

記録された値が、元の記録の中に在るかを確かめる

— 一致率 **55～65%** の母集団と **1～8%** の母集団では、価格差の中央値が **0.000** と **18.05** で **70** 倍以上違う — 時刻をずらして走査すると、一方には棘が立ち、他方はどこでも平らである — [論文428]

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

二次的な記録は、元の記録と突き合わせられる。本稿が示すのは、その突き合わせが「時刻のずれ」と「そもそも在らない値」を分けられることである——分けるのは、ずらして走査したときの形である。新しい定理も法則も主張しない。第一に、一致率は母集団によって **55～65%** と **1～8%** に分かれ、**6.9** 倍以上違う（第2節）。第二に、価格差の中央値は **0.000～0.05** と **3.5～18.05** で、**70** 倍以上違う（第2節）。第三に、これが本稿の芯である。時刻をずらして走査すると、一方は **0** で棘が立ち（隣の **3.73** 倍）、他方はどこへずらしても平らである（第3節）。第四に、平らな側の最大値 **0.09** は、棘の立つ側の隣の値 **0.15** にも届かない（第3節）。第五に、判別子は倍率ではなく高さである。倍率で比べると平らな側のほうが大きく出てしまう（第3節）。第六に、分離子は「時刻が違うのか、値が無いのか」である（第4節）。

本稿の射程（射程注記）：新しい定理も法則も主張しない——記録の突合せ、時刻同期の検証、相互相関による遅延推定はすべて既知である。誰も、何も、名指さない——個人、組織、製品、事例は主題ではなく、一切登場しない。扱うのは「二次的な記録を元の記録に照合する」という手続きだけである。意図を論じない——一致しないことの原因が何かは、この手続きからは決まらない。記録の丸め、単位の違い、別系統の値、転記の誤りなど、いくつもの説明がありうる。本稿は「一致するか」だけを測る。元の記録の正しさを保証しない——元の側に欠けや誤りがあれば、一致率は下がる。片方を真とする根拠を本稿は持たない。判定の閾値を決めない——どの一致率なら合格かは、用途が決める。特定の分野を論じない——数は時系列の照合から採ったが、対象は主題ではない。既刊との関係：論文426は先読みが期間を分けても消えず、時刻をずらして初めて見えることを示した——同じ「時刻をずらす」という操作が、ここでは別のものを分ける。論文420は偏りの無い欠測でも符号が動くことを示した——そこでは記録の欠けが効いた。ここでは記録の食い違いを測る。論文300は同根か別根かが判定できることを示した——「時刻がずれている」と「その値がどこにも無い」は、その基準で別根である。論文421は先行して見えるものが先行していないことを示した——そこでもラグ走査の形が答を出した。加えたのは、二つの母集団の一致率と価格差を並べたこと、ずらし走査で棘と平らを分けたこと、判別子が倍率ではなく高さであることを示したこと、作り物で本物の記録と作った記録を同じ手続きにかけたこと、分離子を「時刻か、値の不在か」に置いたことである。

1. 記法と問い

設定：元の記録は、時刻ごとの値の列 $\{x_i\}$ である。二次的な記録は、時刻と値の対 (τ_i, p_i) の集まりである。許容差 ε とずらし量 s に対し、一致率を次で定める。

$$H(s) = \frac{1}{n} \#\{i : |x_{\tau_i+s} - p_i| \leq \varepsilon\}$$

観察 1.1

二次的な記録は、たいてい元の記録と照合されない。整合していることが前提にされる——照合には元の記録が要り、それを持っていないことが多いからである。

観察 1.2

元の記録があれば、照合は一行で書ける。時刻を引いて、値を比べるだけである——規約を一つも仮定せずにできる。

問い：一致しなかったとき、それは時刻がずれているのか、値が無いのか。

2. 一致率も価格差も、桁で分かれる

計算 2.1：二つの母集団を、同じ許容差で照合する。

母集団	件数	一致率	値の差の中央値
甲	2087	0.55～0.65	0.000～0.05
乙	3403	0.01～0.08	3.5～18.05

観察 2.2

一致率は **6.9** 倍以上違う。**0.55** 対 **0.08**——最も甲に不利な組で比べても、この差がある。

観察 2.3

値の差は **70** 倍以上違う。**3.5** 対 **0.05**——一致するかどうかの問題ではなく、桁の問題である。

観察 2.4

甲では差の下端がちょうど **0.000** である。完全に一致する組が実在する——照合の仕組みが働いていることの、直接の証拠である。

観察 2.5

甲の一致率も **100%** ではない。**55～65** パーセント——元の記録にも欠けがあるので、一致しない組が残るのは当然である。だから低い一致率だけでは何も言えない。

3. ずらして走査すると、形が分かれる

計算 3.1：時刻を前後にずらしながら、一致率を測る。

母集団	ずらさないとき	隣の時刻
甲	0.56～0.63	0.05～0.15
乙	0.01～0.09（どこでも）	同じ

甲では棘が立ち、その高さは隣の **3.73** 倍である。乙はどこへずらしても平らである。

観察 3.2

棘が立つことは、時刻が正しいことの証明である。ずらすと下がる——ずれていたなら、別の位置に棘が立つはずである。

観察 3.3

平らであることは、時刻では説明が付かないことを意味する。どこへずらしても上がらない——時刻を直せば一致する、という説明が閉じる。

観察 3.4

判別子は倍率ではなく高さである。乙の凹凸を最大と最小の比で測ると **9.00** 倍になり、甲の棘 **3.73** 倍より大きく出てしまう——小さい数どうしの比は暴れるからで、比べるべきは棘の高さ **0.56** と、平らな側の最大 **0.09** である。

観察 3.5

そして平らな側の最大 **0.09** は、棘の立つ側の隣の値 **0.15** にも届かない。甲の一番悪いところより、乙の一番良いところが低い——重ならないので、分離できる。

計算 3.6：作り物で、同じ手続きを二つの記録にかける。

ずらし	元から写した記録	元に無い値の記録
−40	0.018	0.020
−10	0.067	0.074
0	0.957	0.001
+10	0.058	0.062
+40	0.016	0.022

観察 3.7

作り物でも同じ形が出る。一方は 0 で 0.957 の棘、他方は最大でも 0.074——手続きが両方を分けられることの確認である。

4. 分離子

分離子：時刻が違うのか、値が無いのか。

観察 4.1

時刻の違いは、ずらせば直る。棘が別の位置に立つ——直したあとは高い一致率が出る。

観察 4.2

値の不在は、ずらしても直らない。どの位置にも棘が立たない——その値は、どの時刻にも無い。

観察 4.3

これは論文300 の基準で別根である。どちらも「一致しない」と呼ばれる——一方は対応づけの問題、一方は存在の問題である。

観察 4.4

そして一致しない理由までは、この手続きからは決まらない。丸め、単位の違い、別系統の値、転記の誤り——いくつもの説明がありうる。本稿が測るのは「一致するか」だけである。

観察 4.5

元の記録を真とする根拠も、本稿は持たない。元の側に欠けや誤りがあれば、一致率は下がる——甲が100%でないのは、その現れでもある。

観察 4.6

正直に言えば——低い一致率だけでは何も言えない。元の記録の欠けでも下がるからである——効いているのは、ずらし走査の形のほうである。棘が立つか、平らか。

観察 4.7

最後に一つ——照合できる記録は、照合しておくとい。規約を一つも仮定せずにできる——そして一致した部分については、そこから先の計算を安心して積める。

5. 点呼と結び

1. 一致率は **6.9** 倍以上違う（観察 2.2）：**0.55** 対 **0.08**。
2. 値の差は **70** 倍以上違う（観察 2.3）：**3.5** 対 **0.05**。
3. 甲の差の下端はちょうど **0.000**（観察 2.4）：完全一致が実在。
4. 甲の一致率も **100%** ではない（観察 2.5）：元にも欠けがある。
5. 甲の棘は隣の **3.73** 倍（計算 3.1）：時刻は正しい。
6. 乙はどこでも平ら（計算 3.1）：時刻では説明が付かない。
7. 倍率で測ると乙が **9.00** 倍（観察 3.4）：判別子は高さ。
8. 乙の最大 **0.09** < 甲の隣 **0.15**（観察 3.5）：重ならない。
9. 作り物では **0.957** 対 **0.074**（計算 3.6）：手続きが分ける。

10. 分離子是对应づけか存在か（観察 4.3）：論文300 の基準で別根。

結び：二次的な記録は、たいてい元の記録と照合されない——整合していることが前提にされるからで、照合には元の記録が要り、それを持っていないことが多い。だが元の記録があれば照合は一行で書ける——時刻を引いて値を比べるだけで、規約を一つも仮定せずにできる。照合すると、桁で分かれる——一致率は 0.55 対 0.08 と 6.9 倍以上、値の差の中央値は 3.5 対 0.05 と 70 倍以上違う。一方では差の下端がちょうど 0.000 で、完全に一致する組が実在する——照合の仕組みが働いていることの直接の証拠である。ただしその一致率も 55 ～65 パーセントで 100% ではない。元の記録にも欠けがあるので一致しない組は残り、だから低い一致率だけでは何も言えない。分けるのは、ずらして走査したときの形である——一方はずらさないところに棘が立ち、その高さは隣の 3.73 倍である。ずらすと下がるのだから、時刻は正しい。他方はどこへずらしても 0.01～0.09 で平らである。上がる位置が無いのだから、「時刻を直せば一致する」という説明は閉じる。判別子は倍率ではなく高さである——平らな側の凹凸を最大と最小の比で測ると 9.00 倍になり、棘の 3.73 倍より大きく出てしまう。小さい数どうしの比は暴れるので、比べるべきは棘の高さ 0.56 と平らな側の最大 0.09 である。そしてその 0.09 は、棘の立つ側の隣の値 0.15 にも届かない。一番悪いところより一番良いところが低いので、重ならず分離できる。作り物でも同じ形が出て、元から写した記録は 0 で 0.957 の棘を立て、元に無い値の記録は最大でも 0.074 にとどまる。分けているものは一つ——時刻が違うのか、値が無いのか。時刻の違いはずらせば直り、棘が別の位置に立つ。値の不在はずらしても直らず、どの位置にも棘が立たない。論文300 の基準で別根であり、一方是对应づけの問題、一方は存在の問題である。正直に言えば——一致しない理由までは、この手続きからは決まらない。丸め、単位の違い、別系統の値、転記の誤りなど、いくつもの説明がありうる。本稿が測るのは「一致するか」だけである。元の記録を真とする根拠も本稿は持たない——元の側に欠けや誤りがあれば一致率は下がり、一方が 100% でないのはその現れでもある。最後に一つ——照合できる記録は、照合しておくときよい。規約を一つも仮定せずにできるし、一致した部分については、そこから先の計算を安心して積める。

参考文献

1. Fellegi, I. P., & Sunter, A. B. (1969). A Theory for Record Linkage. *Journal of the American Statistical Association*.
2. Christen, P. (2012). *Data Matching*. Berlin: Springer.
3. Redman, T. C. (1998). The Impact of Poor Data Quality on the Typical Enterprise. *Communications of the ACM*.
4. Knuth, D. E. (1998). *The Art of Computer Programming: Sorting and Searching* (2nd ed.). Reading: Addison-Wesley.
5. 山岸勇輝 (2026). 『三百本目の点呼 — 同根か別根かは、判定できる』[論文300]. Zenodo. 10.5281/zenodo.22171914
6. 山岸勇輝 (2026). 『偏りを生まない欠測でも、符号は動く』[論文420]. Zenodo. 10.5281/zenodo.22699994
7. 山岸勇輝 (2026). 『「先行して見える」ものは、先行していない』[論文421]. Zenodo. 10.5281/zenodo.22700022
8. 山岸勇輝 (2026). 『先読みは、期間を分けても高い t 値でも消えない』[論文426]. Zenodo. 10.5281/zenodo.22700184

— *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)