

乱数の軸で遅くなる確率は分散だけで上から押さえられるが、その上限は実際よりずっと甘い

— 比べる回数のばらつきの平均に対する割合は n とともに縮み、平均の 1.5 倍を超える確率の上限もゆっくり下がるが、実際に超える割合はその上限を大きく下回る — 分離子は、分散だけで押さえるか、分布の形まで見るか — [論文952]

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/952>

— *From thinking to syncing.* —

要旨

軸を乱数で選ぶクイックソートは平均では速いが、運が悪ければ遅くなりうる。本稿は、比べる回数の平均・分散・三次の積率を漸化式で正確に出し、平均の 1.5 倍や 2 倍を超える確率の上限を分散だけから出して、乱数の試行で実際に超えた割合と比べる。新しい定理も法則も主張しない。第一に、平均と分散は漸化式で正確に出せ、知られた式と一致する。 $n = 1000$ で平均 10985.912663、分散 409117.768339 で、平均は式と相対 10^{-12} 、分散は式と相対 10^{-9} 以内で一致した。分散を n の二乗で割ると $n = 2000$ で 0.414001 だった（極限 0.420264）（第2節）。第二に、ばらつきの平均に対する割合は n とともに縮む。標準偏差÷平均は $n = 100$ で 0.091816、1000 で 0.058222、2000 で 0.052037 だった（第2節）。第三に、分散だけで、遅くなる確率を上から押さえられる。平均の 1.5 倍を超える確率は、 $n = 100$ で 0.032621 以下、1000 で 0.013378 以下、2000 で 0.010715 以下で、2 倍を超える確率は 2000 で 0.002701 以下だった（第3節）。第四に、実際に超える割合は上限をずっと下回る。 $n = 100$ で 40000 回試すと 1.5 倍を超えたのは 13 回（0.000325）で、上限の 0.009963 倍だった。 $n = 1000$ では 4000 回で 0 回だった（第3節）。第五に、分離子は、分散だけで押さえるか、分布の形まで見るかである。分散だけの上限は、同じ分散を持つあらゆる分布に当てはまる最悪の値である。実際の分布は上に長い裾を持つが（歪み 0.867748）、その裾は分散だけの上限よりずっと薄い（第4節）。キーワード：クイックソート、比べる回数、分散、カンテリの不等式、裾の確率、乱数の軸。

本稿の射程（射程注記）：新しい定理も法則も主張しない。乱数の軸のクイックソートの比べる回数の平均 $2(n+1)H_n - 4n$ と分散の式、分散を n の二乗で割った値の極限 $7 - 2\pi^2/3$ 、分散だけで裾を押さえるカンテリの不等式、より強い裾の上限が知られていることは、いずれも既知である。比べる回数だけを数える。記憶の読み書き、定数倍、再帰の深さは扱わない（論文927と同じ）。強い上限を導かない。分布の形を使った指数の裾の上限は名前を挙げるにとどめ、ここでは分散だけの上限を出す。試行の 0 回を無いとは言わない。 $n = 1000$ の 4000 回で 1.5 倍を超えなかったことは、超えないことを意味しない。既刊との関係：論文927の射程注記は、乱数の運を上から押さえず、乱数の軸でも遅くなる確率がどれほど小さいかは主張しない、試行の平均を並べるだけであると書いた。本稿はその席に座る（先客なし）。論文927は、先頭を軸にするクイックソートは整った入力で最も遅く、軸を乱数で選ぶと平均の遅さが入力によらなくなることを示した。ここでは平均だけでなく、平均から外れる確率を上から押さえる。加えたのは、平均・分散・三次の積率を漸化式で正確に出し式と突き合わせたこと、標準偏差+平均の縮み方を並べたこと、1.5 倍と 2 倍を超える確率の上限を出したこと、乱数の試行の割合と上限を比べたこと、分離子を「分散だけで押さえるか、分布の形まで見るか」に置いたことである。

1. 記法と問い

設定：長さ n の列を、軸を一様な乱数で選ぶクイックソートで並べる。比べる回数を C_n とすると、軸が K 番目のとき $C_n = (n-1) + C_{K-1} + C'_{n-K}$ （二つは独立）である。このことから、平均・二次・三次の積率を $n = 2000$ まで漸化式で正確に出す。平均と分散の式は次のとおり知られている。

$$E C_n = 2(n+1)H_n - 4n, \quad \text{Var } C_n = 7n^2 - 4(n+1)^2 H_n^{(2)} - 2(n+1)H_n + 13n$$

分散だけで、平均から t 以上大きくなる確率を上から押さえる（カンテリの不等式）。

$$P(C_n - E C_n \geq t) \leq \frac{\sigma^2}{\sigma^2 + t^2}$$

t を平均の 0.5 倍（平均の 1.5 倍を超える）と 1 倍（2 倍を超える）に置く。乱数の試行は $n = 100$ で 40000 回、 $n = 1000$ で 4000 回。

問い：乱数の軸で遅くなる確率は、どれだけ小さいと言い切れるか。その言い切りは実際とどれだけ離れているか。

2. 平均と分散

計算 2.1：漸化式で出した正確な値。

n	平均	分散	分散÷ n^2	標準偏差÷平均	歪み
10	24.437302	15.475117	0.154751	0.160977	1.050731
100	647.850259	3538.266678	0.353827	0.091816	0.959517
500	4806.409077	100179.209402	0.400717	0.065852	0.891185
1000	10985.912663	409117.768339	0.409118	0.058222	0.876740
2000	24729.829151	1656005.576805	0.414001	0.052037	0.867748

五つの n で、平均は式と相対 10^{-12} 、分散は式と相対 10^{-9} 以内で一致した。分散÷ n^2 の極限は $7 - 2\pi^2/3 = 0.420264$ である。

観察 2.2

ばらつきは n に比例し、平均に対する割合は縮む。分散÷ n^2 は 0.414001 まで上がって極限 0.420264 に近づき、標準偏差は n にほぼ比例する。平均は n より速く伸びるので、標準偏差÷平均は 0.160977 から 0.052037 へ縮んだ。歪みは正で、上に長い裾を持つ。

3. 遅くなる確率の上限と実際

計算 3.1：カンテリの不等式による上限。

n	平均の 1.5 倍を超える確率の上限	2 倍を超える確率の上限
10	9.391919×10^{-2}	2.525903×10^{-2}
100	3.262105×10^{-2}	8.359791×10^{-3}
500	1.705012×10^{-2}	4.317744×10^{-3}
1000	1.337787×10^{-2}	3.378363×10^{-3}
2000	1.071521×10^{-2}	2.700506×10^{-3}

計算 3.2：乱数の試行。

$n \cdot$ 回数	平均 (正確・標準誤差)	分散 (正確)	1.5 倍を超えた回数 (割合)	上限に対する比
100・40000	648.169 (647.850・0.298)	3561.287 (3538.267)	13 (0.000325)	0.009963
1000・4000	10989.062 (10985.913・10.031)	402510.179 (409117.768)	0 (0.000000)	0.000000

二つの n とも、試行の平均は正確な平均と標準誤差の 3 倍以内で合った。

観察 3.3

上限は言い切れるが、実際よりずっと甘い。分散だけで、平均の 1.5 倍を超える確率は $n = 2000$ で 0.010715 以下と言い切れ、その上限は n とともに下がった。しかし $n = 100$ で実際に超えた割合は 0.000325 で、上限の 0.009963 倍にすぎなかった。 $n = 1000$ では 4000 回で一度も超えなかったが、これは超えないことを意味しない。

4. 分離子

観察 4.1

分離子は、分散だけで押さえるか、分布の形まで見るかである。カンテリの不等式は、同じ平均と分散を持つあらゆる分布に当てはまる最悪の値を返す。クイックソートの比べる回数は上に長い裾を持つが ($n = 2000$ で歪み 0.867748)、その裾は最悪の分布よりずっと薄く、 $n = 100$ では上限の 0.009963 倍しか超えなかった。より強い上限は、分布の形を使って得られることが知られている。

観察 4.2

言わないこと。分布の形を使った強い上限の導き方、実装の速さ、 $n = 1000$ で超える確率の値は扱わない。示したのは、分散だけで遅くなる確率を正確に上から押さえられることと、その上限と実際の割合の隔たりである。

5. 点呼と結び

1. 比べる回数の漸化式とカンテリの不等式（第1節）：平均と分散から上限。
2. 平均と分散が式と一致（計算 2.1）：分散 $\div n^2$ は **0.414001**。
3. 標準偏差 $\div$ 平均 **0.052037**（計算 2.1）： n とともに縮む。
4. **1.5 倍**を超える確率 **0.010715** 以下（計算 3.1）：分散だけで言い切れる。
5. 実際は上限の **0.009963** 倍（計算 3.2）：上限はずっと甘い。
6. 分散だけか、分布の形までか（観察 4.1）：分離子・論文927 の射程注記。

結び：乱数の軸で遅くなる確率は分散だけで上から押さえられるが、その上限は実際よりずっと甘い。比べる回数の平均と分散を漸化式で正確に出すと、標準偏差 $\div$ 平均は $n = 2000$ で **0.052037** まで縮み、平均の **1.5 倍**を超える確率は **0.010715** 以下と言い切れた。 $n = 100$ で実際に超えた割合は **0.000325** で、上限の **0.009963** 倍だった。分離子は、分散だけで押さえるか、分布の形まで見るかである。既刊との位置：論文927 が試行の平均を並べるにとどめた乱数の運を、分散から上に押さえ、その押さえがどれだけ甘いかを数えた。正直に言えば、強い上限は導いていない。分散だけでどこまで言い切れ、実際とどれだけ離れているか、そこまでである。

参考文献

1. Hoare, C. A. R. (1962). Quicksort. *The Computer Journal*.
 2. McDiarmid, C. J. H., & Hayward, R. B. (1996). Large deviations for Quicksort. *Journal of Algorithms*.
 3. Rösler, U. (1991). A limit theorem for Quicksort. *RAIRO Theoretical Informatics and Applications*.
 4. 山岸勇輝 (2026). 『先頭を軸にするクイックソートは整った入力で最も遅く、軸を乱数で選ぶと平均の遅さは入力によらなくなる』[論文927]. YOI: <https://www.yamagishiyuuki.com/yoi/927>
-

— *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)