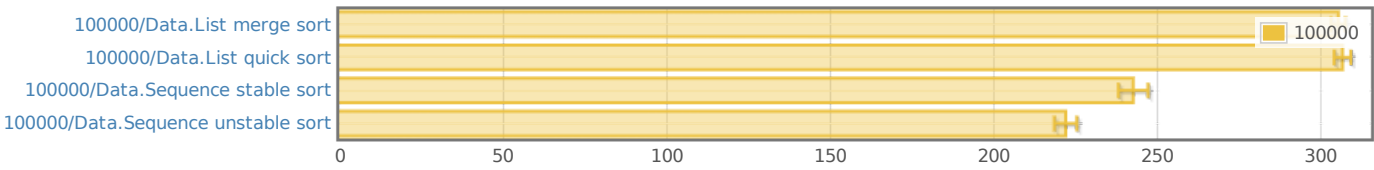# criterion performance measurements
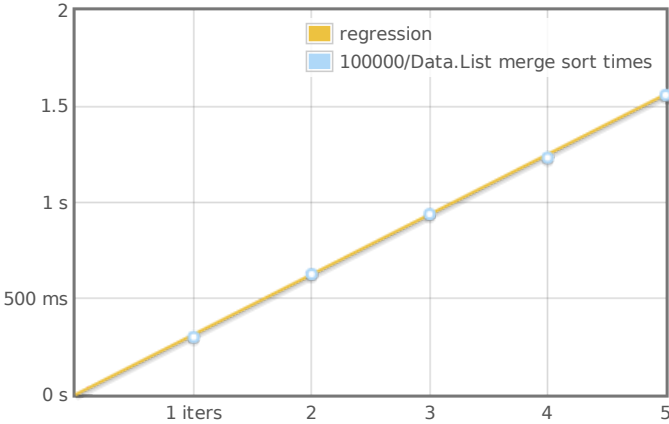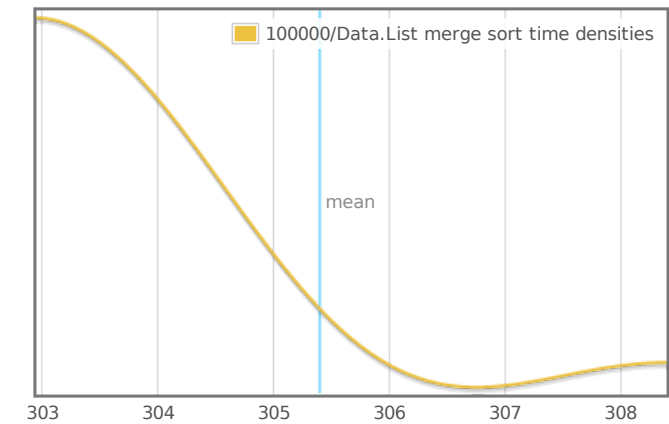
overview

want to understand this report?
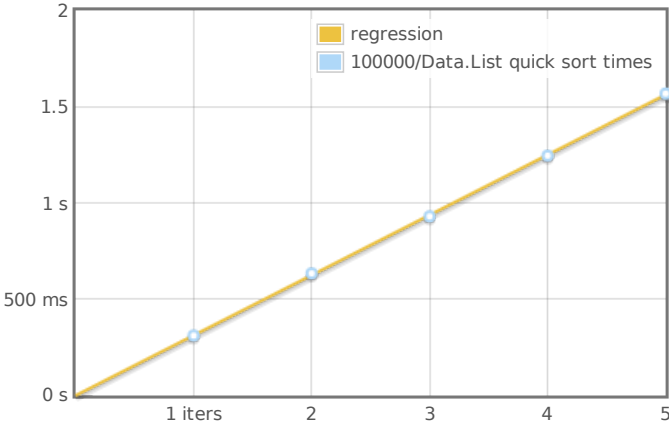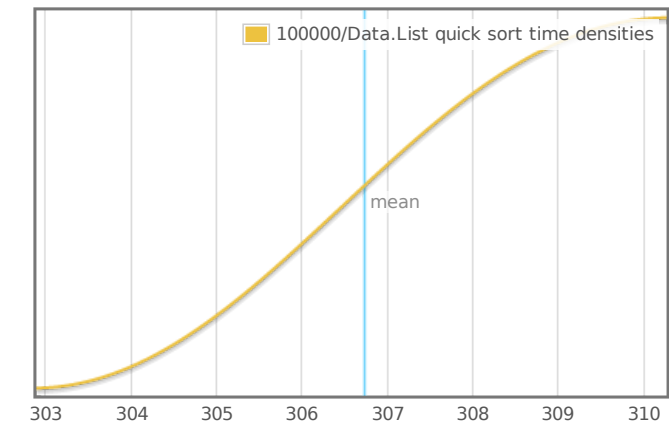

100000/Data.List merge sort

	lower bound	estimate	upper bound
OLS regression	302 ms	313 ms	321 ms
R² goodness-of-fit	0.999	1.000	1.000
Mean execution time	303 ms	305 ms	307 ms
Standard deviation	1.50 ms	2.22 ms	2.62 ms

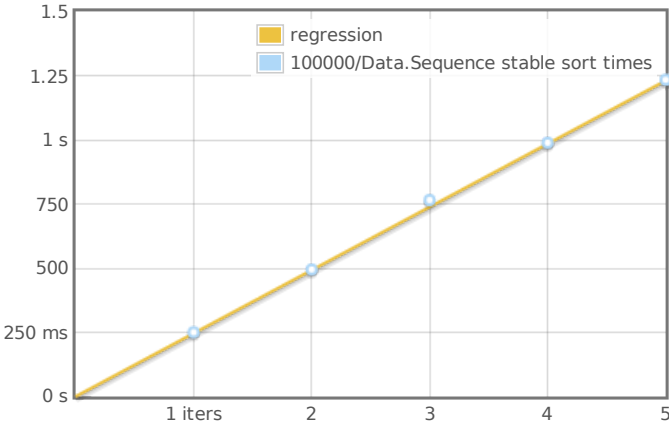Outlying measurements have moderate (16.0%) effect on estimated standard deviation.

100000/Data.List quick sort

	lower bound	estimate	upper bound
OLS regression	304 ms	312 ms	319 ms
R² goodness-of-fit	0.999	1.000	1.000
Mean execution time	304 ms	307 ms	309 ms
Standard deviation	1.33 ms	2.54 ms	3.54 ms

Outlying measurements have moderate (16.0%) effect on estimated standard deviation.

100000/Data.Sequence stable sort
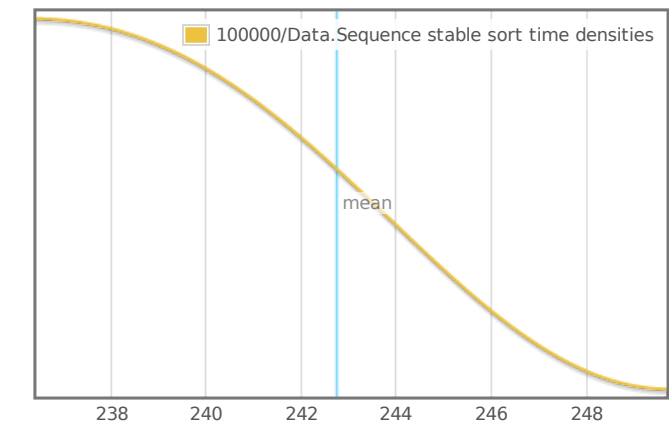
lower bound   estimate   upper bound

	lower bound	estimate	upper bound
OLS regression	234 ms	246 ms	258 ms
R² goodness-of-fit	0.998	0.999	1.000
Mean execution time	239 ms	243 ms	247 ms
Standard deviation	2.30 ms	4.53 ms	6.37 ms

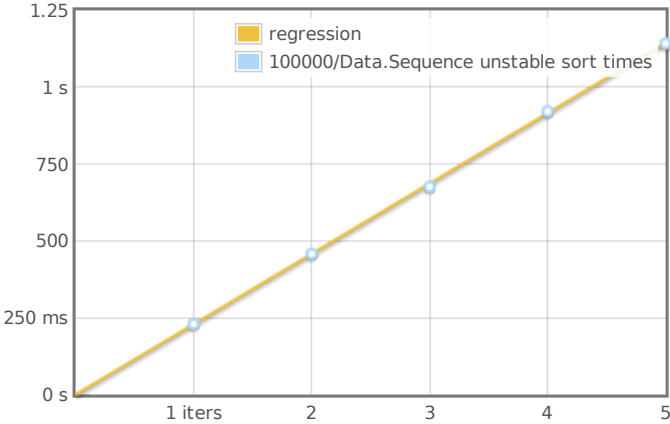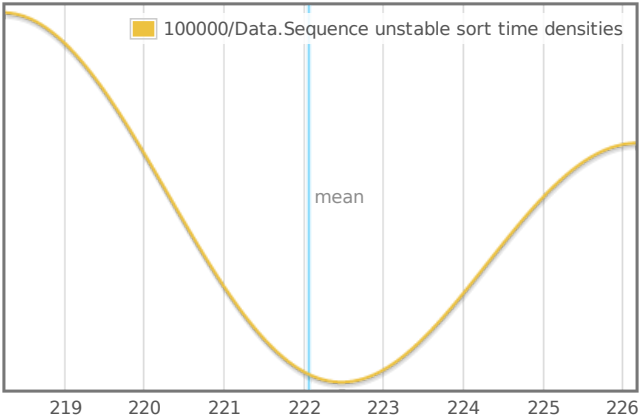Outlying measurements have moderate (16.0%) effect on estimated standard deviation.

100000/Data.Sequence unstable sort

	lower bound	estimate	upper bound
OLS regression	223 ms	229 ms	235 ms
R² goodness-of-fit	0.999	1.000	1.000
Mean execution time	219 ms	222 ms	225 ms
Standard deviation	2.84 ms	3.46 ms	3.80 ms

Outlying measurements have moderate (16.0%) effect on estimated standard deviation.

understanding this report

In this report, each function benchmarked by criterion is assigned a section of its own. The charts in each section are active; if you hover your mouse over data points and annotations, you will see more details.

- The chart on the left is a kernel density estimate (also known as a KDE) of time measurements. This graphs the probability of any given time measurement occurring. A spike indicates that a measurement of a particular time occurred; its height indicates how often that measurement was repeated.
- The chart on the right is the raw data from which the kernel density estimate is built. The x axis indicates the number of loop iterations, while the y axis shows measured execution time for the given number of loop iterations. The line behind the values is the linear regression prediction of execution time for a given number of iterations. Ideally, all measurements will be on (or very near) this line.

Under the charts is a small table. The first two rows are the results of a linear regression run on the measurements displayed in the right-hand chart.

- *OLS regression* indicates the time estimated for a single loop iteration using an ordinary least-squares regression model. This number is more accurate than the *mean* estimate below it, as it more effectively eliminates measurement overhead and other constant factors.
- *R² goodness-of-fit* is a measure of how accurately the linear regression model fits the observed measurements. If the measurements are not too noisy, R² should lie between 0.99 and 1, indicating an excellent fit. If the number is below 0.99, something is confounding the accuracy of the linear model.
- *Mean execution time* and *standard deviation* are statistics calculated from execution time divided by number of iterations.

We use a statistical technique called the bootstrap to provide confidence intervals on our estimates. The bootstrap-derived upper and lower bounds on estimates let you see how accurate we believe those estimates to be. (Hover the mouse over the table headers to see the confidence levels.)

A noisy benchmarking environment can cause some or many measurements to fall far from the mean. These outlying measurements can have a significant inflationary effect on the estimate of the standard deviation. We calculate and display an estimate of the extent to which the standard deviation has been inflated by outliers.