

SARS-CoV-2 Growth Rates

Why growth rates for outbreak.info?

- In pandemic tracking, a huge volume of data needs to be sorted through
- One of the most important questions users need to answer is
 - “What data should I be looking at?”
 - Fast-growing lineages are more likely to become variants of concern so a high growth rate is a good heuristic for importance
- Growth rates provide a point of direct comparison across our divergent data set
- And growth rates also provide a clear view into the dynamics of the pandemic

Goals

Provide variant-specific growth rates estimates and UI elements for outbreak.info that:

- provide useful insight
- are directly interpretable
- are statistically meaningful
- are robust to extremely various data conditions
- are automated and efficiently scalable

Uncertainty
Estimation

Minimum model
approach

Data Sources

- Sequence counts by variant (main data source)
- Case counts, death counts, population (secondary data sources)

Principles

- Operate effectively on as little data as possible
- Aggregate late – aggregating early risks losing information
- Return a value if at all possible, even a very uncertain one
- Estimates are probability distributions; uncertainty is just as important as data
- We can use uncertainty propagation to translate information about uncertainty

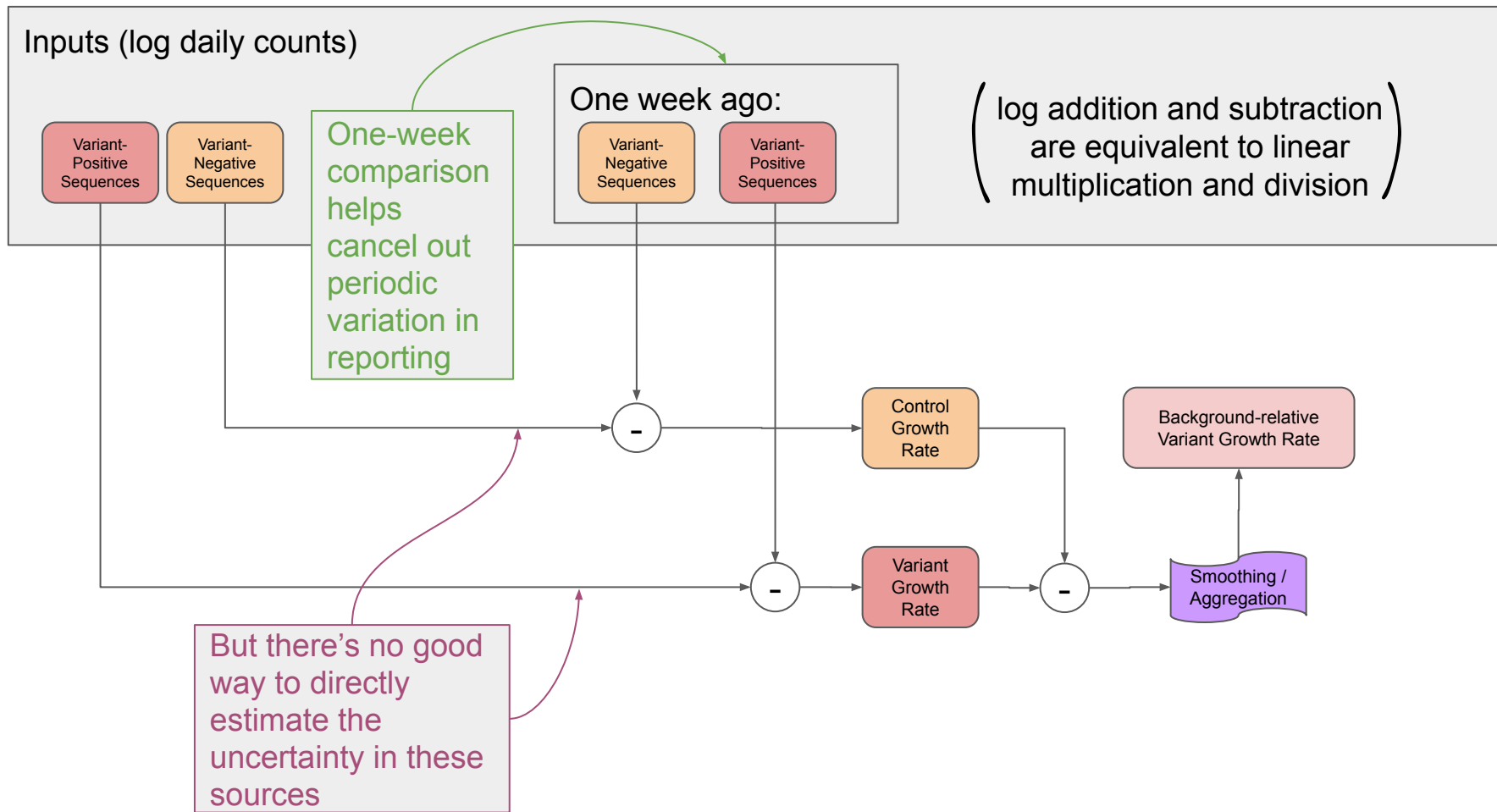
$$c = a + b$$

$$(\sigma c)^2 = (\sigma a)^2 + (\sigma b)^2 + 2ab (\sigma ab)$$

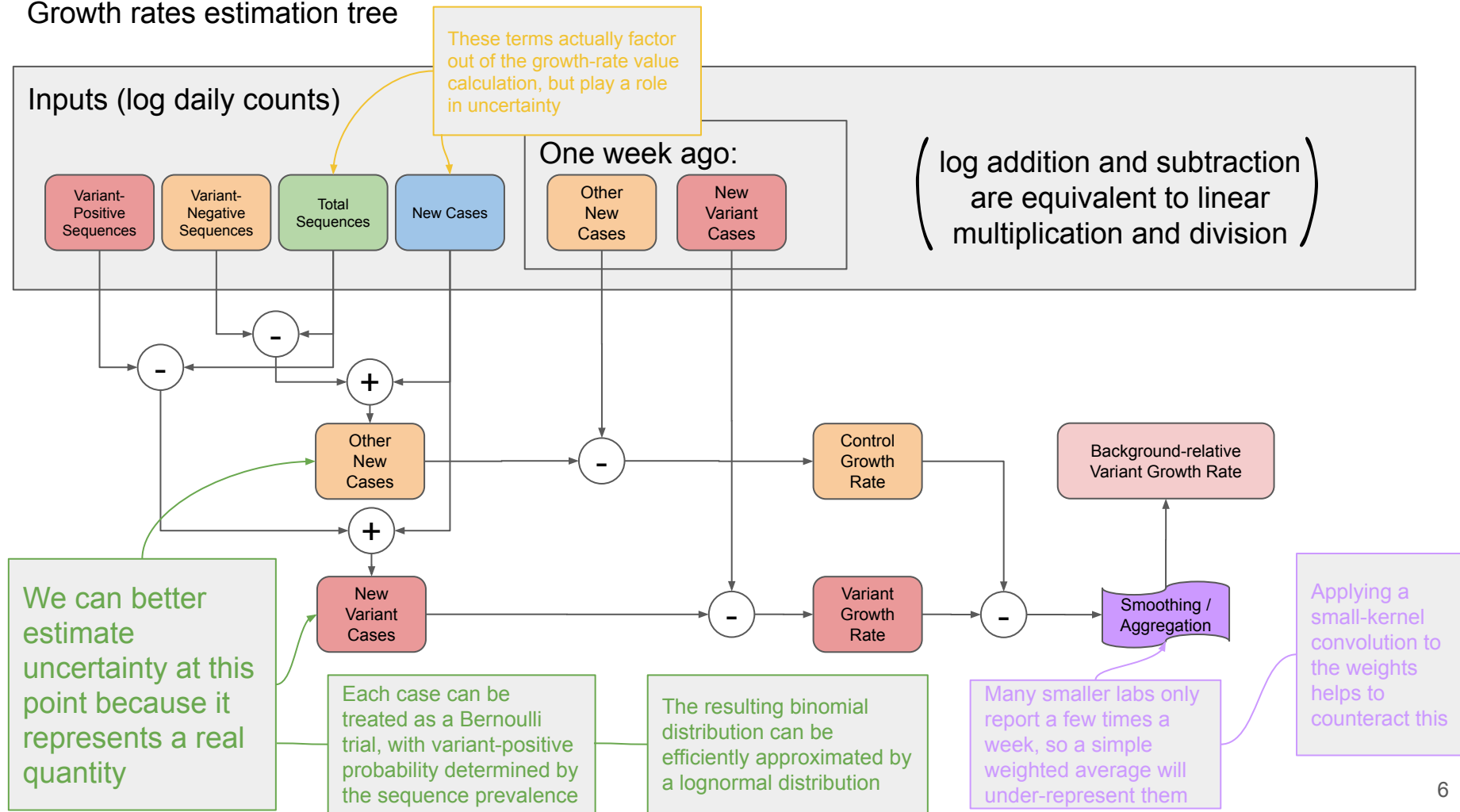


Difficult to estimate
in most cases

Growth rates estimation tree

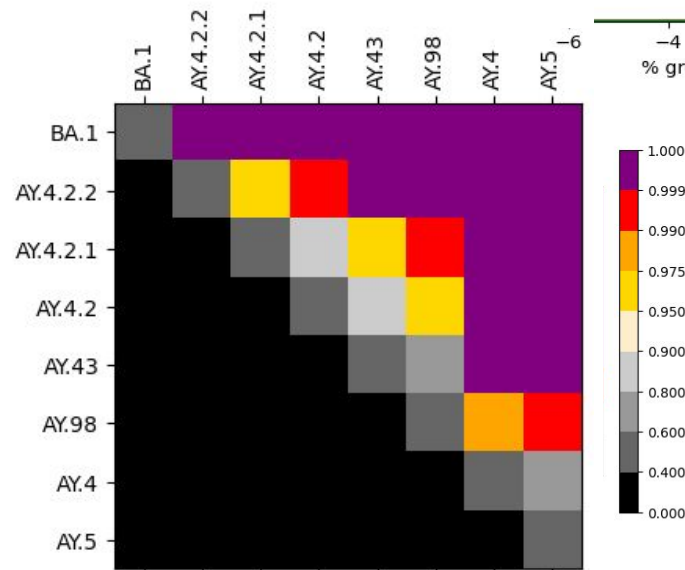
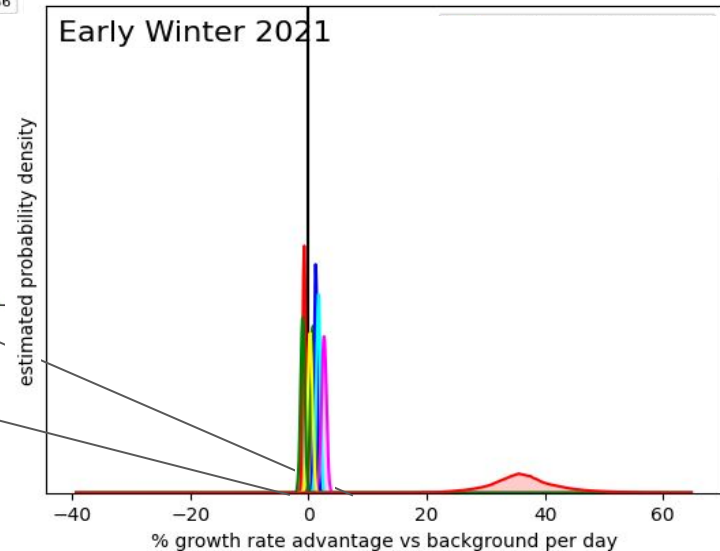
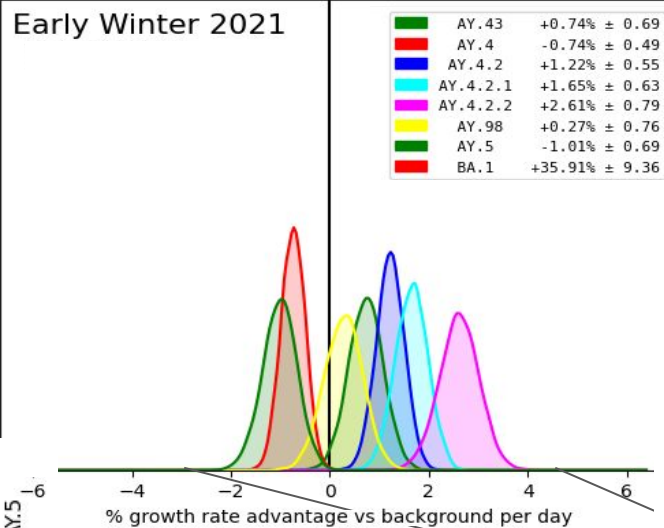


Growth rates estimation tree



The result is a lognormal probability distribution for each variant / location / time-window, representing our estimated knowledge about that variant's growth rate relative to all other strains circulating in the same population

estimated probability density



By taking the sliding product of two of these PDFs (a cross correlation), we can derive a new PDF describing our belief over the difference, allowing us to compare variants directly.

We can use this to calculate R_0 values by comparing to a known lineage and raising to the exponent of the generation time

If a variant has a growth-rate advantage of 33%, then each case will cause about one third more cases in a single day than a typical non-variant strain