

Variance Alone Bounds the Chance That Random-Pivot Quicksort Runs Slow, but the Bound Is Far Looser Than Reality

— the spread of comparisons relative to the mean shrinks with n , and the bound on exceeding one and a half times the mean falls slowly, yet the share that actually exceeds it lies far below the bound —the separator is whether one bounds with the variance alone or looks at the shape of the distribution — [Paper 952]

Yuuki YAMAGISHI (山岸 勇輝)

Yamagishi Complexity Torus Lab, Japan

September 2026

YOI: <https://www.yamagishiyuuki.com/yoi/952>

— *From thinking to syncing.* —

Abstract

Quicksort with a randomly chosen pivot is fast on average, but can be slow with bad luck. This paper computes the mean, variance and third moment of the number of comparisons exactly by recursion, bounds from the variance alone the probability of exceeding one and a half or two times the mean, and compares these with the share that actually exceeds them in random trials. No new theorem or law is claimed. First, the mean and variance can be computed exactly by recursion and agree with the known formulas: at $n = 1000$ the mean was 10985.912663 and the variance 409117.768339; the mean matched its formula within a relative 10^{-12} and the variance within 10^{-9} . The variance divided by n squared was 0.414001 at $n = 2000$ (limit 0.420264) (Section 2). Second, the spread relative to the mean shrinks with n : standard deviation over mean was 0.091816 at $n = 100$, 0.058222 at 1000 and **0.052037** at 2000 (Section 2). Third, the variance alone bounds the chance of running slow: the probability of exceeding one and a half times the mean was at most 0.032621 at $n = 100$, 0.013378 at 1000 and **0.010715** at 2000, and of exceeding twice the mean at most 0.002701 at 2000 (Section 3). Fourth, the share that actually exceeds lies far below the bound: in 40000 trials at $n = 100$, 13 exceeded one and a half times the mean (0.000325), **0.009963** of the bound. At $n = 1000$, 0 of 4000 trials exceeded it (Section 3). Fifth, the separator is whether one bounds with the variance alone or looks at the shape of the distribution: a bound from the variance alone is the worst case over all distributions with that variance. The actual distribution has a long upper tail (skewness 0.867748), but that tail is far thinner than the variance-only bound (Section 4). **Keywords:** quicksort, number of comparisons, variance, Cantelli's inequality, tail probability, random pivot.

Scope of this paper (scope note): No new theorem or law is claimed: the mean $2(n+1)H_n - 4n$ and the variance formula for comparisons in random-pivot quicksort, the limit $7 - 2\pi^2/3$ of the variance divided by n squared, Cantelli's inequality bounding tails by the variance alone, and that stronger tail bounds are known are all established. Only comparisons are counted: memory access, constant factors and recursion depth are not treated (as in Paper 927). No strong bound is derived: exponential tail bounds that use the shape of the distribution are only named; here the variance-only bound is derived. Zero occurrences in trials are not taken as absence: that none of 4000 trials at $n = 1000$ exceeded one and a half times the mean does not mean it cannot happen. Relation to earlier papers: Paper 927, in its scope note, did not bound the luck of the random pivot from above and did not claim how small the probability of running slow is even with a random pivot, only listing trial means; this paper sits in that seat (no earlier occupant). Paper 927 showed that quicksort with the first element as pivot is slowest on sorted input and that choosing the pivot at random makes its average cost independent of the input. Here not only the mean but the probability of departing from it is bounded from above. What is added: computing the mean, variance and third moment exactly by recursion and checking them against formulas, setting out how standard deviation over mean shrinks, deriving bounds on exceeding one and a half and two times the mean, comparing the shares in random trials with the bounds, and placing the separator on "variance alone, or the shape of the distribution".

1. Notation and the Question

Setting: a list of length n is sorted by quicksort with a pivot chosen uniformly at random. With C_n the number of comparisons, when the pivot is the K -th element, $C_n = (n-1) + C_{K-1} + C'_{n-K}$ (the two independent). From this, the first three moments are computed exactly by recursion up to $n = 2000$. The mean and variance are known to be:

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

The variance alone bounds the probability of exceeding the mean by t or more (Cantelli's inequality).

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

t is set to 0.5 times the mean (exceeding one and a half times the mean) and 1 times the mean (exceeding twice).

Random trials: 40000 at $n = 100$ and 4000 at $n = 1000$.

Question: how small can the probability of running slow with a random pivot be asserted to be, and how far is that assertion from reality?

2. Mean and Variance

Computation 2.1: exact values from the recursion.

n	mean	variance	$\text{variance}/n^2$	sd/mean	skewness
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

For all five n , the mean matched its formula within a relative 10^{-12} and the variance within 10^{-9} . The limit of $\text{variance}/n^2$ is $7 - 2\pi^2/3 = 0.420264$.

Observation 2.2

the spread grows in proportion to n and shrinks relative to the mean. $\text{Variance}/n^2$ rose to 0.414001, approaching its limit 0.420264, so the standard deviation is nearly proportional to n : the mean grows faster than n , so standard deviation over mean shrank from 0.160977 to 0.052037. The skewness is positive, with a long upper tail.

3. Bounds and Reality for Running Slow

Computation 3.1: bounds from Cantelli's inequality.

n	bound on exceeding 1.5 times the mean	bound on exceeding 2 times
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}

Computation 3.2: random trials.

n , trials	mean (exact, standard error)	variance (exact)	exceeding 1.5 times (share)	ratio to bound
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

For both n the trial mean matched the exact mean within three standard errors.

Observation 3.3

the bound can be asserted, but it is far looser than reality. From the variance alone, the probability of exceeding one and a half times the mean can be asserted to be at most 0.010715 at $n = 2000$, and the bound fell as n grew: but at $n = 100$ the share that actually exceeded was 0.000325, only 0.009963 of the bound. At $n = 1000$ none of 4000 trials exceeded it, which does not mean it cannot happen.

4. The Separator

Observation 4.1

the separator is whether one bounds with the variance alone or looks at the shape of the distribution. Cantelli's inequality returns the worst case over all distributions with the same mean and variance: the number of comparisons in quicksort has a long upper tail (skewness 0.867748 at $n = 2000$), but that tail is far thinner than the worst-case distribution, exceeding only 0.009963 of the bound at $n = 100$. Stronger bounds are known to be obtained by using the shape of the distribution.

Observation 4.2

what is not said. How to derive strong bounds using the shape of the distribution, implementation speed, and the value of the probability of exceeding at $n = 1000$ are not treated. What is shown is that the variance alone gives an exact upper bound on the probability of running slow, and how far that bound lies from the actual share.

5. Roll Call and Closing

1. recursion for comparisons and Cantelli's inequality (Section 1): **bounds from mean and variance.**
2. mean and variance match the formulas (Computation 2.1): **variance/ n^2 is 0.414001.**
3. sd/mean **0.052037** (Computation 2.1): **shrinks with n .**
4. exceeding 1.5 times at most **0.010715** (Computation 3.1): **asserted from variance alone.**
5. actual share **0.009963** of the bound (Computation 3.2): **the bound is far looser.**
6. variance alone, or the shape of the distribution (Observation 4.1): **the separator; the scope note of Paper 927.**

Closing: Variance alone bounds the chance that random-pivot quicksort runs slow, but the bound is far looser than reality. Computing the mean and variance of comparisons exactly by recursion, standard deviation over mean shrank to 0.052037 at $n = 2000$, and the probability of exceeding one and a half times the mean could be asserted to be at most 0.010715: at $n = 100$ the share that actually exceeded was 0.000325, 0.009963 of the bound. The separator is whether one bounds with the variance alone or looks at the shape of the distribution. Placed among the earlier papers: the luck of the random pivot, which Paper 927 only listed as trial means, was bounded from above by the variance, and how loose that bound is was counted. To be honest: strong bounds are not derived. How far the variance alone lets one assert, and how far that is from reality, and no further.

References

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. Yamagishi, Y. (2026). *Quicksort with the First Element as Pivot Is Slowest on Sorted Input, and Choosing the Pivot at Random Makes Its Average Cost Independent of the Input* [Paper 927]. YOI: <https://www.yamagishiyuuki.com/yoi/927>
-

— *From thinking to syncing.* —

Yuuki YAMAGISHI (山岸 勇輝)

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