RGR MS

Chris H. Wilson

March 2, 2019

### Background and Rationale

Analyzing growth of individuals is fundamental in many areas of ecology and biology. A common situation is the need to intercompare multiple individuals across genotypes or species in experimental or observational settings where variations in initial sizes and environmental factors both contribute to observed variations in size at any point in time. In this setting, a common default practice is to re-express growth as a relative measure, dividing the growth increment by the initial size. In the limit as the time period goes to zero, this can be represented as

Without explicit specification of a time-varying dynamic, e.g. some kind of non-linear growth function, this representation corresponds to exponential growth. That is, the quantity obtained by integration of

over some time period, and given some initial size

is simply the familiar exponential equation

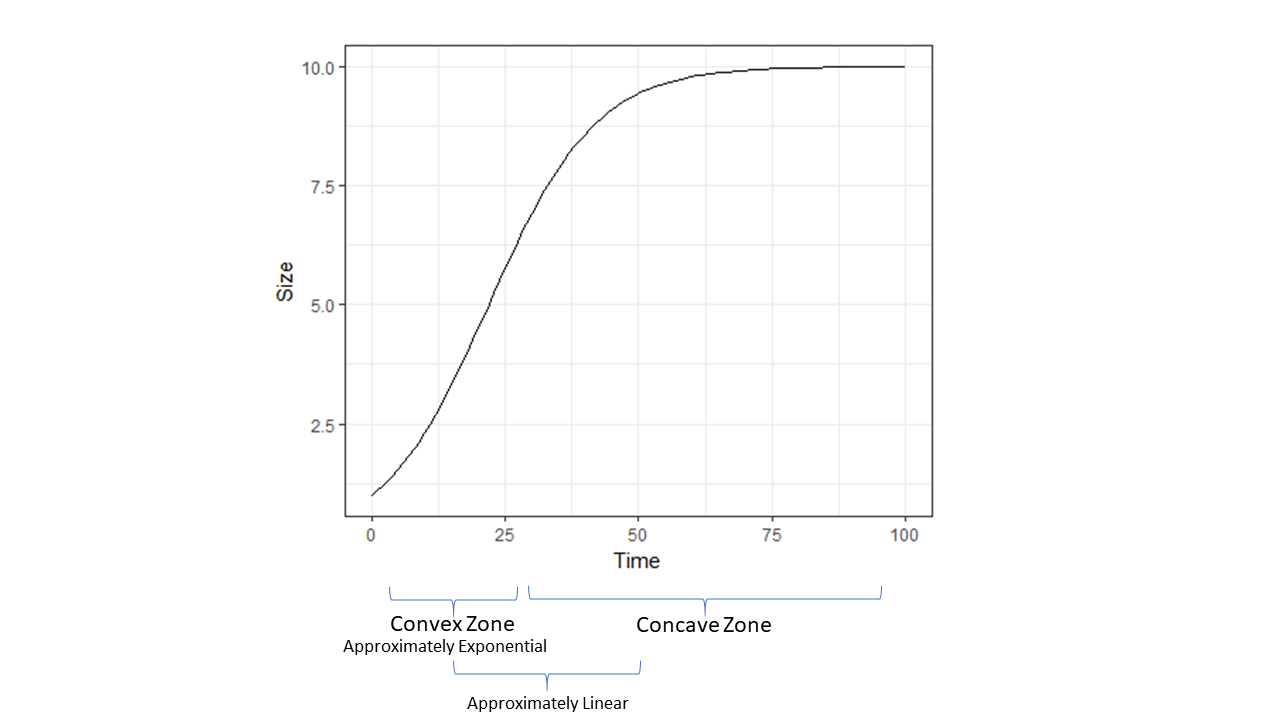
The quantity in equation 1 is often referred to as relative growth rate (RGR), and the usual method of quantification, hereafter the “log measure” corresponds to the solution in 2, as is readily checked. The log measure is, simply

The log measure is very frequently utilized as a default in place of taking the simply difference . XXX et al. (2012) summarized several flaws of the log measure and recommended instead to fit non-linear growth functions. I wholeheartedly concur with this advice. However, ecologists are often confronted with datasets where only 2 or 3 time periods are available, thus precluding effective fitting of non-linear functions.

In this note, I demonstrate that the log measure should also not be used in the data limited setting of only 2 or 3 observation times. Instead, the linear difference is to be preferred for three reasons: 1) simplicity of interpretation, 2) ecological/biological validity, and 3) ease of use.

### Conceptual Overview

First, we assume a basic theoretical framework for growth: the sigmoidal curve. Nearly every biologically motivated growth model follows sigmoidal behavior. For instance, West et al. (2001) famously derived a sigmoidal equation for growth from metabolic scaling theory. Although the universality of their particular model has been challenged, saturating non-linear growth has not. Therefore, we will compare the linear difference and the log measure to a sigmoidal curve, which is itself presumed to better approximate underlying biological/ecological reality. Before proceeding to a brief mathematical exposition, let’s consider a heuristic/intuitive argument. The following figure encapsulates my entire argument:

 In the convex zone, the usual log measure is a fine approximation. However, the zone of approximate linearity is just as large (if not larger), where the linear difference is to be preferred. Finally, neither approximation is great in the upper portion of the concave zone, although as demonstrated below, the linear approximation is uniformly superior there.

Mathematically, the argument can be boiled down for any generic equation for growth over time: . Using Taylor Series, we can approximate around some value to second order with a generic function as:

As noted above, the canonical log measure corresponds to exponential dynamics . Use of exponential dynamics to approximate obviously only works well where both the first and second derivative of are positive (i.e. where function is convex). Given that the second derivative of any sigmoidal curve flips from positive to negative, this approximation error grows rapidly outside of a narrow zone.

### Mathematical Analysis of Log Measure

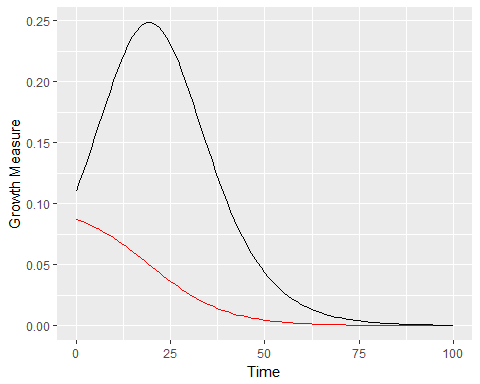
We take the familiar logistic equation as a reasonable representaton for sigmoidal growth, while noting that many options are available:

Given this representation of underlying growth, the first question is: what does the canonical log measure correspond to? In other words, we want to ask what happens given exact measurements of and , given that they are sampled from above equation. For We have:

If we have size observations and from two times, and ,the difference between them is:

For any given interval :

One flaw of the canonical log measure (as pointed out previously by XXXX) is that it is really time-varying, but in effect treated as though time constant (by necessity given the limitation of data). If we want to investigate how this quantity varies with sampling of arbitrary timepoints and along the sigmoidal curve, we re-express and let

Graphically, the comparison with the observed growth increments is: 

This comparison, while visually striking, is fundamentally misleading since the log measure does not really have the same dimensions as the observed growth increment. It is really a **rate** with dimensions of , and the usual re-expression as is unhelpful at best. Thus, the only way to compare use of the canonical log measure to the simple linear measure is to reformulate the problem in terms of implied dynamics.

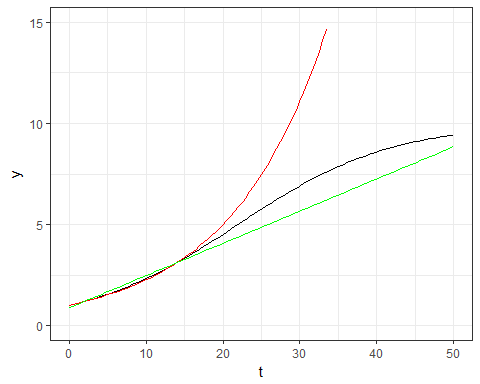
### Comparing log measure and linear measure in terms of dynamics

Use of the linear difference corresponds to assumption of a static linear growth rate dynamic, just as use of corresponds to assuming a static exponential growth rate dynamic. In the latter case, the log-measure has the nice property of representing an ergodic observable (sense Peters and Gell-man 2016). While widely (and rightly) dismissed as unrealistic, the linear dynamic may in fact be a generally superior measure for ecological analysis where no time series of size/biomass data is available.

The comparison made here is the goodness of fit implied by replacing the sigmoidal with either a linear approximation or an exponential approximation, given sampling of size from two pairs of time points: 1) from the early (“exponential”) portion of sigmoid curve, and 2) from the middle (“linear”) portion of sigmoidal curve, and 3) from the saturating part of curve.

1. and :

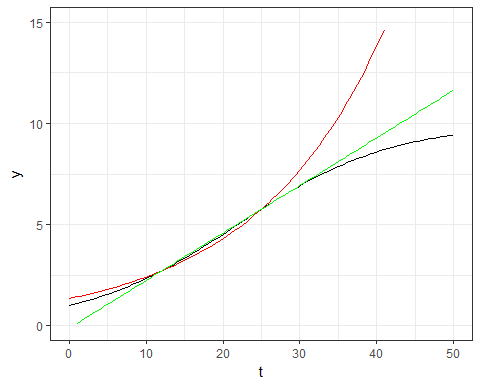
## Warning: Removed 33 rows containing missing values (geom\_path).

 As expected, the exponential approximation works better with data from the convex portion of growth curve. However, the improvement is marginal in absolute value, and quickly diverges outside of the convex portion (in accord with our intuitive model).

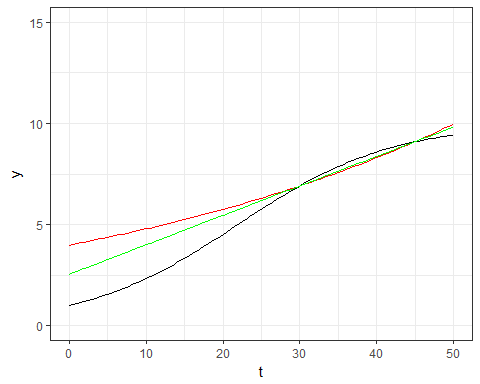
1. and :

## Warning: Removed 18 rows containing missing values (geom\_path).

## Warning: Removed 2 rows containing missing values (geom\_path).

 As can be seen, the linear model is a better approximation where data are taken from within the center part of the growth cycle. Again, the improvement is marginal, but real. Forecast accuracy is much higher, and backcast accuracy marginally worse.

1. and

 As expected, in this scenario, the linear approximation is uniformly better and thus always to be preferred.

### Statistical Properties

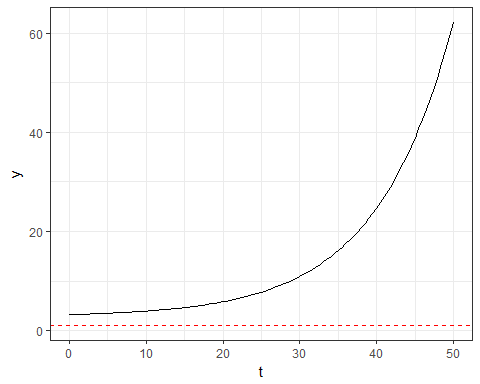
We can derive the sampling distribution of the log measure based on a Taylor Series’ approximation. Specifically, we consider measurements and with normally distributed error, where the variance scales with the mean (a fairly typical property in biological/ecological data). The distribution of - is then simply the difference of two Normals. Next, we approximate the moments of the distribution of as:

. Using the delta method for variance, we have:

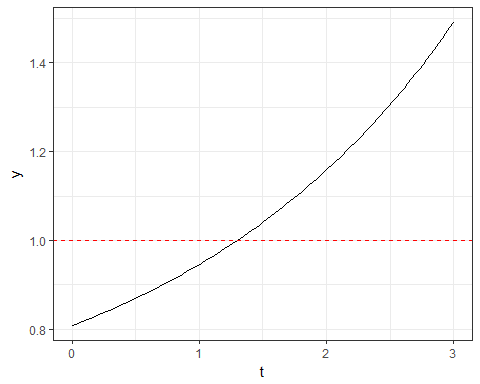
Given a constant coefficient of variation (reflecting variance scaling with mean on the original scale) the sampling distribution of (setting to unit scale), is therefore:

The CV of the log measure is then related to the expectaiton of the Z scores of the new sampling distribution, and is inversely proportional to statistical power:

Thus, where over a unit time increment, the log measure should have greater statistical power, while it loses statistical power as the log measure declines (which as we saw in section XXX above occurs faster than the linear difference of course). This can be reformulated as , revealing a *scale free* property of statistical analysis of the log measure. Specifically, wherever the multiplicative growth increase on a unit time scale , the log measure has worse statistical properties than the linear difference.

For our previous parameter value simulations above, here is the curve of with time , expressed in multiples of : 

As can be seen, in this situation, it is always worse! What happens if we accelerate growth rate considerably (5X)



There is a small window in early growth where it is expected to have better properties for estimating observed differences, but it rapidly loses power.

### Case Study: A Situation Statistical Analysis of Log Measure and Linear Measure Diverge

Maybe not necessary here. Theoretical demonstration feels complete.

### Conclusions and Recommendations

In summary, the chief virtue of the measure is that it effectively linearizes the differences in size from the convex portion of biological growth curve. Thus, it arguably might increase the ability to infer subtle but consequential differences in growth rates based on small differences in observed sizes. However, analysis of the statistical sampling distribution suggests that inference on this quantity is actually more difficult than inference on the linear measure. In fact, ability to detect differences rapidly declines with position on the sigmoidal curve.

The much maligned linear measure is a superior default on three grounds therefore. First, it corresponds far more directly with current ecological reality. It is a measure with an interpretable biological dimension (usually mass or length, whereas the log measure has dimensions of ). This enables us to more literally describe our system. Second, interpreted in terms of dynamics, it is *less wrong* than the exponential approximation implied by the log measure. Third, it has superior statistical properties almost everywhere.

The widespread use of the log measure of “RGR” as an *a priori* preferred default for the data-limited situation should be abandoned. Where only two or three time points are available, fitting a linear growth trend is just as good if not better than estimating an exponential growth rate. The ideal scenario is to collect a proper time series (5-7+) and fit a proper growth model. Where data are at all limiting, we recommend careful incorporation of literature values and other external information as priors in a fully Bayesian analysis in order to regularize inferences.