バックテストの成績は、戦略の質ではなく、試した回数を測っている

— 実力がまったく無い戦略でも、5 年の日次データで測った年率シャープレシオは標準誤差 0.447214 でばらつく — 分離子は、有効試行回数が報告されているか — [論文706]

山岸 勇輝 (Yuuki YAMAGISHI)
Yamagishi Complexity Torus Lab, Japan｜山岸勇輝ラボ
2026年9月

— From thinking to syncing. —

要旨

バックテストの成績は、その戦略がどれだけ良いかを示すものとして読まれる。だが実力がまったく無くても、たくさん試せば良い成績が必ず出てくる。何個試せばどれだけ出るのかを数える。新しい定理も法則も主張しない。 第一に、実力ゼロでも、成績はばらつく——5 年の日次データで測った年率シャープレシオは、標準誤差 **0.447214** でばらつく（第2節）。 第二に、これが本稿の芯である。たくさん試して最良を取れば、必ず良い数字が出る——$N = 1000$ 個試せば、実力ゼロでも最良は **1.449610** になる（第2節）。 第三に、要る試行数は、驚くほど少ない——最良が 1.0 を超えるのに要るのは、わずか **49** 個である（第3節）。 第四に、増え方は、対数でしかない——$N = 10$ で **0.688147**、$N = 100$ で **1.121425**、$N = 10000$ でも **1.722490** で、1000 倍試しても 2.503086 倍にしかならない（第2節）。 第五に、二通りの出し方が一致する——数値積分の 1.449610 に対し、乱数 4000 回では **1.451801** で、差は **0.002191** である（第3節）。 第六に、分離子は、有効試行回数が報告されているかである——報告されていなければ、成績の数字は原理的に解釈できない（第4節）。 キーワード：バックテストの過適合、シャープレシオ、多重比較、最大値の期待値、有効試行回数、選択の偏り。

本稿の射程（射程注記）：新しい定理も法則も主張しない——多重比較による選択の偏り、シャープレシオの推定誤差、最大値の期待値の対数的な増え方は、いずれも既知である（ベイリーほか 2014、シャープ 1994）。実在の戦略を扱わない——実在の戦略にも市場のデータにも当てはめない。実力がまったく無い戦略だけを考える。推定誤差を単純化する——年率シャープレシオの標準誤差は $1/\sqrt{\text{年数}}$ の目安に固定した。自己相関や裾の厚さは入れていない。防ぐ方法を評価しない——期間外検定やデフレートしたシャープレシオには触れるにとどめ、比べない。個別のことを言わない——特定の運用、商品、人については何も述べない。既刊との関係：論文317は、有意水準が「一回あたり」の約束であって研究あたりではないことを示した——そこでは多重比較の代償を二値の判定（有意か否か）で数えた。ここでは同じ代償を連続量の最大値で数える。二値なら「有意ではない」と言われて止まるが、連続量はそのまま成績の数字として報告されてしまう。加えたのは、試行数ごとの最良シャープレシオを数値積分で出したこと、同じ値を乱数で出して差 **0.002191** で突き合わせたこと、1.0 に届くのに **49** 個で足りることを示したこと、上界 $\sqrt{2\ln N}$ が実測より大きいことを示したこと、分離子を「有効試行回数が報告されているか」に置いたことである。

1. 記法と問い

設定：期待収益がちょうどゼロの戦略を考える。T 年の日次データで測った年率シャープレシオは、目安として標準誤差 $1/\sqrt{T}$ でばらつく。N 個試して最良を取る。

$$\mathrm{SE} = \frac{1}{\sqrt{T}}, \qquad \mathbb{E}\left[\max_{1\le i\le N} Z_i\right], \qquad \mathbb{E}\left[\max \mathrm{SR}\right] = \mathrm{SE}\cdot\mathbb{E}\left[\max Z_i\right]$$

Z_i は標準正規に従う。$T = 5$ 年とすると $\mathrm{SE} = 0.447214$ である。

問い：良い成績は、戦略の質を測っているのか、試行の数を測っているのか。

2. 実力ゼロでも、良い数字は出る

計算 2.1：試行数ごとに、最良シャープレシオの期待値を数値積分で出した。

試行数	最良シャープレシオ	上界 $\sqrt{2\ln N}$ 換算
1	−0.000004	—
10	0.688147	0.959705
100	1.121425	1.357228
1000	1.449610	1.662258
10000	1.722490	1.919410

> **観察 2.2**
>
> これが本稿の芯である。$N = 1$ ではゼロなのに、$N = 1000$ では 1.449610 になる。戦略は一つも変わっていない。変わったのは試した回数だけである——報告された数字は、戦略の質ではなく試行の数を測っている。

> **観察 2.3**
>
> 増え方は、対数でしかない。N を 10 から 10000 へ 1000 倍にしても、最良は 0.688147 から 1.722490 へ 2.503086 倍になるだけである——だから「たくさん試した」と聞いても、どれだけ試したかは成績からは読めない。

> **観察 2.4**
>
> 上界は、実測より大きい。$\sqrt{2\ln N}$ 換算は $N = 1000$ で 1.662258 で、実測の 1.449610 を上回る——上界で見積もると過大になるので、期待値そのものを出す必要がある。

3. 要る試行数は、少ない

計算 3.1：最良が 1.0 を超えるのに要る試行数を求めると、49 個であった。

> **観察 3.2**
>
> 49 個は、意図して探さなくても届く数である。指標を数個、期間を数通り、閾値を数通り組み合わせれば、すぐに超える——「1000 個試した」と自覚している人は少ないが、49 個なら誰でも試している。

計算 3.3：同じ値を乱数 4000 回でも出した。$N = 1000$ で 1.451801 となり、数値積分の 1.449610 との差は 0.002191 である。

> **観察 3.4**
>
> 二通りで出したのは、数の出どころを一つにしないためである。積分の刻みが粗ければ積分が狂い、回数が少なければ乱数が狂う。どちらも狂っていれば一致しない——一致は、両方が正しいことの証拠になる。

> **観察 3.5**
>
> だから「有効試行回数」が要る。報告された成績だけからは、N が 1 だったのか 1000 だったのかが区別できない——同じ 1.4 という数字が、実力の証拠にも、試行の数の証拠にもなる。

4. 分離子

> **観察 4.1**
>
> 分離子は、有効試行回数が報告されているかである。報告されていれば、成績をその数で割り引いて読める。されていなければ、その数字は原理的に解釈できない——足りないのは精度ではなく、情報の一項目である。

> **観察 4.2**
>
> 論文317 と並べる。317 は、有意水準が「一回あたり」の約束であって研究あたりではないことを示した——そこでは多重比較の代償が二値の判定に出た。ここでは連続量の最大値に出る。二値なら「有意ではない」と言われて止まるが、連続量はそのまま成績の数字として報告されてしまう。止め方が無いぶん、こちらのほうが危うい。

> **観察 4.3**
>
> 言わないこと。実在の戦略にもデータにも当てはめず、防ぐ方法も比べていない——示したのは、実力ゼロでも試行数だけで数字が出ることと、要る試行数が少ないことである。

5. 点呼と結び

1. 標準誤差 **0.447214**（第1節）：5 年の日次データ。
2. $N = 1$ で **−0.000004**（計算 2.1）：実力ゼロ。
3. $N = 1000$ で **1.449610**（計算 2.1）：本稿の芯。
4. $N = 10$ でも **0.688147**（計算 2.1）：すぐ出る。
5. $N = 10000$ で **1.722490**（計算 2.1）：増え方は対数。
6. 上界は **1.662258**（計算 2.1）：実測より大きい。
7. **1.0** に **49** 個（計算 3.1）：意図せず届く。
8. 乱数では **1.451801**（計算 3.3）：二通りで出した。
9. 差は **0.002191**（計算 3.3）：一致が証拠になる。
10. 有効試行回数が報告されているか（観察 4.1）：分離子・論文317。

結び：バックテストの成績は、戦略の質ではなく、試した回数を測っている。実力がまったく無い戦略でも、5 年の日次データで測った年率シャープレシオは標準誤差 0.447214 でばらつく——そこから N 個試して最良を取れば、$N = 10$ で 0.688147、$N = 1000$ で 1.449610、$N = 10000$ で 1.722490 が出る。戦略は一つも変わっていない。変わったのは試した回数だけである——しかも最良が 1.0 を超えるのに要るのは、わずか 49 個である。増え方は対数でしかないので、1000 倍試しても 2.503086 倍にしかならず、成績からは試行数が読めない。数値積分の 1.449610 と乱数 4000 回の 1.451801 は差 0.002191 で一致した——二通りで出したのは、数の出どころを一つにしないためである。分離子は、有効試行回数が報告されているかである。既刊との位置——論文

317 では多重比較の代償が二値の判定に出た。ここでは連続量の最大値に出るので、止め方が無い。正直に言えば——実在の戦略にもデータにも当てはめず、推定誤差も目安に固定した。

参考文献

1. Bailey, D. H., Borwein, J., Lopez de Prado, M., & Zhu, Q. J. (2014). Pseudo-mathematics and financial charlatanism: the effects of backtest overfitting on out-of-sample performance. *Notices of the American Mathematical Society*.
2. Sharpe, W. F. (1994). The Sharpe ratio. *The Journal of Portfolio Management*.
3. 山岸勇輝 (2026).『有意水準は「一回あたり」の約束であって、研究あたりではない』[論文317]. Zenodo. 10.5281/zenodo.22208819

— 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)