データを増やすほど、罰則つきの規準は複雑なほうを選ぶようになる

— 周転円が勝ちはじめる n は AIC で **53.777041**（閉じた式 $2\Delta k/\ln$（残差の比）と **6** 桁一致）、BIC で **131.109261** — 分離子は、罰則が n とともにどう増えるか — [論文701]

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

複雑な模型は罰されるはずだ、と言われる。だが罰則つきの規準に入れても、データを増やすほど複雑なほうが勝つ。罰則と当てはまりが n に対してどう伸びるかを、交差点の数で示す。新しい定理も法則も主張しない。 第一に、当てはまりだけで選べば、複雑なほうが勝つ——残差は周転円 0.004、楕円 0.005 で、母数を 11 使える側が勝つ（第2節）。 第二に、罰則を入れても、$n = 100$ では二つの規準が食い違う——AIC は周転円 -990.663110 対 楕円 -980.348755 で周転円を、BIC は -962.006238 対 -967.322904 で楕円を選ぶ（第2節）。 第三に、これが本稿の芯である。データを増やすほど、複雑なほうが勝つ——周転円が勝ちはじめる n は AIC で 53.777041、BIC で 131.109261 であり、それより上ではずっと周転円が選ばれる（第3節）。 第四に、交差点には閉じた式がある——AIC では $2\Delta k$ を残差の比の対数で割った値で、二分法の 53.777041 と 6 桁一致した（第3節）。 第五に、理由は、伸び方の違いだけである——当てはまりの差は n に比例して伸びるのに、罰則の差は定数（AIC）か対数（BIC）でしか伸びない。$n = 5000$ で AIC の差は $+1103.717757$ になる（第3節）。 第六に、分離子は、罰則が n とともにどう増えるかである——罰則があるかどうかではない。n に比例して増える罰則でなければ、いずれ追い越される（第4節）。 キーワード：模型選択、AIC、BIC、罰則項、過適合、標本数の効き。

本稿の射程（射程注記）：新しい定理も法則も主張しない——AIC・BIC の定義、罰則の項の伸び方、過適合は、いずれも既知である（赤池 1974、シュワルツ 1978）。天文データに当てはめない——周転円と楕円の母数と残差は前提として置く代表値であり、歴史上の当てはまりの良さを主張しない。選択の方法を並べない——交差検証や予測誤差による選択には触れるにとどめ、比較はしない。パラダイム論を扱わない——科学史における理論の交代の理論は論じない。量の話だけをする。残差の扱いを明示する——残差平方和は模型の性質として n に依らず固定した。実際には n とともに増えるので、そこは理想化である。既刊との関係：論文536 は、式の形を借りたのか導いたのかは、当てはめた範囲の外に出れば分かることを示した——そこでは範囲の外が分かれ目だった。ここでは範囲の内側にとどまるかぎり、罰則を入れても分かれないこと、しかもデータを増やすほど誤った側へ寄ることを数える。536 の主張の裏側にあたる。加えたのは、$n = 100$ で AIC と BIC の判定が食い違うことを示したこと、交差点を **53.777041** と **131.109261** で出したこと、AIC の交差点が閉じた式と 6 桁一致することを示したこと、差が n に比例して開くことを $n = 5000$ の **+1103.717757** で示したこと、分離子を「罰則が n とともにどう増えるか」に置いたことである。

1. 記法と問い

設定：同じデータに二つの模型を当てる。母数の数を k、残差平方和を S とし、二つの規準を置く。

$$\mathrm{AIC} = n \ln \frac{S}{n} + 2k, \qquad \mathrm{BIC} = n \ln \frac{S}{n} + k \ln n$$

周転円 5 段を $k = 11$・$S = 0.004$、楕円＋摂動を $k = 5$・$S = 0.005$ とする。S は模型の性質として n に依らず固定した。

問い：罰則を入れれば、複雑な模型は退けられるのか。

2. $n = 100$ で、二つの規準が食い違う

計算 2.1：$n = 100$ での値を出した。

規準	周転円	楕円	選ばれるのは
残差	**0.004**	**0.005**	周転円
AIC	**−990.663110**	**−980.348755**	周転円
BIC	**−962.006238**	**−967.322904**	楕円

> **観察 2.2**
>
> 同じデータ、同じ二つの模型で、規準を替えると答が変わる。AIC は周転円を、BIC は楕円を選ぶ——「罰則つきの規準で選んだ」という一文は、どの規準かを言わなければ何も言っていない。

> **観察 2.3**
>
> 差の出どころは、罰則の項だけである。当てはまりの項は両者で共通の形なので、食い違いは $2k$ と $k \ln n$ の違いから出る——$n = 100$ では $\ln n = 4.6$ で、2 より大きい。だから BIC のほうが複雑さを重く罰する。

3. データを増やすほど、複雑なほうが勝つ

計算 3.1：n を動かして、どちらが選ばれるかを出した（差が正なら周転円）。

n	AIC の差	BIC の差
10	−9.768564	−11.584075
50	−0.842822	−12.314960
100	+10.314355	−5.316666
500	+99.571776	+74.284127
5000	+1103.717757	+1064.614597

計算 3.2：周転円が勝ちはじめる n は、AIC で 53.777041、BIC で 131.109261 である。AIC の交差点は閉じた式 $2\Delta k$ を残差の比の対数で割った値と 6 桁一致した。

> **観察 3.3**
>
> これが本稿の芯である。向きが、直感と逆である。データが少ないうちは簡単なほうが選ばれ、増やすほど複雑なほうが選ばれる——「データを増やせば正しい模型に落ち着く」は、残差に差がある二つの模型のあいだでは成り立たない。

観察 3.4

理由は伸び方だけである。当てはまりの差は n に比例し、残差の比 1.250000 の対数を掛けた分だけ毎回増える。罰則の差は AIC なら定数、BIC なら対数でしか増えない——比例に定数も対数も勝てない。だから交差点は必ず有限の n に存在する。

観察 3.5

だから罰則は「複雑さを罰する」ものではない。罰しているのは母数の数であって、当てはめすぎそのものではない——当てはまりが本当に良い模型は、罰則をいくら払っても勝つ。問題は、その当てはまりの良さが本物かどうかを、同じデータの中では言えないことである。

4. 分離子

観察 4.1

分離子は、罰則が n とともにどう増えるかである。罰則があるかどうかではない。n に比例して増える罰則でなければ、いずれ当てはまりの差に追い越される——AIC と BIC の違いも、定数か対数かという伸び方の違いにすぎず、どちらも比例には届かない。

観察 4.2

論文536 と並べる。536 は、式の形を借りたのか導いたのかは、当てはめた範囲の外に出れば分かることを示した——そこでは範囲の外が分かれ目だった。ここでは範囲の内側にとどまるかぎり、罰則を入れても分かれず、しかもデータを増やすほど誤った側へ寄る。二つ合わせて、模型を分けるのは範囲の外だけだということになる。

観察 4.3

言わないこと。天文データにも科学史にも当てはめず、残差は n に依らず固定した理想化である——示したのは、罰則の伸び方が比例に届かないという一点である。

5. 点呼と結び

1. 残差は 0.004 対 0.005（計算 2.1）：当てはまりでは周転円。
2. AIC は -990.663110 対 -980.348755（計算 2.1）：周転円を選ぶ。
3. BIC は -962.006238 対 -967.322904（計算 2.1）：楕円を選ぶ。
4. 規準を替えると答が変わる（観察 2.2）：どの規準かを言う。
5. AIC の交差点 53.777041（計算 3.2）：本稿の芯。
6. BIC の交差点 131.109261（計算 3.2）：遅いが、来る。
7. 閉じた式と 6 桁一致（計算 3.2）：二分法と突き合わせた。
8. $n = 5000$ で $+1103.717757$（計算 3.1）：差は開き続ける。
9. 残差の比は 1.250000（観察 3.4）：その対数が毎回効く。
10. 罰則がどう増えるか（観察 4.1）：分離子・論文536。

結び：データを増やすほど、罰則つきの規準は複雑なほうを選ぶようになる。同じ 100 点に対し、AIC は周転円 -990.663110 対 楕円 -980.348755 で周転円を選び、BIC は -962.006238 対 -967.322904 で楕円を選ぶ——「罰則つきの規準で選んだ」は、どの規準かを言わなければ何も言っていない。周転円が勝ちはじめる n は AIC で 53.777041、BIC で 131.109261 で、それより上ではずっと周転円が選ばれる。AIC の交差点は閉じた式と 6 桁一致し、$n = 5000$ では差が $+1103.717757$ まで開く——当てはまりの差は n に比例して伸びるのに、罰則の差は定数か対数でしか伸びないからである。比例に定数も対数も勝てないので、交差点は必ず有限の n に存在する。分離子は、罰則が n とともにどう増えるかである。既刊との位置——論文536 は範囲の外に出れば分かると示した。ここでは範囲の内側では分からず、増やすほど誤った側へ寄ることを示した。正直に言えば——残差を n に依らず固定した理想化であり、母数も残差も代表値である。

参考文献

1. Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*.
2. Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*.
3. 山岸勇輝 (2026). 『式の形を借りたのか、導いたのかは、境目の外で分かる』[論文536]. Zenodo. 10.5281/zenodo.23014123

— 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)