

The Log Weighted Average for Measuring Printer Throughput

The log weighted average balances the different time scales of various plots in a test suite. It prevents an overemphasis on plots that take a long time to print and allows adjustments according to the expected user profile weighting. It is based on percentage changes rather than absolute plot times.

by John J. Cassidy, Jr.

The HP DeskJet 1600C printer is designed to be used for a variety of documents, from simple memos to complex color graphics. One of the main characteristics on which the printer will be judged is throughput. We needed a way to measure throughput across a wide range of plots that would reflect a user's subjective perception of the product.

The two most common metrics—simple average and simple weighted average—had serious problems when applied to the disparate plots in our test suite. A simple and common mathematical technique was used to overcome these problems, resulting in a metric called the *log weighted average*.

This paper explains how to calculate the log weighted average, and why it is a good metric.

The Problem

We use a standard set of plots to measure the speed of the HP DeskJet 1600C printer. For the sake of this paper, I simplify the test suite down to four plots—we actually use 15. The actual timings have also been simplified and are not accurate for any version of the printer under development. The four plots are (1) text page, a normal letter or memo, (2) business graphic, some text with an embedded multi-color bar chart, (3) spreadsheet with color highlighting of some of the numbers, and (4) scanned image, a complex, full-page, 24-bit color picture.

For a given version of the HP DeskJet 1600C printer, call it version 3.0, let's say the time to process and print each of these pages is as follows:

Text page	10 seconds
Business graphic	20 seconds
Spreadsheet	45 seconds
Scanned image	10 minutes (600 seconds)

There are various things we can do to the printer to change the speed of each of these plots. Often a change will speed up one plot while slowing down another. What we need to do is compare alternative possible version 3.1s and see which one is faster overall.

Simple Average

The simple average is calculated by adding up the time for each of the plots and dividing by the number of plots. The formula for this is:

$$\text{Simple Average} = \sum T_i / n,$$

where n is the number of plots and T_i is the time to process plot number i .

For version 3.0 above, the sum of the four times is 675 seconds which divided by four gives a simple average of 169 seconds (rounding from 168.75).

The problem with the simple average is that it gives equal importance to each of the seconds spent on each of the plots. If a version 3.1a saved five seconds on the scanned image, this plot would go down from 600 seconds to 595 seconds and the user would barely notice. But if a version 3.1b saved 5 seconds from the text plot, this plot would go from 10 seconds to 5 seconds, twice as fast! The user would be very, very happy with the text speed.

The simple average tells me that these two changes are of equal value. So if I am using this metric, I'll go for the easy change of speeding up the scanned image by a little bit (less than 1% faster), instead of the much more difficult and more useful speedup of the text page (50% faster).

Simple Weighted Average

A common way to improve the simple average is to make use of the fact that we know how often the user is going to print each type of plot (at least we make good guesses). We know, for example, that someone in our target market will print a lot more simple text pages than complex scanned graphic pages.

The simple weighted average applies a weight to each of the plots, corresponding to the proportion of time the user will be printing that type of plot. In mathematical terms:

$$\text{Simple Weighted Average} = \frac{\sum (T_i W_i)}{\sum W_i}$$

where W_i is the weight for plot i . If the W_i add up to 1.0, the denominator can be ignored.

For the HP DeskJet 1600C printer, let's say half of the plots will be like the text page, one-fifth like the business graphic, one-fifth like the spreadsheet, and one-tenth like the scanned image. This gives the following calculation:

Plot	Time (s)	Weight	$T_i W_i$ (s)
Text Page	10	0.5	5
Business Graphic	20	0.2	4
Spreadsheet	45	0.2	9
Scanned Image	600	0.1	60
Sum		1.0	78

The simple weighted average is 78 seconds.

This method of calculation is much better than the simple average. It takes into account our knowledge of the target market, and any average we come up with needs to be able to do this.

But there are still problems with this average. Say that version 3.1a speeds up the scanned image by 5% (down to 570 seconds), and version 3.1b speeds up the text page by 50% (down to 5 seconds).

We know from our own experience that speeding something up from 10 minutes to 9.5 minutes is not very significant. On the other hand, the 3.1b version, which makes the most frequent task go twice as fast, would represent a very noticeable improvement. However, the simple weighted average rates the two versions very similarly, with the 3.1a winning (at 75 s) over the 3.1b version (at 75.5 s).

Our subjective experience of time is such that we tend to notice changes not in absolute seconds, but in percentages of time. A one-percent speedup of any of the categories would be impossible to detect without a stopwatch, but a twenty-five percent speedup would be dramatic for any plot.

Criteria for a Good Average

A good averaging technique would have the following characteristics:

- It is based on percentage changes. For a short task, a small speedup is significant. For a long task like the scanned image, it takes a big speedup to make a difference. A good average would not focus on how many seconds were saved, but on what percentage of the task was saved.
- It reflects user profile weighting. For the HP DeskJet 1600C printer we need to emphasize text speed, since that is the center of our market. But for another printer aimed at another market, the spreadsheet or the scanned image might be most important. The average has to allow tailoring.
- It is invariant under a many-for-one substitution. If instead of one text page weighted at 0.5, we substituted five text pages each weighted at 0.1 into the calculation (to avoid dependence on the quirks of a single document), and if each of the five text pages took the same time as the original one (10 s) to print, the average should not change.

Log Weighted Average

The log weighted average fulfills the above criteria. Its general principle is to use a standard mathematical technique (logarithms) for keeping large and small numbers on the same scale.

The formula for the log weighted average is:

$$\text{Log Weighted Average} = \exp\left(\frac{\sum (\ln T_i) W_i}{\sum W_i}\right),$$

where \ln is the natural logarithm (log to the base e), and \exp is the exponent function, e to the x . As before, if the sum of the weights is 1.0,

$$\text{Log Weighted Average} = \exp\left(\sum (\ln T_i) W_i\right).$$

For our example, the calculation would be:

Plot	T_i (s)	$\ln T_i$	Weight	$(\ln T_i) W_i$
Text Page	10	2.30	0.5	1.15
Business Graphic	20	3.00	0.2	0.60
Spreadsheet	45	3.81	0.2	0.76
Scanned Image	600	6.40	0.1	0.64
Sum				3.15

$$\text{Log Weighted Average} = e^{3.15} = 23.4 \text{ s}$$

One of the first things you notice about the log weighted average (aside from the fact that it took an extra step to do the calculation) is that the result of 23 seconds is shorter than the results of the other two calculations. The simple average gave 169 seconds, and the simple weighted average gave 78 seconds. This is because the more sophisticated averages do a progressively better job of moderating the influence of the very long 10-minute scanned image plot. Also, this example was artificially constructed with a wide variation in plot times. Often we deal with plots that are more similar than these. If the plots were very similar and every plot in the test suite had exactly the same timing, say 30 seconds, then it wouldn't matter which method you used. All three methods would give the same average: 30 seconds.

Rule of Thumb

The biggest drawback of the log weighted average is that it is less intuitive than the other two methods. There is something basically counterintuitive about using logarithms if you aren't a professional mathematician. They tend to throw off our mental approximations of what is reasonable.

However, there is a relatively simple rule of thumb to help us know what to expect when doing comparisons: *A small percentage change in one component is equivalent to the same percentage change in another component, multiplied by the ratio between their weights.*

In our example, this means that a small change in the text page (with a weight of 0.5) would be five times as important as a change in the scanned image (with a weight of 0.1), and two and a half times as important as a change in the spreadsheet or business graphic (with a weight of 0.2). Thus, we would expect a 1% change in the text page to be equivalent

to a 5% change in the scanned image or a 2.5% change in the other two plots.

This approximation is very close. A 1% speedup in the text page, from 10 s to 9.9 s, reduces the overall log weighted average from 23.4 to 23.3 seconds. The equivalent change required for one of the other plots to get the average down to 23.3 is shown in Table I.

Table I
Equivalent Speedups (Small Deltas)

Text page	10 s → 9.9 s = 1.00% faster
Business Graphic	20 s → 19.5 s = 2.48% faster
Spreadsheet	45 s → 43.9 s = 2.48% faster
Scanned image	600 s → 571 s = 4.90% faster

As changes get bigger, the rule of thumb becomes less accurate. If you make a big change in one of the components, like speeding up the scanned image by 40%, you stray farther from the expected equivalent speedups of 20% (half as much) for the spreadsheet and business graphic, or 8% (one fifth as much) for the text page. This change brings the log weighted average down to 22.2 seconds. Table II shows the equivalent speedups for larger changes.

Table II
Equivalent Speedups (Larger Deltas)

Text page	10 s → 9.03 s = 9.7% faster
Business Graphic	20 s → 15.5 s = 22.5% faster
Spreadsheet	45 s → 34.9 s = 22.5% faster
Scanned image	600 s → 360 s = 40.0% faster

The Exact Rule

Exact calculation of equivalent speedups for any situation using the log weighted average can be done using the following rule: *Multiplying the time for component A by a factor r is equivalent to multiplying component B by r raised to the power W_A/W_B , the ratio of the weights of the two components.*

For example, if we multiply the text page time by 1.2 (slowing it down by two seconds), that would raise the log weighted average from 23.4 seconds to 25.6 seconds. To get an equivalent change by altering the scanned image time, we would have to multiply it by 1.2 to the fifth power (the ratio of the text page weight to the scanned image weight is five), or $600 \times 1.2^5 = 1493$. Thus, by changing the scanned image time to 1493 seconds, we could also raise the average from 23.4 to 25.6 seconds.

For very large changes in any of the components, the log weighted average gives results that can conflict with intuition. For example, speeding up the text page from ten seconds to one second would improve the average dramatically. Such a speedup is wildly improbable for the HP DeskJet 1600C printer, but can be anticipated for some comparable printer to be developed in our lifetime.

To get an equivalent improvement in the average by only changing the scanned image, we would have to print it in $600 \times 0.1^5 = 0.006$ second (which probably violates some laws of physics). You can use the exact rule to verify that the same sort of numerical blowup results when you try to compare any two printers that are greatly dissimilar. This is not a particular problem for us. Greatly dissimilar printers also have dissimilar weighting profiles, and we don't know any way to compare them well.

Usefulness of the Log Weighted Average

The log weighted average is designed around a user's subjective perception of printer speed. It assumes the common situation in which a user is working at a computer, sends something to the printer, and somehow notices how long it takes to come out. There is also an assumption that if something takes twice as long, the user is unhappy and if something takes half as long, the user is happy, and the unhappiness in the first situation is roughly equivalent in intensity to the happiness in the second situation.

There are some situations for which this isn't true and the log weighted average is the wrong average to use. For example, you could have a printer in continuous use with no stopping except to add paper and change pens. This might be at a real estate office producing a large number of personalized letters and envelopes each day and a smaller number of scanned house photos. For a customer like this, the subjective perception of speed is not important. Two seconds saved on a text page is no more important than two seconds saved on a scanned image. The simple weighted average would be the correct average to use here.

Our success with this technique resulted from regular application. On the HP DeskJet 1600C project, we timed the 15-plot test suite twice a month. This helped us quickly identify and resolve issues that might otherwise have caused problems.

Conclusion

The log weighted average does a good job of balancing the different time scales of various plots in a test suite. It prevents an overemphasis on plots that take a long time to print, and allows adjustments according to the expected user profile.

The main cost of the log weighted average is that it is less intuitive than other methods. The rule of thumb and the exact rule are good guides as to how the average will react.

The log weighted average has limits, but for comparing two reasonably similar printers in a normal home or office environment, it gives extremely helpful results.

Acknowledgment

Thanks to Jeff Best of the San Diego Printer Division for his comments and discussion.