計算 3.1）：上がっていく。
7. やめる方法の伸び **9.914480** 倍（計算 3.1）：時間に比例。
8. **30** 回中 **10** 回居着く（観察 3.3）：はじめの運。
9. 十万回でやめる方法は **18.750409** 倍（観察 3.3）：**UCB** に対して。
10. ある回数の損か、伸び方か（観察 4.1）：分離子・論文**387**の席。

結び：探し方の優劣は、ある時点の損ではなく、損の伸び方で決まる。三本の腕を一万回引くと、割合 **0.1** で探す方法の損 **122.396667** が UCB の **145.970000** より少ないが、十万回では **1023.296667** と **266.916667** になり、比は **3.833768** である。UCB の十倍ごとの伸びは **1.828572** 倍まで縮み、割合で探す方法は **8.360495** 倍まで上がるからである。探すのをやめると **30** 回中 **10** 回は最良でない腕に居着き、伸びは **9.914480** 倍で時間に比例する。決まった割合で探し続けても、最良の腕を見つけたあとの探しが損として積もる。既刊との位置：論文 **387** が射程注記で空けた探索と活用の席に、採餌の模型ではなく多腕バンディットで座った。正直に言えば、腕は三本、くり返しは **30** 回で、損の理論上の上界は扱っていない。

参考文献

1. Auer, P., Cesa-Bianchi, N., & Fischer, P. (2002). Finite-time Analysis of the Multiarmed Bandit Problem. *Machine Learning*.
2. Sutton, R. S., & Barto, A. G. (2018). *Reinforcement Learning: An Introduction* (2nd ed.). MIT Press.
3. 山岸勇輝 (2026). 『いつ出るかは、その場所の中では決まらない』 [論文387]. YOI: <https://www.yamagishiyuuki.com/yoi/387>
・ DOI: 10.5281/zenodo.22699131

— *From thinking to syncing.* —

山岸 勇輝 (Yuuki YAMAGISHI)

© 2026 Yamagishi Complexity Torus Lab, Japan · Licensed under Creative Commons Attribution 4.0 International (CC BY 4.0)