Counting-based inference of mutant growth rates from pooled sequencing across growth regimes

Deniz Sezer*[1] and Erdal Toprak[1]

[1]Department of Pharmacology, UTSW Medical Center, TX 75390, USA

October 14, 2025

Abstract

Simultaneously inferring the growth rates of thousands of genetically distinct cellular variants coexisting in a single culture is now possible by tracking their relative abundances over time through deep sequencing, or other counting methods. This capability provides a quantitative window into increasingly broad regions of the vast combinatorial fitness landscape—well beyond those explored by natural evolution. While identifying a few engineered variants that survive under extreme selection is valuable, quantitative mapping of the fitness landscape requires accurate inference of growth rates for the entire variant population. Here, we revisit three established approaches for estimating growth rates from sequencing count data: least-squares fitting, maximum likelihood estimation, and variational Bayesian inference. In contrast to previous treatments, we introduce the *softmax* transformation to link the probabilistic model of counting noise to the deterministic model of variant growth. By clearly delineating these two modeling components—jointly required for inferring growth rates from time-resolved sequencing counts—we show how to replace exponential growth with any alternative model defined by a small number of variant-specific parameters. While we illustrate this flexibility using the logistic and Gompertz models of growth, application to more mechanistic models, in which growth rates are expressed in terms of microscopic biochemical parameters, will open the door to high-throughput inference of fundamental kinetic and biophysical constants from sequencing data.

1 Introduction

Recent advances in sequencing and gene synthesis technologies have enabled the simultaneous interrogation of thousands of genetically distinct pooled variants growing under nearly identical conditions within the same culture. Depending on the experimental design, these variants may represent distinct gene knockouts [1] or knockdowns [2] of a parental genome, or they may differ in the sequence of one [3;4], two [5], or multiple genes whose products influence growth under the studied conditions.

On a qualitative level, screening thousands of variants in parallel increases the chance of identifying a few exceptionally fit mutants among a combinatorially vast space of possible genotypes. Quantitatively, estimating the fitness of each variant from changes in its relative abundance over time—as measured through sequencing-based counting—provides a systematic framework for mapping epistatic interactions between mutations. [4] Elucidating the mechanistic basis of these interactions, in turn, offers deeper insights into the underlying biological processes.

This work focuses on the quantitative estimation of variant growth rates from sequencing-based counts. Specifically, we consider growth *competition* experiments in which a diverse population of pre-existing variants is propagated in parallel for several generations—a time scale short enough that the spontaneous occurrence of new mutations can be neglected. This setup contrasts with *evolution* experiments, which typically span hundreds of generations and are designed to study the emergence and fixation of novel genetic changes. [6]

*dzsezer@gmail.com

Estimating growth rates requires sequencing at least two time points along the competition trajectory, and many studies sequence only at the beginning and end of the experiment.[3] With two points only, assuming exponential growth between them, the growth rate can be inferred by fitting a line through two log-transformed abundance values. When sequencing data are collected at more than two time points, the question arises of how best to integrate the additional temporal information. One intuitive approach, implemented for example in the Enrich2 package for analyzing deep mutational scanning data,[7] is to fit a weighted least-squares line through the time points. In contrast, likelihood-based approaches model sequencing counts at each time point through either a Poisson,[6] negative-binomial,[8] or multinomial[9] distribution, and use exponential growth to link the parameters of the distribution across time. This approach has recently been extended to Bayesian inference of growth rates, where uncertainties of the estimates are an integral part of the statistical model.[9]

Because all these analytical approaches assume sustained exponential growth, maintaining such conditions experimentally is essential. This can be achieved using a turbidostat,[2;5] which continuously dilutes the culture to maintain constant cell density, or through frequent dilutions.[10] In fixed-duration batch protocols (e.g., 12 or 24 hours), however, nutrient depletion and waste accumulation progressively limit growth, making the assumption of uninterrupted exponential behavior—including between dilutions—increasingly unrealistic.

More generally, a central assumption in estimating growth rates from sequencing-based pooled competition assays is that variants do not interact with one another, despite sharing the same physical environment. This stands in contrasts to selection experiments on microbial communities, where interspecies interactions that modulate individual growth rates are essential for determining both the resilience and the long-term composition of the community.[11;12;13] Similar inter-variant interactions may also occur in pooled competition assays. A well-studied example arises in the context of beta-lactam antibiotics: resistant bacteria expressing beta-lactamase enzymes hydrolyze the antibiotic, lowering its concentration in the medium and thereby reducing selection pressure for all members of the mixed population, including the susceptible ones.[14;15] Because of this complication, a recent study from our group on the beta-lactamase TEM-1 used the area under the variant growth curve as a proxy for fitness, rather than calculating explicit growth rates.[4]

Here, as a first step toward accommodating potential inter-variant interactions in pooled competition assays, we extend the sequencing-based analysis of growth rates to growth models with saturation, thereby moving beyond the assumption of variant independence and exponential growth. Unlike the specific interactions central to microbial community studies, the reduction of growth near saturation represents a non-specific interaction that influences all variants. To distinguish the initial, near-exponential phase from the subsequent saturation regime, sequencing at more than two time points is required. We therefore revisit existing strategies—least-squares fitting and time-parametrized likelihood methods—for analyzing time-resolved sequencing data.

In contrast to previous studies, we express the variant fractions as a *softmax* transformation of the logarithms of the variant abundances [eq. (11)]. Under exponential growth, this parametrization yields an exact, closed-form expression for the time-dependence of the variant fractions [eq. (12)]. Without this transformation, the dynamics of the fractions cannot be integrated directly—even in exponential growth—because they depend explicitly on the mean population growth rate.[6;8;9;10] As a result, likelihood-based methods,[6;8;10] including the recent Bayesian formulation,[9] face the problem of estimating the mean growth rate as a prerequisite to estimating the variant-specific growth rates.[6;8;9;10] Our parametrization removes explicit reference to the mean population growth rate, thereby eliminating the need to estimate it as an intermediate step. In the context of least-squares fitting, the *softmax* transformation also ensures that fits are invariant to the choice of reference variant, unlike the linear regression used in the Enrich2 analysis tool.[7]

When modeling saturation, we integrate the growth dynamics numerically for given parameter values of the growth model, and use automatic differentiation[16] to compute gradients of the loss function (likelihood, or evidence lower bound, in the case of variational Bayesian inference) with respect to these parameters. Although only the sequencing time points ultimately contribute to the loss function and its gradients, the numerical integration step size is independent of the temporal resolution of the sequencing data and can be chosen to be much smaller. This finer integration effectively interpolates the growth trajectories between

sequencing time points, enabling potential comparisons with complementary high-resolution measurements such as optical density or metabolomics data. If available, such data could serve as additional constraints for parameter estimation in the growth model. (This possibility is not explored in the present paper.)

Given the current costs of deep sequencing, the sampling frequency along the time axis is likely to remain limited in the near future. While a relatively small number of sequencing time points is sufficient for estimating growth rates in the absence of direct inter-variant interactions, it does not permit the simultaneous estimation of interaction parameters, whose number can be very large. For instance, in the case of pairwise interactions, the number of parameters scales quadratically with the number of variants and becomes prohibitively large even for tens of variants. For this reason, the estimation of such specific pairwise interactions is not attempted in the present work.

When modeling the noise in the experimental counts, we focus on counting noise arising from the partitioning of a finite number of sequencing reads among the variants. Other sources of variability—such as noise from the PCR amplification or physiological heterogeneity among isogenic variants—are inevitable,[6] but are not considered in our analysis.

We model variant growth deterministically. Because the cell is the minimal replicating unit, we assume that each cell can be assigned to a single genetic variant, with cells sharing identical genotypes belonging to the same variant. Situations that complicate this simple correspondence—such as ongoing genome replication without cell division, or plasmid-borne variants with variable copy numbers—should be analyzed on a case-by-case basis. With the exception of Sec. 6, we focus on exponential growth, not only because of its experimental relevance, but also because of its nice mathematical properties that allow analytical insight.

In Sec. 3 we examine the weighted least-squares procedure implemented, for example, in Enrich2.[7] In Sec. 4, we formulate the problem as maximum-likelihood estimation, thereby eliminating the need for least-squares fitting. In Sec. 5, we adopt a Bayesian inference framework to quantify the uncertainties of the estimated growth rates, following the pioneering work of Razo-Mejia and colleagues.[9] In these latter two sections, we also analyze the limiting case of only two sequencing time points—a scenario that remains experimentally attractive—and derive new analytical expressions for the growth-rate estimates and their uncertainties. Finally, in Sec. 6, we apply the developed inference framework to the logistic and Gompertz growth models, which serve as minimal examples of systems exhibiting inter-variant interactions. Our conclusions are summarized in Sec. 7, and additional technical details are provided in the Supporting Information (hereafter referred to as SI).

2 Describing exponential growth

When modeling the stochasticity of the sequencing read counts using a probability distribution—such as Poisson,[6] negative binomial,[8] or multinomial[9]—it is most natural to parametrize the distribution in terms of the true fractions of the variants in the culture. If $N_k(t)$ is the number of "cells" of variant k at time t, and

$$N(t) = \sum_{k=1}^{K} N_k(t) \tag{1}$$

is the total number of cells across all K variants, then the true variant fractions are

$$f_k(t) = \frac{N_k(t)}{N(t)}. \tag{2}$$

In a growth competition experiment, these fractions change with time, and thus the parameters of the probability distribution have different values at different sequencing time points. These values, however, are connected through the underlying growth dynamics of the variants.

When all variants grow exponentially, their cell numbers are described by the differential equations

$$\frac{\mathrm{d}N_k(t)}{\mathrm{d}t} = \lambda_k N_k(t), \tag{3}$$

where the specific growth rare λ_k is a constant characterizing variant growth at the given condition. The aim of the statistical analysis is to estimate the growth rates of all K variants from the numbers of sequencing reads over time.

From the differential equations of the cell numbers, analogous differential equations can be derived for the variant fractions, which most intuitively parametrize the probability distribution of the sequencing counts. Specifically, differentiating both sides of (2) with respect to time and making use of (3) yields

$$\frac{\mathrm{d}f_k(t)}{\mathrm{d}t} = [\lambda_k - \bar{\lambda}(t)]f_k(t), \tag{4}$$

where

$$\bar{\lambda}(t) := \sum_{k=1}^{K} \lambda_k f_k(t) \tag{5}$$

is the mean growth rate of the culture. (See SI, Sec. S1.2 for the missing steps.) For a changing variant composition, the mean growth rate changes with time, even if the individual variant growth rates are constant. Consequently, while the exponential growth of each variant is independent from the others in (3), the dynamics of the variant fractions in (4) are neither exponential nor decoupled.

Because the variant fractions parametrize the probability distribution of the sequencing counts, there is a tendency in the literature to also describe the growth dynamics in terms of them.[6;8;9;10] Specifically, the differential equation in (4) is commonly integrated between two successive sequencing time points as[9]

$$f_k(t_{\tau+1}) = f_k(t_\tau)\mathrm{e}^{[\lambda_k - \bar{\lambda}(t_\tau)](t_{\tau+1} - t_\tau)}, \tag{6}$$

where we have denoted the T times of sequencing by t_τ, with $\tau = 0, 1, \ldots, T-1$. (Placing the origin of the time axis at the first sequencing point, i.e., $t_0 = 0$, motivates starting the index τ from zero.) Note that (6) is actually an approximation to the true dynamics, since the integration from t_τ to $t_{\tau+1}$ assumes that the mean growth rate remains constant in this time interval.

When estimating variant growth rates using (6), it becomes necessary to first estimate the mean growth rate $\bar{\lambda}(t_\tau)$, before estimating the individual growth rates.[6;8;9;10] In evolution experiments, where the spontaneously occurring mutants initially comprise a tiny fraction of the population, the mean growth rate can be safely identified with the growth rate of the wild type.[6] However, such an approximation is not justified in growth competition experiments where all variants may be present from the start at similar abundances. One workaround, which effectively recreates the situation in evolution experiments, is to ensure that the wild type constitutes at least 95% of the population throughout the duration of the competition experiment.[9;10] Biologically, this also ensures that all other low-abundance variants grow in a wild-type context. It is important, however, to distinguish this biological requirement from the mathematical condition that the mean growth rate is approximately constant. While the wild-type context could be desired in itself, we argue that the mathematical approximation is actually not needed for the purposes of estimating growth rates.

Our approach is to describe exponential growth not directly in terms of variant fractions but in terms of the natural logarithms of the cell counts, which we call *log abundances*:

$$y_k(t) := \log N_k(t). \tag{7}$$

Differentiating (7) with respect to time and using (3) gives

$$\frac{\mathrm{d}y_k(t)}{\mathrm{d}t} = \lambda_k, \tag{8}$$

which can be integrated exactly, i.e., without assuming that the mean population growth rate is constant over the integration interval. With the initial log abundances defined as

$$\delta_k := \log N_k(0), \tag{9}$$

integration from $t_0 = 0$ to t yields

$$y_k(t) = \lambda_k t + \delta_k. \tag{10}$$

Thus, under exponential growth, the log abundances of the K variants evolve along linear trajectories determined by the $2K$ *growth parameters* $(\lambda_1, \ldots, \lambda_K)$ and $(\delta_1, \ldots, \delta_K)$.

For compact notation, we write variant-specific properties as K-dimensional (column) vectors in bold. For example, $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_K)$ collects the growth rates of all variants. For time-dependent quantities, we denote evaluation at the sequencing times t_τ with a subscript, e.g., $f_{k\tau} = f_k(t_\tau)$, and in vector form $\mathbf{f}_\tau = (f_{1\tau}, \ldots, f_{K\tau})$.

At this stage, we have two complementary representations: *variant fractions*, which naturally parametrize the probabilistic noise model and can therefore be directly estimated from sequencing counts; and *variant log abundances*, which conveniently describe the growth dynamics through their linear dependence on the growth parameters. These two descriptors are straightforwardly linked by combining their definitions in (2) and (7):

$$f_{k\tau} = \frac{\mathrm{e}^{y_{k\tau}}}{\sum_{j=1}^{K} \mathrm{e}^{y_{j\tau}}} := \phi_k(\mathbf{y}_\tau). \tag{11}$$

The vector-valued mapping $\boldsymbol{\phi}(\cdot)$ defined on the right-hand side of (11) is widely known in the machine-learning literature as the *softmax* transformation.[17] It is also referred to as the logistic function in the present context,[18] but because this latter term is more commonly used for the special case of a single scalar variable, here we use the former name.

For exponential growth, where the log abundances are linear in the growth parameters,

$$f_{k\tau} = \frac{\mathrm{e}^{\lambda_k t_\tau + \delta_k}}{\sum_{j=1}^{K} \mathrm{e}^{\lambda_j t_\tau + \delta_j}}. \tag{12}$$

Unlike (6), here the time dependence of the variant fractions is exact. Moreover, because the mean growth rate does not appear explicitly on the right-hand side of (12), the individual growth rates can be estimated directly, without the need to first estimate the mean growth rate of the culture, as we demonstrate in the following sections.[a]

In summary, the *softmax* transformation in (11) provides a bridge between the intuitive parameters of the probabilistic noise model (fractions on the left) and the mathematically convenient description of biological growth (log abundances on the right). Building on this observation, in the following sections we revisit some of the established approaches for estimating growth rates from sequencing read counts. Specifically, in Sec. 3 we examine the weighted least-squares procedure implemented in tools such as Enrich2,[7] in Sec. 4 we reformulate the problem as maximum-likelihood estimation, eliminating the need for least-squares fitting, and in Sec. 5 we adopt a Bayesian framework to assign uncertainties to the estimated growth parameters.[9]

When working with the *softmax* transformation, it is important to recognize that it does not possess a unique inverse. Adding the same constant to the log abundances of all variants in (11) leaves the variant fractions unchanged. Thus, when experimental data provide information only about variant fractions (so-called *compositional* data[19]), the log abundances can be determined only up to an arbitrary global shift. From the definition in (7), such a shift is equivalent to changing the units for $N_k(t)$—like expressing abundance in optical density (OD) instead of absolute cell counts. To obtain unique numerical values of the log abundances—for instance, to perform explicit calculations or report parameter estimates—this degree of freedom must be removed by imposing an additional, yet arbitrary, constraint. As an example, the values across all variants could be required to sum to zero. Often referred to as gauge fixing,[20] this step effectively amounts to selecting a particular unit for the abundances.

In the special case of exponential growth, both the growth rates and the initial log abundances in (12) can be shifted independently without altering the variant fractions. Consequently, each set of parameters must

[a] Equation (12) can be derived more directly by integrating (3) as $N_k(t) = \mathrm{e}^{\lambda_k t + \delta_k}$, and substituting into (2). Our detour through log abundances is motivated by the desire to generalize the statistical analysis beyond exponential growth. For such models, the general *softmax* relationship in (11) still holds, since it is an identity that follows from the definitions of $f_k(t)$ and $y_k(t)$, independently of the growth dynamics.

be constrained separately—for example, by fixing the sum of the growth rates to zero and the exponential sum of the initial log abundances to one. The latter choice corresponds to $\sum_k \mathrm{e}^{\delta_k} = N(0) = 1$, meaning that the total initial population size is taken as the unit for measuring variant abundances.

A different choice, which is often made in the literature, is to designate one variant—typically the wild type—as a reference, and then apply global shifts such that both the growth rate and initial log abundance of this reference variant are zero. To illustrate how this emerges in practice, we take the logarithm of both sides of the *softmax* function in (11):

$$\log f_{k\tau} = y_{k\tau} - \log \sum_{j=1}^{K} \mathrm{e}^{y_{k\tau}}. \tag{13}$$

While the first term on the right-hand side of the equality is specific to variant k, the second one—due to the normalization factor in the denominator of the *softmax* transformation—is the same for all variants (but can change at different time points t_τ). To cancel this common term, one can take the difference with another variant ρ, chosen as reference:

$$\log f_{k\tau} - \log f_{\rho\tau} = y_{k\tau} - y_{\rho\tau}. \tag{14}$$

Being obtained directly from the *softmax* transformation in (11), this equality also does not depend on the growth model.

For exponential growth, (14) becomes

$$\log f_{k\tau} - \log f_{\rho\tau} = \lambda_{k|\rho}\, t_\tau + \delta_{k|\rho}, \tag{15}$$

where we have introduced the *relative* properties (i.e., with respect to the reference)

$$\lambda_{k|\rho} := \lambda_k - \lambda_\rho, \qquad \delta_{k|\rho} := \delta_k - \delta_\rho. \tag{16}$$

The subscripts are intended to highlight that the growth parameters on the right-hand side of (15) are in a gauge where variant ρ has zero growth rate and zero initial log abundance. This choice, however, should not make the reference variant somehow special. In other words, the ultimate estimates of the growth parameters and their associated uncertainties should not depend on the choice of reference variant.

3 Combining estimates at different sequencing time points through least-squares fit

Because they link variant fractions to the parameters of exponential growth, the equalities (12) and (15) provide a natural starting point for fitting the analytical growth model to experimental sequencing data. In the current section, we estimate the left-hand sides of these equalities from the read counts at each sequencing time point. We then define a mean-squared error between the estimates at different time points and the predictions of the growth model on their right-hand sides. The parameters of the growth model are estimated by minimizing this error. This least-squares fitting approach is used, for example, in Enrich2.[7]

Concretely, let $f_{k\tau}^{\mathrm{ML}}$ denote the maximum-likelihood estimate of the fraction of variant k at time point t_τ (given in Sec. 3.1 below). Using (15), we enforce agreement between the estimates at all time points and the growth model by minimizing the mean-squared error

$$E_{k|\rho}(\lambda_{k|\rho}, \delta_{k|\rho}) = \sum_{\tau=0}^{T-1} \left[(\log f_{k\tau}^{\mathrm{ML}} - \log f_{\rho\tau}^{\mathrm{ML}}) - (\lambda_{k|\rho}\, t_\tau + \delta_{k|\rho}) \right]^2 \tag{17}$$

with respect to the growth parameters $\lambda_{k|\rho}$ and $\delta_{k|\rho}$. Because the objective involves a sum over all sequencing time points, the resulting estimates incorporate information from the entire temporal dataset.

In (17), each squared difference between empirical estimates and analytical predictions contributes equally to the error across all time points. Implicitly, this assumes that the random fluctuations of the maximum-likelihood estimates around the exact values from the growth model have the same variance at every time

point. To relax this assumption, the contributions should instead be normalized by their variances, which requires estimating from the data not only the means of the fluctuations but also the variances.

In Sec. 3.2, we use a Bayesian framework to infer the full probability distribution of the variant fractions, from which both means and variances can be calculated as needed. We then replace the error (17) with a *weighted* mean-squared error (Sec. 3.2.2), thereby also recovering the form implemented in Enrich2.[7] Finally, we repeat the procedure starting from the equality (12) to obtain an alternative mean-squared error with more favorable fitting properties (Sec. 3.2.3).

3.1 Maximum-likelihood estimates of the variant fractions at one time point

Let $n_{k\tau}$ be the number of sequencing reads assigned to variant k at time t_τ, and

$$n_\tau = \sum_{k=1}^{K} n_{k\tau} \tag{18}$$

be the total number of reads at this time. In vector notation, $\mathbf{n}_\tau$ collects the sequencing counts of all variants at time t_τ. We denote the complete experimental dataset of $K \times T$ counts by the matrix $\mathbf{N} = [n_{k\tau}]$, whose τth column is the vector $\mathbf{n}_\tau$. To similarly refer to the kth row of $\mathbf{N}$, which contains the counts of variant k across all sequencing time points, we use the notation $\vec{n}_k$.

We model the experimental read counts at sequencing time point t_τ as random draws from a multinomial distribution, parametrized by the variant fractions:

$$\mathbf{n}_\tau \sim \text{Multinomial}(\mathbf{f}_\tau; n_\tau). \tag{19}$$

In effect, the multinomial distribution partitions the total number of reads, n_τ, among the different variants according to the fractions, $\mathbf{f}_\tau$, but does this in a probabilistic manner. Specifically, the probability of observing a given combination of counts is

$$p(\mathbf{n}_\tau|\mathbf{f}_\tau) = C(\mathbf{n}_\tau) \prod_{k=1}^{K} (f_{k\tau})^{n_{k\tau}}, \tag{20}$$

where the normalization factor $C(\cdot)$ is not of direct relevance here, as it does not depend on the fractions that we aim to estimate.

Importantly, the multinomial model of counting noise inherently respects the compositional constraint that variant fractions must sum to one.[19] Its characteristic negative covariances capture the fact that an increase in the fraction of one variant necessarily entails a decrease in the fractions of the others. This interdependence is lost when the counting noise of individual variants is modeled independently with Poisson[6] or negative-binomial distributions.[8]

To estimate the parameters $\mathbf{f}_\tau$ from a given realization of the counts $\mathbf{n}_\tau$, we maximize the probability in (20), now reagreded as a function of the variant fractions and referred to as *likelihood*. Taking the logarithm gives the log likelihood

$$\mathcal{L}(\mathbf{f}_\tau|\mathbf{n}_\tau) := \log p(\mathbf{n}_\tau|\mathbf{f}_\tau) = \sum_{k=1}^{K} n_{k\tau} \log f_{k\tau}, \tag{21}$$

where only terms depending on the fractions have been retained.

Maximizing this log likelihood subject to the constraint that the fractions of all variants sum to one at t_τ, yields the maximum-likelihood (ML) estimates

$$f_{k\tau}^{\text{ML}} = \frac{n_{k\tau}}{n_\tau}. \tag{22}$$

(See Sec. S2.1.1 of the SI for missing steps.) Thus, the simple ratio of the counts of variant k to the total counts—an expression one would compute even without formal statistical analysis—is exactly the maximum-likelihood estimate of the true fraction of that variant. This outcome is, in a sense, inevitable: estimating

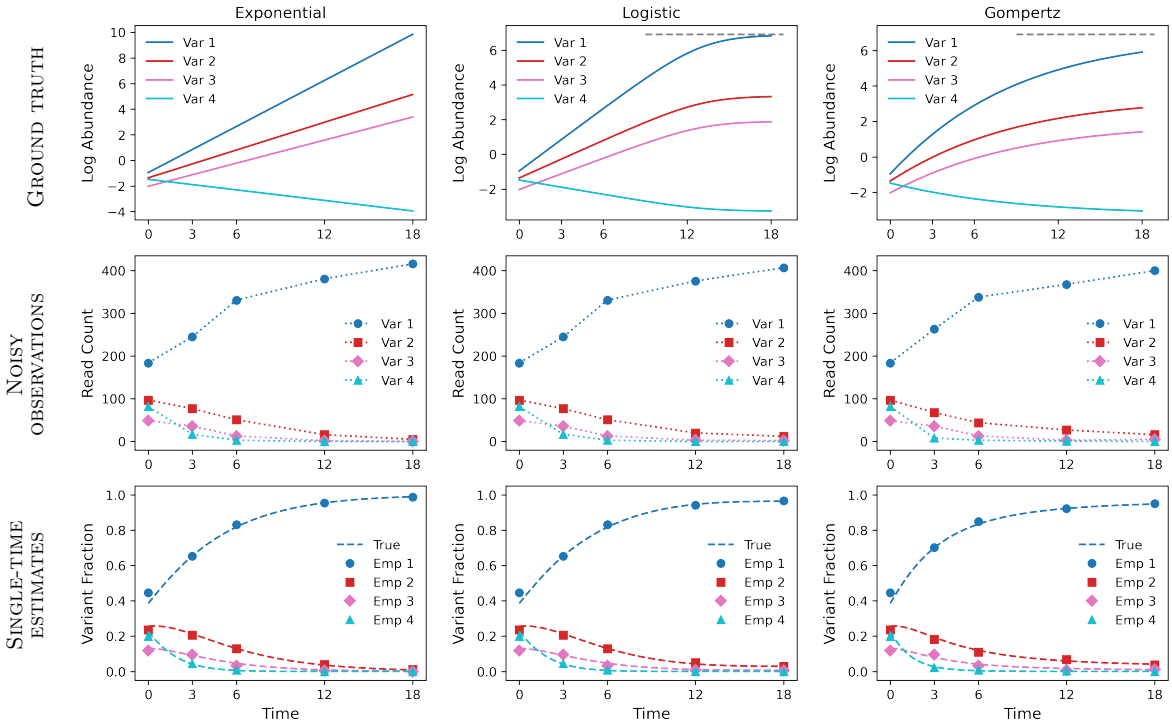

Figure 1: Illustrative example of exponential, logistic and Gompertz growth for $K = 4$ variants. Ground truth: Synthetic growth trajectories of the variant log abundances. Exponential growth lines are calculated analytically using (10). Logistic and Gompertz growth are integrated numerically using Euler's method with integration time step ten times smaller than the unit of time. Carrying capacity is indicated by horizontal dashed lines. Noisy observations: Random multinomial counts generated according to the true fractions of the variants at the five specified time points. Symbols correspond to the numerical values in (27) and (92). Connecting dotted lines are for visual purposes only. Single-time estimates: Comparison between the empirical variant fractions in (22) calculated from the counts data (symbols, Emp), and the true fractions from which the counts were generated (dashed lines, True).

K variant fractions from K observed counts leaves little room for alternative estimators. Because each observed count informs exactly one parameter, the estimates (22) do not benefit from any averaging over the stochastic fluctuations inherent in the counts. Assessing their uncertainty is therefore essential, a task we address in Sec. 3.2.1 using a Bayesian framework.

Substituting these maximum-likelihood estimates into the mean-squared error (17) yields

$$E_{k|\rho}(\lambda_{k|\rho}, \delta_{k|\rho}|\vec{n}_k, \vec{n}_\rho) = \sum_{\tau=0}^{T-1} \left[\log n_{k\tau} - \log n_{\rho\tau} - (\lambda_{k|\rho} t_\tau + \delta_{k|\rho})\right]^2 . \tag{23}$$

Note that the denominator in (22), which is identical for all variants, cancels when taking the difference of the logarithms. As a result, the error (23) depends only on the counts of variant k and ρ across all time points, denoted by $\vec{n}_k$ and $\vec{n}_\rho$. To estimate the complete set of growth parameters, analogous mean-squared errors must be minimized for all non-reference variants. By definition, the parameters of the reference variant are fixed at $\lambda_{\rho|\rho} = \delta_{\rho|\rho} = 0$.

Throughout the paper, we use a simple synthetic example with four variants for illustration. The top-left panel in fig. 1 shows the exponential growth of these variants. (Only the plots in the left-most column correspond to exponential growth; the other two columns depict the two models with saturation analyzed in

Sec. 6.) Because we plot the log abundances (10), the trajectories appear as straight lines.

We normalize the cell numbers so that the initial total cell density equals one, $N(0) = 1$. Dividing this density among the four variants according to the arbitrarily chosen initial fractions

$$\mathbf{f}_0 = \begin{bmatrix} 0.38486 & 0.25597 & 0.13186 & 0.22731 \end{bmatrix} \tag{24}$$

gives the initial log abundances

$$\boldsymbol{\delta} = \begin{bmatrix} -0.9549 & -1.3627 & -2.0260 & -1.4814 \end{bmatrix}, \tag{25}$$

which correspond to the y-intercepts of the Ground truth lines in fig. 1. The growth rates, which correspond to the slopes of these lines, are chosen as

$$\boldsymbol{\lambda} = \begin{bmatrix} 0.5998 & 0.3613 & 0.3006 & -0.1372 \end{bmatrix}. \tag{26}$$

These have units of inverse time. While the time unit here is arbitrary, it is intended to correspond to an hour.

From the linear log abundances in the top row of fig. 1, we compute the variant fractions at the five specified time points (not spaced equally) and use these to generate random multinomial counts. To mimic experimental sequencing, the total counts at each time point are drawn from a Poisson distribution with mean 400, yielding roughly 100 counts per variant on average. One realization is

$$\mathbf{N} = \begin{bmatrix} 184 & 245 & 331 & 381 & 416 \\ 97 & 77 & 51 & 16 & 5 \\ 49 & 36 & 13 & 2 & 0 \\ 82 & 17 & 3 & 0 & 0 \end{bmatrix}, \tag{27}$$

where rows correspond to variants and columns to time points. These values are visualized in the second, Noisy observations row of fig. 1.

From these synthetic counts, we compute the single-time maximum-likelihood estimates (22). These are plotted as symbols in the bottom-left panel of fig. 1. As expected, the empirical fractions closely track the true generating fractions (dashed lines).

In this example, the fastest-growing variant rapidly dominates the population, eventually accounting for nearly all observed counts. As a result, the counts across variants begin to differ by several orders of magnitude, with some variants disappearing altogether. Such pronounced imbalances are also expected in large-scale growth competition experiments involving thousands of variants.

Because rare variants are particularly susceptible to counting noise, a principled approach to estimate model parameters together with their associated uncertainties is needed. In Sec. 5, we use Bayesian inference as a general framework for parameter estimation and uncertainty quantification.[9] In the next subsection, we introduce Bayesian analysis to derive variance-based weight factors for the least-squares fitting procedure.

3.2 Weighted least-squares from Bayesian inference at one time point

Similar to the maximum-likelihood estimation described above, the Bayesian inference performed here is applied to data from a single sequencing time point. Information across time points is subsequently combined through least-squares minimization. As a result, in the presentation of the Bayesian analysis in Sec. 3.2.1, the time index τ may be safely disregarded, although we retain it as a subscript for clarity.

3.2.1 Bayesian analysis

In Bayesian inference, the parameters of the multinomial distribution, $\mathbf{f}_\tau$, are promoted to random variables that have their own probability distribution, just like the counts $\mathbf{n}_\tau$. Unlike the counts, however, the variant fractions are not observed in the experiment; they remain hidden, or latent. The objective is to determine

the probability distribution of these latent variables given the observed data. This *posterior* distribution is denoted by $p(\mathbf{f}_\tau|\mathbf{n}_\tau)$.

The tool leveraged to this end is the Bayes theorem

$$p(\mathbf{f}_\tau|\mathbf{n}_\tau) = \frac{p(\mathbf{n}_\tau|\mathbf{f}_\tau)p(\mathbf{f}_\tau)}{p(\mathbf{n}_\tau)}. \tag{28}$$

Here the generative model of the data, $p(\mathbf{n}_\tau|\mathbf{f}_\tau)$, and the *prior* distribution of the parameters, $p(\mathbf{f}_\tau)$, are treated as known inputs, and the posterior as the desired output. While the prior reflects our knowledge—or lack thereof—about the parameters before seeing the data, the posterior represents the updated knowledge once the data are observed. The denominator in (28) ensures that the posterior is properly normalized over the fractions.

When the generative model in (28) is multinomial, choosing a Dirichlet prior yields a posterior that is also Dirichlet but with different parameters. This conjugacy of the Dirichlet distribution to the multinomial distribution is easily established by substituting the two probability densities into the numerator of (28). A Dirichlet distribution with parameters $\boldsymbol{\alpha}_\tau$ has the probability density

$$p(\mathbf{f}_\tau|\boldsymbol{\alpha}_\tau) = B(\boldsymbol{\alpha}_\tau)\prod_{k=1}^{K}(f_{k\tau})^{\alpha_{k\tau}-1}, \tag{29}$$

where, as before, the exact normalization factor $B(\cdot)$ is not important here because it does not depend on the fractions.

We denote the parameters of the Dirichlet prior by $\tilde{\boldsymbol{\alpha}}_\tau$, and those of the posterior by $\boldsymbol{\alpha}_\tau$. By examining how both sides of (28) depend on the variant fractions, we obtain

$$\alpha_{k\tau} = n_{k\tau} + \tilde{\alpha}_{k\tau} \tag{30}$$

(see Sec. S2.2 of the SI for intermediate steps). In other words, for each variant, the observed count $n_{k\tau}$ is increased by the prior parameter $\tilde{\alpha}_{k\tau}$ to produce the posterior parameter $\alpha_{k\tau}$. The prior parameters therefore act as pseudo-counts that are added to the actual counts.

Because variant counts generally differ across sequencing time points, the posterior parameters also vary with time. However, the same prior parameters can be used for all time points, i.e., $\tilde{\boldsymbol{\alpha}}_\tau = \tilde{\boldsymbol{\alpha}}$. In practice, it is also common to use the same prior value for all variants, i.e., $\tilde{\alpha}_k = \tilde{\alpha}$. As an example, uniform Dirichlet prior corresponds to $\tilde{\alpha} = 1$. Alternatively, the choice $\tilde{\alpha} = 1/2$, which is known as Jeffreys prior, is used in Enrich2.[7]

According to (30), the K parameters of the Dirichlet posterior are determined directly from the K experimental read counts at a given time point. In other words, each posterior parameter $\alpha_{k\tau}$ is informed by a single observed count $n_{k\tau}$, just as in the maximum-likelihood estimate (22). The key difference is that maximum likelihood produced a single point estimate for the fraction of each variant, whereas the Dirichlet posterior defines a full probability distribution. This distribution can be used not only to compute the mean of each variant fraction but also to quantify the spread around the mean, for example, in the form of variance.

For a Dirichlet posterior, the means and variances of the fractions are

$$\mu^f_{k\tau} := \mathbb{E}_{\mathrm{Dir}}[f_{k\tau}] = \frac{\alpha_{k\tau}}{\alpha_\tau}, \qquad \mathrm{Var}_{\mathrm{Dir}}[f_{k\tau}] = \frac{\mu^f_{k\tau}(1-\mu^f_{k\tau})}{\alpha_\tau + 1}, \tag{31}$$

where the denominators contain the sum over all Dirichlet parameters

$$\alpha_\tau = \sum_{k=1}^{K}\alpha_{k\tau}. \tag{32}$$

In these expressions, $\mathbb{E}[\cdot]$ and $\mathrm{Var}[\cdot]$ are operators that take, respectively, the expectation and variance of their arguments. We have included the subscript 'Dir' as a reminder that these are calculated for a Dirichlet posterior.

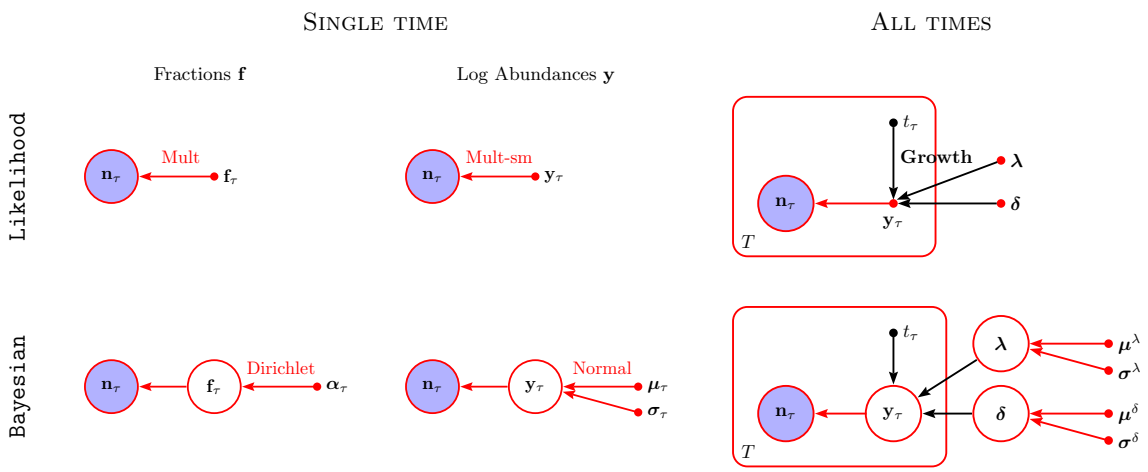

Figure 2: Graphical representation of the probabilistic models examined in this work. Random variables are shown with circles and parameters with points.[17] Blue shading indicates that samples of the random variable are available as experimental data. Empty circles signify that the random variables are not observed and thus remains *hidden* or *latent*.[17] Such latent variables arise naturally in `Bayesian` inference, where the parameters of the `likelihood` are promoted to random variables. Red arrows are used for probabilistic relationships between the variables and black arrows for deterministic functional dependence through the growth model. Variables enclosed in rectangular frame are repeated T times, with the index τ running over the repetitions. Note that only the time t_τ changes between the T repetitions, while the growth parameters $\boldsymbol{\lambda}$ and $\boldsymbol{\delta}$ are shared across all time points. The task of inference consists of estimating values for the parameters represented with red points from given samples of the count variables in blue-shaded circles.

In (31), the expected fraction of variant k equals the ratio between its Dirichlet parameter, $\alpha_{k\tau}$, and the sum of the Dirichlet parameters of all variants, α_τ. For $\tilde{\alpha} = 0$ (i.e., zero pseudo-counts) this ratio becomes the maximum-likelihood estimate (22). These expected variant fractions remain unchanged if all posterior Dirichlet parameters are scaled by a common constant. In contrast, the variances in (31) are sensitive to such rescaling and therefore depend on the absolute magnitude of the observed counts. Consequently, $K-1$ degrees of freedom of the Dirichlet distribution determine the mean variant fractions, while the remaining degree of freedom—the sum of all K Dirichlet parameters in (32)—controls the scale of the deviations from these means. Indeed, α_τ is often referred to as the *precision* of the Dirichlet distribution, as it reflects how tightly the distribution is concentrated around its mean.[21]

Before turning to the weighted least-squares fit to the data, we provide a visual summary of the transition from likelihood-based estimation to Bayesian inference developed thus far. In the top-left panel of fig. 2, we depict the probabilistic model from (19), where the random counts $\mathbf{n}_\tau$ follow a multinomial (Mult) distribution with parameters $\mathbf{f}_\tau$. In this graphical representation, random variables are depicted as circles, while the parameters of their probability distributions are indicated by points. The transition to `Bayesian` modeling, in which the parameters $\mathbf{f}_\tau$ are treated as random variables, is shown in the lower-left panel of fig. 2. These random variables follow a Dirichlet distribution with parameters $\boldsymbol{\alpha}_\tau$, which are represented by the point. The circle of $\mathbf{f}_\tau$ is left white to indicate that these random variables are latent. In contrast, random realizations of the counts $\mathbf{n}_\tau$ are available as experimental data, which is indicated by the blue shading of their circle. In inference, the purpose is to estimate the values of the parameters (red points) from the data (blue-shaded circles).

In anticipation of what is to come, we note that the two additional models in the top row of fig. 2 are discussed in Sec. 4, while the two remaining models in the bottom row appear in Sec. 5.

3.2.2 Fitting the logarithms of the fractions

Having inferred a posterior distribution, we replace the error (17) with the weighted mean-squared error

$$E_k^{\log}(\lambda_{k|\rho}, \delta_{k|\rho}) = \sum_{\tau=0}^{T-1} \frac{1}{\mathrm{Var}_{\mathrm{Dir}}[\log f_{k\tau} - \log f_{\rho\tau}]} \left\{\mathbb{E}_{\mathrm{Dir}}[\log f_{k\tau} - \log f_{\rho\tau}] - (\lambda_{k|\rho} t_\tau + \delta_{k|\rho})\right\}^2 . \tag{33}$$

Here, the maximum-likelihood estimates of $\log f_{k\tau} - \log f_{\rho\tau}$ are replaced by posterior expectations, and the squared deviations are normalized by the corresponding posterior variances. This weighting accounts for heterogeneity in statistical uncertainty across time points, ensuring that more precise estimates contribute more strongly to parameter inference.

For a Dirichlet distribution, the means and variances in (33) are available in closed form in terms of the digamma and the trigamma functions, $\psi_0(\cdot)$ and $\psi_1(\cdot)$:

$$\mathbb{E}_{\mathrm{Dir}}[\log f_{k\tau} - \log f_{\rho\tau}] = \psi_0(\alpha_{k\tau}) - \psi_0(\alpha_{\rho\tau}), \qquad \mathrm{Var}_{\mathrm{Dir}}[\log f_{k\tau} - \log f_{\rho\tau}] = \psi_1(\alpha_{k\tau}) + \psi_1(\alpha_{\rho\tau}), \tag{34}$$

with the condition $k \neq \rho$ for the variance.

Although it does not bring much additional insight, the expressions in (34) can be substituted into the weighted mean-squared error (33) in order to explicitly show the dependence on the experimental counts. For the case of a single pseudo-count parameter $\tilde{\alpha}$,

$$E_k^{\log}(\lambda_{k|\rho}, \delta_{k|\rho} | \vec{n}_k, \vec{n}_\rho) = \sum_{\tau=0}^{T-1} \frac{\left[\psi_0(n_{k\tau} + \tilde{\alpha}) - \psi_0(n_{\rho\tau} + \tilde{\alpha}) - (\lambda_{k|\rho} t_\tau + \delta_{k|\rho})\right]^2}{\psi_1(n_{k\tau} + \tilde{\alpha}) + \psi_1(n_{\rho\tau} + \tilde{\alpha})}. \tag{35}$$

As far as we are aware, this expression has not been used in the literature. What has been used, for example in Enrich2,[7] is its limit for large variant counts, which we derive next.

The digamma and trigamma functions admit the following asymptotic expansions:

$$\psi_0(z) \sim \log z - \frac{1}{2z} + \ldots, \qquad \psi_1(z) \sim \frac{1}{z} + \frac{1}{2z^2} + \ldots . \tag{36}$$

As the argument z increases, the accuracy of the partial sums improves. Retaining only the first terms, we can approximate the error (35) as

$$E_k^{\log}(\lambda_{k|\rho}, \delta_{k|\rho} | \vec{n}_k, \vec{n}_\rho) \approx \sum_{\tau=0}^{T-1} \frac{\left[\log(n_{k\tau} + \tilde{\alpha}) - \log(n_{\rho\tau} + \tilde{\alpha}) - (\lambda_{k|\rho} t_\tau + \delta_{k|\rho})\right]^2}{\frac{1}{n_{k\tau}+\tilde{\alpha}} + \frac{1}{n_{\rho\tau}+\tilde{\alpha}}}. \tag{37}$$

For $\tilde{\alpha} = 1/2$, this reduces to the weighted least-squares error used in Enrich2.[7]

Since the error for variant k is affected only by the parameters $\lambda_{k|\rho}$ and $\delta_{k|\rho}$ of that variant, a separate minimization should be preformed for each non-reference variant. Alternatively, one can pool the individual errors together, and minimize the single total error

$$E^{\log}(\boldsymbol{\lambda}_{|\rho}, \boldsymbol{\delta}_{|\rho} | \mathbf{N}) = \sum_{k=1}^{K} E_k^{\log}(\lambda_{k|\rho}, \delta_{k|\rho} | \vec{n}_k, \vec{n}_\rho) \tag{38}$$

with respect to all growth parameters, keeping those of the reference variant equal to zero.

For the illustrative example with four variants, we minimized the total error formed by the sum of the errors (35), with variant $\rho = 2$ as reference and $\tilde{\alpha} = 1$ for the prior (i.e., one additional pseudo-count). The fits for the three non-reference variants are shown as lines in the upper half of fig. 3. In addition, we show the target posterior means as orange points and the posterior standard deviations as orange error bars. When the standard deviation is large, the squared difference between the linear model and the posterior mean contributes less to the total error. As expected, the largest standard deviations correspond to zero

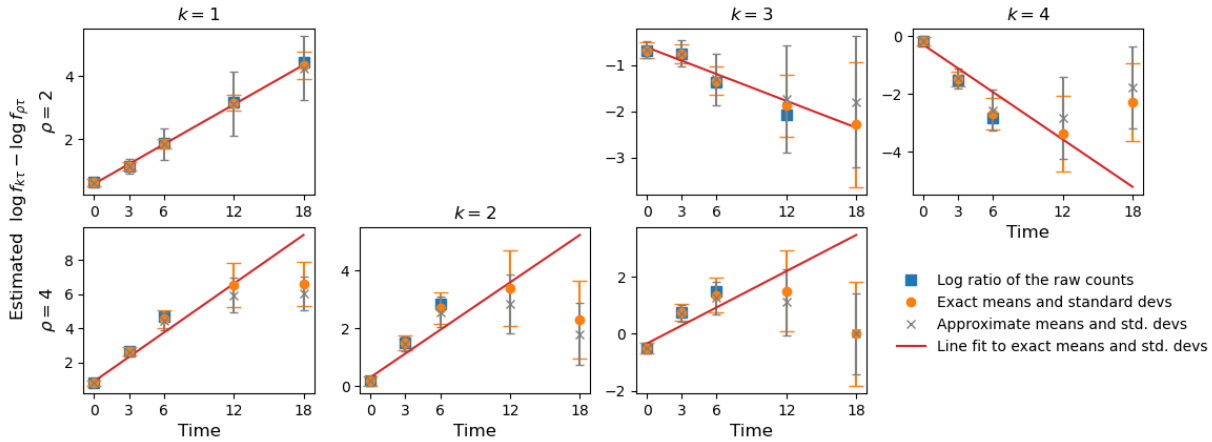

Figure 3: Linear weighted least-squares fits with either variant 2 (top row) or variant 4 (bottom row) serving as reference. Exact means and standard deviations from (34) are in orange and approximations used in the error (37) are in gray. Blue squares show $\log n_{k\tau} - \log n_{\rho\tau}$, which are the maximum-likelihood estimates used without any pseudo-counts in the error (23).

variant counts (last point of variant 3 and last two points of variant 4). In these cases, the magnitudes of the standard deviations are mainly dictated by the pseudo-count parameter $\tilde{\alpha}$.

For comparison, we also show the approximate means and standard deviations used in the error (37), again calculated with a uniform prior ($\tilde{\alpha} = 1$). To assess the effect of the pseudo-counts, we additionally plot the maximum-likelihood estimates used in the error (23), which does not include any pseudo-counts (blue squares). Note that these are available only when the counts of both the analyzed variant and the reference variant are non-zero, since otherwise the logarithms evaluate to minus infinity.

For the non-zero counts, both the approximate means (gray crosses) and the log ratios of the raw counts (blue squares) agree well with the exact posterior means (orange points). Not surprisingly, the discrepancy between the exact and approximate posterior means is larger for zero counts, since the approximation obtained by terminating the asymptotic expansions at the first term applies for large count numbers.

Focusing now on the proximity of the fitted line to the target points, from the top-right panel in fig. 3 we observe that the last two points of variant 4, which lack any actual counts, pull the fitted line away from the linear trend of the first three points. As a result, at three out of the five points—including two points with non-zero counts—the fitted line is more than one standard deviation away from the means.

The fit of the most abundant variant 1 in the top-left panel of fig. 3 is perfect. What is interesting in this case is the increase of the standard deviations at the last two time points, in spite of the fact that the counts of variant 1 increase steadily with time and attain their largest values there. Somehow counterintuitively, the largest counts of variant 1 at the last two time points contribute to the total error with smaller weights than the comparatively smaller counts at the first three time points. The explanation of this seemingly strange phenomenon lies in the fact that the error bars correspond to the variances in (34), which involve the reference variant. Indeed, the counts of variant 2 decrease steadily with time and become relatively small at the last two time points, thus increasing the standard deviations.

To better demonstrate the influence of the reference variant on the fits, in the lower half of fig. 3 we show the same analysis but with $\rho = 4$ used as a reference. In this case, even for the most abundant variant 1 the line lies more than one standard deviation from two of the five target points shown in orange. For all variants, the last two points visibly deviate from the linear trend of the first three points.

The dramatic effect of the reference variant on the fits, as observed in this simple example, contradicts our expectation that the selection of reference is needed only to fix the trivial global shift in growth rates and initial log abundances. We defer the discussion of the underlying reason for this discrepancy to the end of the next subsection, after first introducing an alternative mean-squared error that aligns with our

expectation.

3.2.3 Fitting the fractions

The mean-squared error of the previous subsection was constructed from the mathematical identity (15). Here we form a different mean-squared error, starting with the *softmax* identity (12). For consistency with the previous analysis, we also fix the gauge of the *softmax* function by subtracting the growth rate and initial log abundance of a reference variant, thus again working with values relative to ρ. The resulting error for variant k is

$$E_k^{\text{frac}}(\boldsymbol{\lambda}_{|\rho}, \boldsymbol{\delta}_{|\rho}) = \sum_{\tau=0}^{T-1} \frac{1}{\text{Var}_{\text{Dir}}[f_{k\tau}]} \left(\mathbb{E}_{\text{Dir}}[f_{k\tau}] - \frac{\mathrm{e}^{\lambda_{k|\rho} t_\tau + \delta_{k|\rho}}}{\sum_{j=1}^{K} \mathrm{e}^{\lambda_{j|\rho} t_\tau + \delta_{j|\rho}}} \right)^2 , \tag{39}$$

where the expectations and variances of the fractions should be estimated from the data. For the Bayesian analysis with Dirichlet posterior, these means and variances are given in (31). We have used the superscript 'frac' to distinguish this new error, which fits the fractions, from the error (33), which fits the log differences of the fractions.

Differently from the situation before, the error of variant k in (39) depends on the growth parameters of all variants, since these are used in the denominator of the *softmax* function for the purposes of normalization. To minimize all parameters simultaneously, we form a single, scalar loss function, giving equal weights to all variants:

$$E^{\text{frac}}(\boldsymbol{\lambda}_{|\rho}, \boldsymbol{\delta}_{|\rho}) = \sum_{k=1}^{K} E_k^{\text{frac}}(\boldsymbol{\lambda}_{|\rho}, \boldsymbol{\delta}_{|\rho}). \tag{40}$$

Note that the sum over k includes the reference variant, which means that it is also part of the minimization. Its growth parameters, however, should be manually set to zero after each iteration step in order to impose the chosen gauge. (The reference-variant gauge is selected here for consistency with the previous analysis, but is not required by the error E^{frac}. In fact, any other gauge can be imposed at this stage.) Substituting the means and variances from (31), the total error becomes

$$E^{\text{frac}}(\boldsymbol{\lambda}_{|\rho}, \boldsymbol{\delta}_{|\rho}|\mathbf{N}) = \sum_{k=1}^{K} \sum_{\tau=0}^{T-1} \frac{(\alpha_\tau + 1)}{\frac{\alpha_{k\tau}}{\alpha_\tau}\left(1 - \frac{\alpha_{k\tau}}{\alpha_\tau}\right)} \left[\frac{\alpha_{k\tau}}{\alpha_\tau} - \phi_k(\boldsymbol{\lambda}_{|\rho} t_\tau - \boldsymbol{\delta}_{|\rho}) \right]^2 , \tag{41}$$

where the dependence on the raw counts $\mathbf{N}$ comes from the posterior Dirichlet parameters $\alpha_{k\tau}$ according to (30).

For the illustrative example with four variants, we minimized the error (41) with $\tilde{\alpha} = 1$, again selecting variant $\rho = 2$ as a reference. The result of the fit is shown in fig. 4. On the vertical axes we now plot the variant fractions, since their means (orange points) are the target values of the fit. The fractions are plotted on a linear scale in the upper half of the figure, and on a logarithmic scale in the lower half. The fitted analytical expression (blue lines) is no longer linear but is given by the *softmax* function.

Looking at the top panels first, we see that the error bars of variant 1 decrease with time, reflecting the increase of its counts. Interestingly, the error bars of all other variants also decrease with time, although their counts are gradually depleted, and eventually become zero for variants 3 and 4. As a result, the zero counts of these two variants have tiny error bars, and thus contribute to the total error with large weights, in contrast to the linear fits in the top row of fig. 3.

However, it is not fair to directly compare the error bars in the upper halfs of figs. 3 and 4, because the latter shows the variant fractions and the former the logarithms of these fractions. To be able to visually assess the quality of the two alternative fitting strategies, in the lower half of fig. 4 we show the same variant fractions on a logarithmic scale. These lower panels are therefore directly comparable to those in fig. 3, where the logarithms of the fractions are on a linear scale.

Satisfyingly, for variant 4 in the lower-right panel of fig. 4, the error bars of the last two time points are much larger than the error bars of the first three points, in qualitative agreement with what we have in the upper-right panel of fig. 3. However, while in fig. 3 these zero-count points were pulling the fitted red line

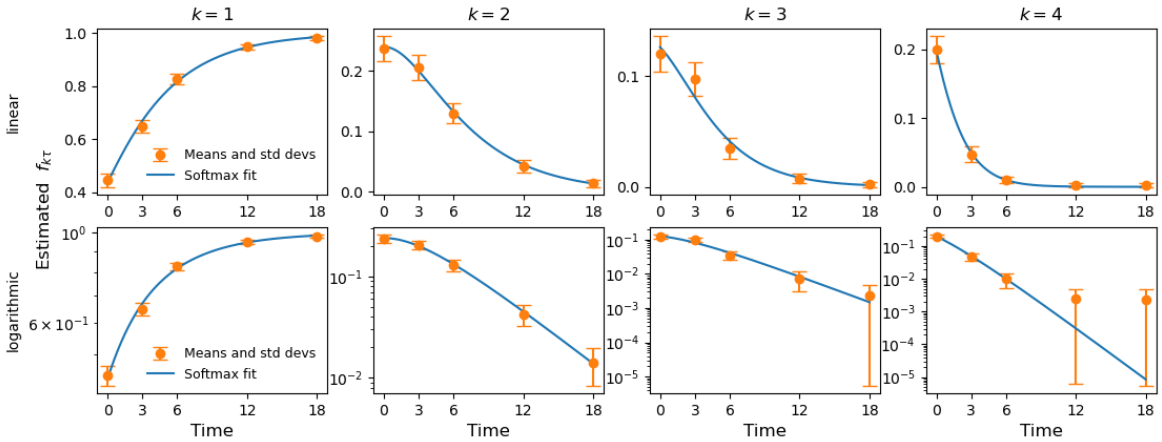

Figure 4: Non-linear weighted least-squares fit obtained by minimizing the error (41). The same variant fractions are displayed on linear (top row) and logarithmic (bottom row) vertical scales. Although variant $\rho = 2$ is used as a reference, identical fits are obtained when ρ is any other variant.

away from the trend of the first three points, here the fitted blue curve not only goes straight through the first three points, but also remains within one standard deviation of the last two points.

In the case of variant 1 in the lower-left panel of fig. 4, the error bars at later times are smaller than those at earlier times, which is the opposite of what we had in the upper-left panel of fig. 3 for the linear fit of this variant. Recall, however, that the error bars in the linear fit depend also on the counts of the reference variant, while those of the *softmax* fit do not. Indeed, the error bars of variant 2 in the lower part of fig. 4, increase with time on a log scale, in line with the increase of the error bars in the upper half of fig. 3 where this variant is used as a reference.

The above comparison shows that the mean-squared errors (35) and (41) treat the reference variant in fundamentally different ways. While the linear fits depend strongly on the choice of reference, as demonstrated in fig. 3, the *softmax* fits do not. Indeed, exactly the same fits as those shown in fig. 4 are obtained when the minimization is repeated with $\rho = 4$, even though its last two counts are zero. Therefore, for the purposes of fitting exponential growth, the *softmax* fit is more satisfying than the linear fit.

The dramatically different performance of the two alternatives stems from the different normalization choices. The *softmax* transformation (12), on which the error (41) is based, cancels the effect of the global shift through the division by the sum over all variants. Correspondingly, the estimated fractions contain normalization by α_τ, which also sums over all variants. Ironically, the motivation for arriving at (15), on which the linear fit is based, was to cancel this normalization by all variants. Subtracting the reference for that purpose, amounts to normalization by the reference variant, which becomes apparent when the difference of the logarithms is written as $\log(f_{k\tau}/f_{\rho\tau})$. While, analytically, the two normalizations are as good as any other way of fixing the gauge, from the perspective of estimation normalization by a single variant is more susceptible to counting noise than normalization by all variants.

Historically, a linear fit may have been preferred over a non-linear, *softmax* fit for computational reasons, since minimizing the error in linear regression requires a single matrix inversion. By contrast, minimizing the *softmax* error requires iterative numerical optimization, such as gradient descent. Moreover, the linear fit can be performed independently for each variant, optimizing only two parameters at a time, whereas the *softmax* fit requires the simultaneous optimization of all $2K$ growth parameters. With the advent of modern computational resources and widespread availability of automatic differentiation libraries,[16] these computational disadvantages of the weighted *softmax* regression are no longer prohibitive—even when analyzing thousands of variants.

4 Maximum-likelihood estimation of the growth parameters

The weighted least-squares approach to estimating growth parameters—including its *softmax* variant—suffers from two main shortcomings. First, although Bayesian reasoning is invoked to construct the weight factors, the method ultimately produces only point estimates of the growth rates, with no quantitative measure of their uncertainty. This limitation arises because Bayesian inference is applied too early in the analysis pipeline: instead of capturing uncertainty in the final growth-rate estimates, it is directed toward the intermediate variant fractions. Second, the approach aggregates independent maximum-likelihood estimates from individual time points by minimizing a mean-squared error, thereby treating the temporal structure of the data only indirectly, as an afterthought. Yet the time axis can be incorporated directly within the maximum-likelihood framework itself, as we demonstrate in this section. Building on the likelihood function introduced here, we then turn in Sec. 5 to a fully Bayesian treatment of the growth parameters.

4.1 Softmax reparametrization of the multinomial distribution

To obtain maximum-likelihood estimates of the growth parameters directly, the later must appear explicitly in the likelihood of the probabilistic model. However, the multinomial distribution that we use to represent counting noise is parametrized in terms of variant fractions. We therefore reparametrize this distribution in two successive steps. First, using the *softmax* transformation (11), we express the multinomial distribution in terms of the log abundances of the variants, yielding the *multinomial-softmax* model shown in the middle-top panel of fig. 2. The parameters of this model, $\mathbf{y}_\tau$, still vary across sequencing time points. In the second step, we incorporate the growth model to express these time-dependent log abundances in terms of the constant growth parameters. This yields the probabilistic model depicted in the upper-right panel of fig. 2, whose likelihood can now be maximized directly with respect to $\boldsymbol{\lambda}$ and $\boldsymbol{\delta}$.

Changing the parametrization of the multinomial distribution in (20) from variant fractions to log abundances yields what we call the *multinomial-softmax* distribution:

$$p(\mathbf{n}_\tau|\mathbf{y}_\tau) = C(\mathbf{n}_\tau) \prod_{k=1}^{K} [\phi_k(\mathbf{y}_\tau)]^{n_{k\tau}}. \tag{42}$$

Importantly, this change of variables does not require a Jacobian factor, since we are reparametrizing the parameters of the distribution rather than the random variables themselves.

The corresponding log likelihood is

$$\mathcal{L}(\mathbf{y}_\tau|\mathbf{n}_\tau) = \log p(\mathbf{n}_\tau|\mathbf{y}_\tau) = \sum_{k=1}^{K} n_{k\tau} y_{k\tau} - n_\tau \log\left(\sum_{k=1}^{K} \mathrm{e}^{y_{k\tau}}\right), \tag{43}$$

where we omit terms independent of $y_{k\tau}$. Setting the derivative of this log likelihood with respect to $y_{k\tau}$ to zero yields the maximum-likelihood (ML) conditions (see SI, Sec. S2.1.2)

$$\phi_k(\mathbf{y}_\tau^{\mathrm{ML}}) = \frac{n_{k\tau}}{n_\tau}, \tag{44}$$

which are equivalent to the estimates $f_{k\tau}^{\mathrm{ML}}$ in (22).

4.2 Combining all time points

To integrate information across all time points, we assume that the counting noise at one sequencing time is independent of that at any other time. Under this assumption, the single-time log likelihoods (43) can be summed over the sequencing times to obtain the combined log likelihood

$$\mathcal{L}(\mathbf{Y}|\mathbf{N}) = \sum_{\tau=0}^{T-1} \left[\sum_{k=1}^{K} n_{k\tau} y_{k\tau} - n_\tau \log\left(\sum_{k=1}^{K} \mathrm{e}^{y_{k\tau}}\right)\right], \tag{45}$$

which relates the log abundances $\mathbf{Y} = [y_{k\tau}]$ to the observed counts $\mathbf{N}$.

However, summing across time does not by itself achieve a joint analysis: maximizing (45) with respect to the log abundances $y_{k\tau}$ recovers the single-time maximum-likelihood estimates in (44). True integration across time arises only when the growth model is used to express the log abundances at different time points in terms of a common set of growth parameters.

For exponential growth, substituting (10) yields the log likelihood

$$\mathcal{L}(\boldsymbol{\lambda}, \boldsymbol{\delta}|\mathbf{N}) = \sum_{\tau=0}^{T-1} \left[\sum_{k=1}^{K} n_{k\tau}(\lambda_k t_\tau + \delta_k) - n_\tau \log \left(\sum_{k=1}^{K} \mathrm{e}^{\lambda_k t_\tau + \delta_k} \right) \right], \tag{46}$$

which now depends explicitly on the growth parameters in addition to all sequencing counts. Maximizing (46) therefore estimates $2K$ parameter from $K \times T$ observations—corresponding to $T/2$ observed counts per estimated parameter.

To understand how maximum-likelihood estimation integrates information across time, it is instructive to examine the gradients of the log likelihood (46) with respect to the growth parameters. Differentiating with respect to λ_k and δ_k, and setting the derivatives to zero, we find that the optimal parameters $\boldsymbol{\lambda}^{\mathrm{ML}}$ and $\boldsymbol{\delta}^{\mathrm{ML}}$ must satisfy

$$\sum_{\tau=0}^{T-1} t_\tau \left[n_{k\tau} - n_\tau \phi_k(\boldsymbol{\lambda}^{\mathrm{ML}} t_\tau + \boldsymbol{\delta}^{\mathrm{ML}}) \right] = 0, \qquad \sum_{\tau=0}^{T-1} \left[n_{k\tau} - n_\tau \phi_k(\boldsymbol{\lambda}^{\mathrm{ML}} t_\tau + \boldsymbol{\delta}^{\mathrm{ML}}) \right] = 0. \tag{47}$$

In both equations, the bracketed term compares the actual counts of a variant, $n_{k\tau}$, to its counts predicted on the basis of the growth model, $n_\tau \phi_k(\boldsymbol{\lambda} t_\tau + \boldsymbol{\delta})$. This term vanishes when

$$\phi_k(\boldsymbol{\lambda}^{\mathrm{ML}} t_\tau + \boldsymbol{\delta}^{\mathrm{ML}}) = \frac{n_{k\tau}}{n_\tau}, \tag{48}$$

which is exactly the maximum-likelihood estimate (44) obtained by analyzing each time point independently. When the data are analyzed jointly across all time points, however, condition (48) need not hold at every t_τ. Instead, the requirement is that only the time-summed equalities in (47) are satisfied. In other words, joint estimation allows positive and negative deviations of the predicted and observed counts at individual time points to balance each other out across time.

The same type of balancing occurs also for the gradients of the mean-squared error in least-squares fitting. The key difference is that least-squares assumes Gaussian noise, with variances adjusted by explicitly assigning weights to the errors at different time points. By contrast, the conditions in (47) are derived directly from the multinomial sampling model, and the appropriate weighting emerges naturally rather than being imposed externally. In particular, in the second equality of (47) the deviations between observed and predicted counts are summed with equal weight across time points, while in the first equality they are weighted by time, thereby giving progressively greater emphasis to later observations.

Occasionally, the counts of certain variants may drop to zero at specific time points, either because the variants go extinct or because sequencing depth is insufficient to capture very small fractions. This occurs, for example, with variants 3 and 4 in our illustrative data. In the Bayesian analysis of the previous section, zeros are avoided by adding a constant number of pseudo-counts to the observed data. However, the maximum-likelihood estimation procedure does not include pseudo-counts. It is therefore instructive to examine the effect of zero counts on the optimality conditions (47).

When $n_{k\tau} = 0$ for variant k at time t_τ, the corresponding term in the square brackets of (47) contributes negatively to both gradient sums, since the *softmax* output is strictly positive. To satisfy the equalities, these negative contributions must be compensated by positive contributions from other time points where the predicted counts fall below the observed counts. In this way, the estimation procedure reconciles nonzero variant fractions with occasional zeros, without the need for pseudo-count adjustments.

For such cancellations to occur, however, at least one nonzero count of the variant is required. In fact, because two parameters are estimated per variant, at least two nonzero time points are necessary. If fewer than two are available, zero counts cannot be analyzed without the introduction of pseudo-counts.

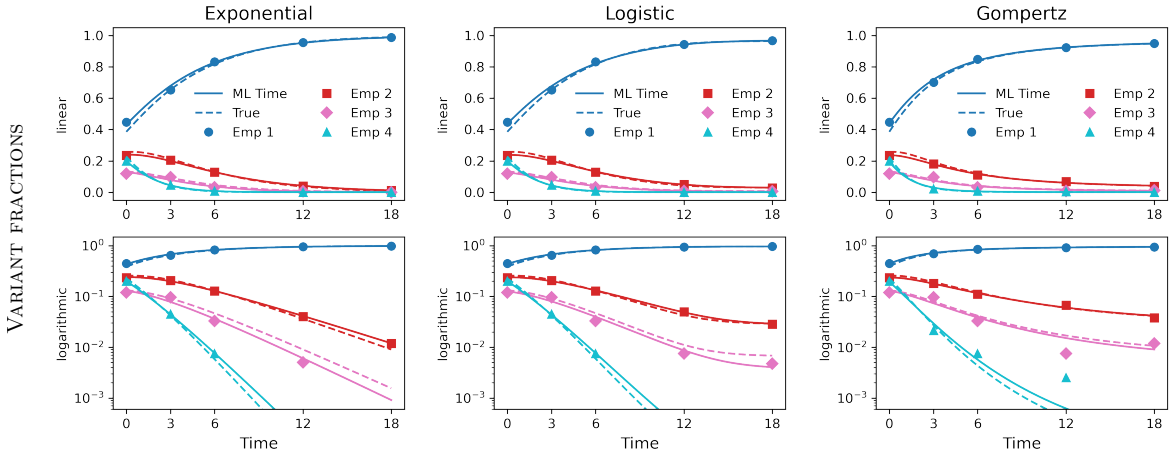

Figure 5: Time trajectories of variant fractions. Calculations based on the all-times maximum-likelihood (ML Time) estimates of the growth rates and initial log abundances are compared to the true (dashed lines) and empirical (symbols) fractions. For each growth model, the fractions are plotted on a linear (top) and logarithmic (bottom) vertical scale to better resolve both many and few counts.

Table 1: Estimates of the variant growth rates and their uncertainties for exponential growth. Counts at all five time points are used for $T = 5$, while only those at $t_0 = 0$ and $t_2 = 6$ are used for $T = 2$. (ML: maximum likelihood; VI: variational inference with ELBO; LB: analytical lower bound to the ELBO.)

	$\boldsymbol{\lambda}$ or $\boldsymbol{\mu}^\lambda$ (zero-sum gauge)	$\boldsymbol{\sigma}^\lambda$ (uncertainty)	
ML ($T = 5$)	$[0.297 \quad 0.084 \quad -0.025 \quad -0.356]$	-	fig. 5, left
ML ($T = 2$)	$[0.293 \quad 0.088 \quad -0.026 \quad -0.356]$	-	eq. (52)
VI ($T = 5$)	$[0.301 \quad 0.086 \quad -0.026 \quad -0.362]$	$[0.012 \quad 0.013 \quad 0.029 \quad 0.062]$	fig. 6, left
LB ($T = 5$)	$[0.300 \quad 0.087 \quad -0.025 \quad -0.362]$	$[0.002 \quad 0.012 \quad 0.028 \quad 0.061]$	fig. 6, right
LB ($T = 2$)	$[0.302 \quad 0.096 \quad -0.023 \quad -0.375]$	$[0.009 \quad 0.023 \quad 0.046 \quad 0.096]$	eq. (83)
True $\boldsymbol{\lambda}$ from (25)	$[0.319 \quad 0.080 \quad 0.019 \quad -0.418]$	-	

In our illustrative four-variant example, we maximized the all-time log likelihood (46) using gradient ascent. After each iteration step, the optimization parameters were adjusted to enforce the gauge conditions: the growth rates were constrained to sum to zero, and the exponential sum of the initial log abundances was normalized to one, corresponding to $N(0) = 1$. The final numerical values of the growth rates are shown in the first row of Table 1 (ML, $T = 5$).

To visualize how well the estimated parameters $\boldsymbol{\lambda}^{\mathrm{ML}}$ and $\boldsymbol{\delta}^{\mathrm{ML}}$ compare with the data, on the left side of fig. 5 we show the dynamical trajectories calculated from these estimates (solid lines). The curves closely track the true variant fractions used to generate the counts (dashed lines), while also aligning with the empirical fractions obtained at individual time points (symbols), which are the single-time maximum-likelihood estimates (22). As expected, the estimated fractions (solid lines) deviate from the true ones (dashed lines) in order to better match the observed fractions (symbols).

Although Gompertz growth is introduced in Sec. 6, we draw attention at this point to the identical analysis presented on the right side of fig. 5. The curves for variants 3 (cyan) and 4 (pink) in the lower-right panel nicely illustrate that counts with lower abundance—or zeros—contribute less than higher counts. This intuitive weighting emerges automatically from the maximum-likelihood framework, in contrast to least-squares fitting, where weights had to be imposed externally.

4.3 Two sequencing time points only

Given that many studies sequence only at the beginning and end of a competition experiment,[3] we now analyze the zero-gradient conditions in (47) for the case $T = 2$. Because of the multiplication by t_τ in the first equality, the initial time point ($t_0 = 0$) does not contribute to this sum. Hence, the condition in (48) must be satisfied entirely by the second time point. Once this holds, however, the second time point contributes nothing to the second equality in (47), since its bracketed term vanishes. Consequently, the first time point alone also satisfies (48). The case $T = 2$ therefore reduces to performing separate maximum-likelihood estimations at the two time points, without any integration of information across time.

Written explicitly, the separate conditions (48) at the first and second time points are

$$\frac{\mathrm{e}^{\delta_k^{\mathrm{ML}}}}{\sum_{j=1}^{K} \mathrm{e}^{\delta_j^{\mathrm{ML}}}} = \frac{n_{k0}}{n_0}, \qquad \frac{\mathrm{e}^{\lambda_k^{\mathrm{ML}} t_1 + \delta_k^{\mathrm{ML}}}}{\sum_{j=1}^{K} \mathrm{e}^{\lambda_j^{\mathrm{ML}} t_1 + \delta_j^{\mathrm{ML}}}} = \frac{n_{k1}}{n_1}. \tag{49}$$

The first equality does not contain the growth rates because they are multiplied by $t_0 = 0$. As a result, the initial log abundances are estimated directly from the initial counts (up to a global shift). With $\boldsymbol{\delta}^{\mathrm{ML}}$ determined from the first time point, the equality at the second time point then fixes the variant growth rates (up to another global shift). Thus, as noted earlier, for $T = 2$ the observed read counts are only just sufficient to determine the y-intercepts and the growth rates, leaving no redundancy for averaging over the counting noise. The only way to go beyond this bare minimum is to sequence at more than two points.

We now solve (49) for the maximum-likelihood estimates of the growth parameters. Since these parameters are identifiable only up to arbitrary global shifts, we discard additive contributions that are common to all variants, i.e., terms without the index k. From the first equality in (49), after removing variant-independent terms, we obtain

$$\delta_k^{\mathrm{ML}} = \log n_{k0}. \tag{50}$$

Applying the same reasoning to the second equality yields

$$\lambda_k^{\mathrm{ML}} t_1 + \delta_k^{\mathrm{ML}} = \log n_{k1}, \tag{51}$$

which leads to

$$\lambda_k^{\mathrm{ML}} = \frac{1}{t_1} \log \frac{n_{k1}}{n_{k0}}. \tag{52}$$

Thus, the maximum-likelihood estimate of the growth rate for variant k depends solely on the ratio of its counts at t_1 and t_0, irrespective of whether the total counts at these two time points differ.

Once obtained, the estimated growth rates and initial log abundances can be shifted collectively to any desired gauge.

As an example, we use (52) to calculate growth rates from the counts at the first and third time points of our synthetic data in (27). (The third time point is the last instance where the counts of all variants are non-zero.) The resulting numerical values, given in the second row of Table 1 (ML, $T = 2$), are practically identical to the maximum-likelihood estimates based on the counts at all five time points, shown above them.

Using the estimates in (50) and (52), one can switch to a gauge where the parameters of a chosen reference variant, ρ, are set to zero:

$$\delta_{k|\rho}^{\mathrm{ML}} = \log \frac{n_{k0}}{n_{\rho 0}}, \qquad \lambda_{k|\rho}^{\mathrm{ML}} = \frac{1}{t_1} \left[\log \frac{n_{k1}}{n_{k0}} - \log \frac{n_{\rho 1}}{n_{\rho 0}} \right] = \frac{1}{t_1} \left[\log \frac{n_{k1}}{n_{\rho 1}} - \log \frac{n_{k0}}{n_{\rho 0}} \right]. \tag{53}$$

The final expression for $\lambda_{k|\rho}$—given here in two equivalent forms—is the standard formula used widely to estimate relative growth rates from sequencing data at two time points. Our analysis demonstrates that this estimator actually yields the maximum-likelihood estimate of the relative growth rate for the case $T = 2$.

It is important to note that the alternative estimator in (52) involves only the counts of variant k, whereas the conventional estimator in (53) also depends on the counts of the reference variant. This distinction has direct implications for the statistical uncertainties of the two estimators, as we show in Sec. 5.2.2, following the introduction of Bayesian inference for the growth parameters.

5 Bayesian inference for the growth parameters

In the previous section, we reparametrized the multinomial distribution in terms of log abundances, yielding the single-time log likelihood (43). We then expressed the log abundances through the parameters of the exponential growth model, leading to the all-time log likelihood (46). These two likelihoods correspond to the two rightmost graphical models in the `Likelihood` row of fig. 2. In the present section, we upgrade the parameters of these probabilistic models to random variables, thereby obtaining the two `Bayesian` counterparts shown directly beneath them in fig. 2.

The Dirichlet distribution serves as a convenient conjugate prior for the multinomial distribution, yielding a Dirichlet posterior. In more general settings, however, no such conjugate prior exists in analytical form, as is precisely the case when the multinomial distribution is reparametrized in terms of log abundances. In these situations, the posterior can instead be approximated using variational inference.[9;22] The core idea is to restrict the prior and posterior distributions to a class of analytically convenient functions, and then select the parameters of the posterior so that it provides the best possible approximation to the exact but intractable posterior within that class.

In Sec. 5.1, we apply variational inference to counts at a single time point and compare the resulting approximate posterior with the exact Dirichlet posterior of Sec. 3.2.1. Since all variables in this setting correspond to the same time t_τ, we simplify the notation by omitting the index τ from all variables. After introducing variational inference in this single-time framework, we then extend it to the joint analysis of all sequencing time points in Sec. 5.2.

5.1 Variational inference at a single time point

The mathematical framework underlying variational inference[9;22] is presented in more detail in Sec. S3 of the SI. Here we focus on describing the final outcome and the assumptions on which it rests. Because the growth model is required only to connect estimates across different sequencing time points, it plays no role in the present subsection, which is limited to the analysis at a single time point.

As standard in Bayesian inference, the goal is to determine the posterior $p(\mathbf{y}|\mathbf{n})$, i.e., the distribution of log abundances given the observed counts. We approximate this intractable posterior within the class of distributions that factorize over the individual latent variables:

$$q(\mathbf{y}) = \prod_{k=1}^{K} q_k(y_k), \tag{54}$$

a choice known as the *mean-field approximation.*[17] Each factor $q_k(\cdot)$ is further taken to be Gaussian, with its own mean and variance:

$$y_k \sim \mathcal{N}(\mu_k, \sigma_k^2). \tag{55}$$

This yields the probabilistic model depicted in the middle of the `Bayesian` row of fig. 2, where the normal distributions in (55) are represented by the two red arrows labeled 'Normal'. Compared with the multinomial-Dirichlet model shown to its left, this model introduces twice as many free parameters that must be inferred from the data—the means $\boldsymbol{\mu}$ and standard deviations $\boldsymbol{\sigma}$.

As we show in Sec. S3 of the SI, for a given realization of the counts $\mathbf{n}$, the variational parameters $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ can be obtained by maximizing the function

$$\mathcal{F}(\boldsymbol{\mu}, \boldsymbol{\sigma}|\mathbf{n}) := \mathbb{E}_{\mathbf{y}\sim q(\mathbf{y}|\boldsymbol{\mu},\boldsymbol{\sigma})}[\mathcal{L}(\mathbf{y}|\mathbf{n})] + \sum_{k=1}^{K} \log \sigma_k. \tag{56}$$

The first term on the left-hand side is the expectation of the log likelihood over the latent variables $\mathbf{y}$, which are distributed according to the approximate posterior $q(\mathbf{y})$ parametrized by $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$, as indicated by the subscript of the expectation operator. The second summand in (56) corresponds to the cumulative entropy of the Gaussian factors forming the posterior.

In maximum-likelihood estimation, the log abundances are treated as parameters and maximizing $\mathcal{L}(\mathbf{y}|\mathbf{n})$ yields the point estimate $\mathbf{y}^{\mathrm{ML}}$. In variational inference, by contrast, the log abundances are random variables with distribution $q(\mathbf{y})$, and the parameters of this distribution are inferred by maximizing the objective (56). To see what this maximization achieves, consider what happens if the entropy term in (56) is ignored and only the expectation of the log likelihood is maximized. This can be attained by placing the means of the Gaussians at the maximum of the likelihood, i.e., $\boldsymbol{\mu} = \mathbf{y}^{\mathrm{ML}}$. Since $\mathcal{L}(\mathbf{y}|\mathbf{n})$ decreases away from its maximum, the expectation is going to be maximized when the Gaussians collapse to a single point, i.e., when $\boldsymbol{\sigma} \to 0$. Thus, the first term in (56) alone drives the approximate posterior towards the maximum-likelihood solution. The second term counteracts this collapse of $q(\mathbf{y})$ by penalizing the decrease of the standard deviations. If only this second term were maximized, on the other hand, the standard deviations would grow indefinitely, pushing the approximate posterior toward a uniform distribution. Maximizing $\mathcal{F}$ therefore strikes a balance between the uniform posterior and the maximum-likelihood solution.[b]

For reasons explained in Sec. S3.1 of the SI, the function $\mathcal{F}$ is often referred to as the ELBO, short for *evidence lower bound*. It can also be viewed as a free energy obtained by adding the entropy of the posterior to the "mean energy" given by the first summand in (56).[23] This interpretation motivates the choice of the letter $\mathcal{F}$.

When maximizing the ELBO either numerically or analytically, one must compute its gradients with respect to the variational parameters. To this end, it is convenient to express the normal random variables in (55) as

$$y_k = y_k(\mu_k, \sigma_k; \epsilon_k) = \mu_k + \sigma_k \epsilon_k, \tag{57}$$

where

$$\epsilon_k \sim \mathcal{N}(0, 1) \tag{58}$$

are *standard* normal variables with zero mean and unit variance. In this parametrization, the randomness of y_k is carried by the auxiliary variable ϵ_k, while the mean μ_k and standard deviation σ_k remain deterministic parameters, which can be handled directly by automatic differentiation. After this so called *reparametization trick*,[24;25] the expectation over $\mathbf{y}$ in (56) can be replaced by expectation over the standard normal variables:

$$\mathcal{F}(\boldsymbol{\mu}, \boldsymbol{\sigma}|\mathbf{n}) = \mathbb{E}_\epsilon[\mathcal{L}(\mathbf{y}(\boldsymbol{\mu}, \boldsymbol{\sigma}; \boldsymbol{\epsilon})|\mathbf{n})] + \sum_{k=1}^{K} \log \sigma_k. \tag{59}$$

So far, we have not used the functional form of the generative model. Substituting the log likelihood (43) of the multinomial-*softmax* distribution, and using $\mathbb{E}_\epsilon[\epsilon_k] = 0$, we finally obtain the ELBO

$$\mathcal{F}(\boldsymbol{\mu}, \boldsymbol{\sigma}|\mathbf{n}) = \sum_{k=1}^{K} n_k \mu_k - n\mathbb{E}_\epsilon \left[\log \left(\sum_{k=1}^{K} \mathrm{e}^{\mu_k + \sigma_k \epsilon_k} \right) \right] + \sum_{k=1}^{K} \log \sigma_k. \tag{60}$$

Maximizing this ELBO with respect to $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ for given experimental counts $\mathbf{n}$, yields estimates for the parameters of the middle `Bayesian` model in fig. 2.

Compared to the `Bayesian` model on its left, the current variational posterior requires twice as many parameters as the Dirichlet posterior. As a result, $2K$ parameters must be determined from only K observed counts. Because estimating more parameters than observations may appear problematic, we analyze how this is achieved by the optimization of the ELBO. To this end, we now examine the ELBO gradients with respect to the variational parameters.

[b] The ELBO in (56) is derived for a uniform prior, arising in the limit of infinite standard deviations. In this limit, the terms containing the prior parameters vanish and the ELBO reduces to (56). For a non-uniform prior, the ELBO includes additional additive contributions from the prior, which ensure that maximizing everything except the expectation term would drive the posterior toward this non-uniform prior. (See Secs. S3.2.2 and S3.2.3 of the SI).

5.1.1 ELBO gradients

Setting the partial derivatives of the ELBO (60) with respect to μ_k and σ_k to zero, we find that the following conditions need to hold at its maximum (the optimal values of the variational parameters are indicated by VI, for *variational inference*):

$$\mathbb{E}_\epsilon[\phi_k(\boldsymbol{\mu}^{\mathrm{VI}} + \boldsymbol{\sigma}^{\mathrm{VI}} \odot \boldsymbol{\epsilon})] = \frac{n_k}{n}, \qquad \sigma_k^{\mathrm{VI}}\, \mathbb{E}_\epsilon[\epsilon_k\, \phi_k(\boldsymbol{\mu}^{\mathrm{VI}} + \boldsymbol{\sigma}^{\mathrm{VI}} \odot \boldsymbol{\epsilon})] = \frac{1}{n}. \tag{61}$$

Here

$$\phi_k(\boldsymbol{\mu} + \boldsymbol{\sigma} \odot \boldsymbol{\epsilon}) = \frac{\mathrm{e}^{\mu_k + \epsilon_k \sigma_k}}{\sum_{j=1}^{K} \mathrm{e}^{\mu_j + \epsilon_j \sigma_j}} \tag{62}$$

is the kth component of the *softmax* function, whose output now is a random variable due to the presence of $\boldsymbol{\epsilon}$ in its argument. The gradients, however, are deterministic, since they involve expectations over $\boldsymbol{\epsilon}$.

The first equality in (61) is directly analogous to the (single-time) maximum-likelihood condition (44), requiring that the variant fraction predicted by the growth model equals the empirical fraction observed in the data. The difference here is that the growth parameters are random variables described by a probability distribution, so this requirement applies to the expectation of the predicted fraction over this distribution.

Now consider scaling the counts of all variants by a common factor, for example due to an increase in sequencing depth. This scaling cancels in the right-hand side of the first equality in (61). The left-hand side must therefore remain invariant as well, implying that this equality carries no information about the absolute magnitude of the counts.

By contrast, the right-hand side of the second equality in (61) scales inversely with the total number of reads. Increasing all counts by a common factor thus forces the left-hand side to shrink proportionally. This in turn drives the inferred standard deviations to decrease, though not necessarily by the same factor, since the expectation over $\boldsymbol{\epsilon}$ also depends on $\boldsymbol{\sigma}^{\mathrm{VI}}$. In this way, the absolute sequencing depth is encoded into the variational standard deviations. Consequently, the K parameters $\boldsymbol{\sigma}$ play a role analogous to the single precision parameter α of the Dirichlet posterior discussed in Sec. 3.2.1.

A related observation is that, in the optimality conditions (61), the means $\boldsymbol{\mu}$ always appear within the argument of the *softmax* function. As a result, their values can be shifted globally without affecting the outcome. By contrast, the standard deviations $\boldsymbol{\sigma}$ also occur outside the *softmax*, which prevents such global shifts and reflects their role in encoding the absolute scale of the counts. Accordingly, when maximizing the ELBO numerically, a gauge must be imposed on the means but not on the standard deviations.

This analysis clarifies the mechanism by which ELBO maximization extracts $2K$ variational parameters from K observations: the means encode relative proportions of the counts, while the standard deviations capture information about the absolute count magnitudes. Because the expectations over the standard normal variables cannot be computed in closed form, however, the optimality conditions (61) cannot be solved analytically for the variational parameters. In Sec. 5.1.3, we address this limitation by replacing the expectation term in (60) with an analytical upper bound, thereby obtaining a tractable lower bound to the ELBO that can be maximized in closed form. Before turning to this bound, we next discuss a numerical strategy for approximating the exact ELBO.

5.1.2 Stochastic calculation of the ELBO

From the perspective of sampling, where one generates samples of the random variables, the expectation term in the ELBO corresponds to the average over infinitely many independent random realizations. Specifically, if we denote one random realization of the K standard normal variables by $\boldsymbol{\epsilon}^{[m]}$, then

$$\mathbb{E}_\epsilon\left[\log\left(\sum_{k=1}^{K} \mathrm{e}^{\mu_k + \sigma_k \epsilon_k}\right)\right] = \lim_{M\to\infty} \frac{1}{M} \sum_{m=1}^{M} \log\left(\sum_{k=1}^{K} \mathrm{e}^{\mu_k + \sigma_k \epsilon_k^{[m]}}\right). \tag{63}$$

By calculating the sample average over a finite number of M realizations, we thus obtain an approximation to the exact expectation. Because the sample mean over another batch of M random realizations will be

Table 2: Variational estimates of the initial log abundances and their uncertainties from the counts at t_0.

	$\boldsymbol{\mu} = \mathbb{E}[\boldsymbol{\delta}]$	$\boldsymbol{\sigma} = \mathrm{Std}[\boldsymbol{\delta}]$
Stochastic ($M = 10^6$)	$\begin{bmatrix}-0.804 & -1.447 & -2.135 & -1.616\end{bmatrix}$	$\begin{bmatrix}0.099 & 0.116 & 0.152 & 0.124\end{bmatrix}$
Lower bound	$\begin{bmatrix}-0.804 & -1.447 & -2.135 & -1.616\end{bmatrix}$	$\begin{bmatrix}0.074 & 0.102 & 0.143 & 0.110\end{bmatrix}$
True $\boldsymbol{\delta}$ from (25)	$\begin{bmatrix}-0.955 & -1.363 & -2.026 & -1.481\end{bmatrix}$	-

different, in general, the obtained approximate value will change stochastically every time it is calculated. As a result, during the numerical maximization of the ELBO, we actually work with its stochastic approximation

$$\mathcal{F}_M(\boldsymbol{\mu}, \boldsymbol{\sigma}|\mathbf{n}) = \sum_{k=1}^{K} n_k \mu_k - n \frac{1}{M} \sum_{m=1}^{M} \log \left(\sum_{k=1}^{K} \mathrm{e}^{\mu_k + \sigma_k \epsilon_k^{[m]}} \right) + \sum_{k=1}^{K} \log \sigma_k. \tag{64}$$

Naturally, the gradients of this stochastic $\mathcal{F}_M$ with respect to the parameters $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$, which are accumulated automatically during the numerical integration of the dynamics, will also be stochastic approximations to the true gradients, and so will continue to fluctuate even at the optimum. This stochasticity of the ELBO and its gradients complicates the assessment of convergence of the maximization. However, the stochastic approximation can be made increasingly more accurate by increasing the number of stochastic samples, which however increases the computational time.

We now apply the variational inference strategy to the counts in the first column of $\mathbf{N}$ in (27), corresponding to $t_0 = 0$. At this initial time, the log abundances are given by $\boldsymbol{\delta}$, so the variational parameters $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ represent the means and standard deviations of $\boldsymbol{\delta}$, inferred from the initial counts only. During the numerical optimization, after each iteration step we shift the means such that their exponential sum over all variants equals one, consistent with $N(0) = 1$, as in the maximum-likelihood estimation of Sec. 4.2.

The values of $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ obtained by maximizing the ELBO with $M = 10^6$ random realizations of $\boldsymbol{\epsilon}$ are reported in the first row of Table 2. This relatively large sample size ensures an accurate approximation to the exact ELBO: indeed, identical numerical values—within the three displayed decimal places—are reproduced when the optimization is repeated with a different random seed. (The effect of reducing the sample size to $M = 100$ is examined in Sec. S3.3.2 of the SI.)

For comparison, the last row of Table 2 lists the true values of the initial log abundances from (25). Since the variational estimates are based on a single realization of the counts, without any averaging over the stochasticity of the counting noise, close agreement between the estimated means and the true values is not expected. The discrepancies, however, should be captured by the estimated standard deviations. Indeed, the true values fall within one standard deviation of the estimated means for all but the first variant, which lies within two standard deviations.

5.1.3 Analytical lower bound to the ELBO

The Jensen inequality—stating that the expectation of a convex (concave) function is greater (smaller) than the function of the expectation—provides a useful tool for deriving analytical bounds in variational inference.[17]

Applying Jensen's inequality to the logarithm inside the expectation operator in (60) gives

$$\mathbb{E}_\epsilon \left[\log \left(\sum_{k=1}^{K} \mathrm{e}^{\mu_k + \sigma_k \epsilon_k} \right) \right] \leq \log \left(\sum_{k=1}^{K} \mathbb{E}_\epsilon \left[\mathrm{e}^{\mu_k + \sigma_k \epsilon_k} \right] \right). \tag{65}$$

Bringing the expectation inside the logarithm makes it possible to compute the resulting expectations over the independent standard normal variables analytically:

$$\mathbb{E}_\epsilon \left[\mathrm{e}^{\mu_k + \sigma_k \epsilon_k} \right] = \mathrm{e}^{\mu_k + \frac{1}{2}\sigma_k^2}. \tag{66}$$

Replacing the second term of the ELBO in (60) with its upper bound from the right-hand side of (65) yields

$$\mathcal{F}_{\mathrm{LB}}(\boldsymbol{\mu}, \boldsymbol{\sigma}|\mathbf{n}) = \sum_{k=1}^{K} n_k \mu_k - n \log \left(\sum_{k=1}^{K} \mathrm{e}^{\mu_k + \frac{1}{2}\sigma_k^2} \right) + \sum_{k=1}^{K} \log \sigma_k. \tag{67}$$

Since $\mathcal{F} \geq \mathcal{F}_{\mathrm{LB}}$, this expression provides a lower bound to the ELBO. It thus offers an analytically tractable surrogate objective for inferring the variational parameters by maximizing $\mathcal{F}_{\mathrm{LB}}$, instead of $\mathcal{F}$. Although this bound is not a controlled approximation to the ELBO, it has the advantage that its maximum can be obtained in closed form.

Setting the gradients of $\mathcal{F}_{\mathrm{LB}}$ to zero yileds

$$\phi_k^* := \phi_k \left(\boldsymbol{\mu}^* + \frac{1}{2}\boldsymbol{\sigma}^{*2} \right) = \frac{n_k}{n}, \qquad (\sigma_k^*)^2 = \frac{1}{n\phi_k^*} = \frac{1}{n_k}, \tag{68}$$

where parameter values at the optimum of $\mathcal{F}_{\mathrm{LB}}$ are denoted by $*$. For each variant, the posterior variance is simply the reciprocal of its counts: the more reads a variant has at a given time point, the smaller the inferred uncertainty of its log abundance is.

Having determined the standard deviations, we can solve the first relation in (68) for the posterior means. Since the argument of the *softmax* function is identifiable only up to a global shift, we discard variant-independent additive terms and obtain

$$\mu_k^* = \log n_k - \frac{1}{2}(\sigma_k^*)^2 = \log n_k - \frac{1}{2n_k}. \tag{69}$$

For the initial counts of our illustrative dataset in (27), we evaluated the posterior means in (69) and the corresponding standard deviations in the last equality of (68). The results are shown in the second row of Table 2, after shifting the means so that their exponential sum equals one. To the reported precision, the means coincide with those obtained by maximizing the exact ELBO, shown above them. The standard deviations from the lower bound, however, are systematically smaller, thus underestimating the uncertainties in the inferred log abundances.

The means and standard deviations in Table 2 are for the log abundances, whereas the Dirichlet analysis in Sec. 3.2.1 provides the corresponding statistics for the variant fractions. A direct comparison between the two approaches can be made via the identity (14), for whose left-hand side the means and variances under the Dirichlet posterior are given in (34). We now write the means and variances of the right-hand side of (14) in terms of the variational parameters $\boldsymbol{\mu}^*$ and $\boldsymbol{\sigma}^*$, and compare the two sets of predictions.

Using (69) and the second equality in (68), we find

$$\begin{aligned} \mathbb{E}_{\mathrm{LB}}[y_k - y_\rho] &= \mu_k^* - \mu_\rho^* = \left(\log n_k - \frac{1}{2n_k} \right) - \left(\log n_\rho - \frac{1}{2n_\rho} \right) \\ \mathrm{Var}_{\mathrm{LB}}[y_k - y_\rho] &= \sigma_k^{*2} + \sigma_\rho^{*2} = \frac{1}{n_k} + \frac{1}{n_\rho}, \end{aligned} \tag{70}$$

with $k \neq \rho$ for the variance. (The subscript LB is a reminder that the expectation and variance are calculated using the posterior obtained by maximizing the lower bound to the ELBO.) Note that, under the mean-field assumption of independence among the log abundances, the variance of the difference equals the sum of the individual variances. As a result, taking a difference relative to a reference variant amplifies the uncertainty, producing estimates that are less precise than those of the individual log abundances.

Since the results in (70) do not contain any pseudo-counts, we compare them to the Dirichlet expressions with $\tilde{\alpha} = 0$. The variance in (70) agrees with the one in (34) if the asymptotic expansion of $\psi_1(\cdot)$ is terminated after the first term, while the expectation in (70) reproduces two terms in the asymptotic expansion of $\psi_0(\cdot)$. Thus, in the regime of large counts, where terminating the asymptotic expansions yields excellent approximations, minimizing the analytical lower bound to the ELBO essentially recapitulates the Dirichlet

analysis with zero pseudo-counts. The variances inferred by maximizing the actual ELBO are expected to be larger.

In summary, the analytical and numerical results of this subsection show that variational inference at a single time point offers a practical alternative to the exact Bayesian analysis with a Dirichlet posterior, despite the need to infer twice as many parameters from the data. Next, we extend variational inference to the joint analysis of the counts at all time points.

5.2 Variational inference across all time points

Variational inference was used in the previous section to transform the second SINGLE TIME - `Likelihood` model in fig. 2 to the `Bayesian` model below it. Having completed this groundwork, we now turn to the ALL TIMES - `Likelihood` model in fig. 2, which is parameterized directly in terms of growth rates and initial log abundances. Promoting these parameters to random variables, and approximating the posterior once again by a product of independent normal distributions, yields the ALL TIMES–`Bayesian` model shown beneath it.

The parameters of this new model are inferred from the data by maximizing the ELBO

$$\mathcal{F}(\boldsymbol{\mu}^\lambda, \boldsymbol{\sigma}^\lambda, \boldsymbol{\mu}^\delta, \boldsymbol{\sigma}^\delta|\mathbf{N}) = \sum_{\tau=0}^{T-1}\left\{\sum_{k=1}^{K} n_{k\tau}\mathbb{E}_\epsilon[y_{k\tau}] - n_\tau \mathbb{E}_\epsilon\left[\log\left(\sum_{k=1}^{K} \mathrm{e}^{y_{k\tau}}\right)\right]\right\} + \sum_{k=1}^{K}(\log\sigma_k^\lambda + \log\sigma_k^\delta). \tag{71}$$

The dependence of this ELBO on the variational parameters is through the log abundances $y_{k\tau}$, which also depend on the standard normal variables ϵ_k^λ and ϵ_k^δ. The operator $\mathbb{E}_\epsilon[\cdot]$ takes expectation over $\boldsymbol{\epsilon}^\lambda$ and $\boldsymbol{\epsilon}^\delta$. As before, the final term represents the entropy of the posterior, now receiving contributions from both the growth rates and the initial log abundances, as indicated by the superscripts. The first term is the "mean energy," which now includes a sum over all time points. This ELBO does not yet incorporate the model of growth, and so can be combined with any growth dynamics, including the saturating growth in Sec. 6.

In the case of exponential growth, the log abundances take the form [from eqs. (10) and (57)]

$$y_{k\tau} = \lambda_k t_\tau + \delta_k = (\mu_k^\lambda + \sigma_k^\lambda \epsilon_k^\lambda)t_\tau + (\mu_k^\delta + \sigma_k^\delta \epsilon_k^\delta). \tag{72}$$

The first expectation in the ELBO can be evaluated analytically, since $\mathbb{E}_\epsilon[\epsilon_k^\lambda] = \mathbb{E}_\epsilon[\epsilon_k^\delta] = 0$:

$$\mathcal{F}(\boldsymbol{\mu}^\lambda, \boldsymbol{\sigma}^\lambda, \boldsymbol{\mu}^\delta, \boldsymbol{\sigma}^\delta|\mathbf{N}) = \sum_{\tau=0}^{T-1}\sum_{k=1}^{K} n_{k\tau}(\mu_k^\lambda t_\tau + \mu_k^\delta) - \sum_{\tau=0}^{T-1} n_\tau \mathbb{E}_\epsilon\left[\log\left(\sum_{k=1}^{K} \mathrm{e}^{y_{k\tau}}\right)\right] + \sum_{k=1}^{K}(\log\sigma_k^\lambda + \log\sigma_k^\delta). \tag{73}$$

Now only the second term retains an expectation operator, mirroring the structure of the single-time ELBO in (60).

As before, this expectation can be approximated stochastically by drawing M samples of $(\boldsymbol{\epsilon}^\lambda, \boldsymbol{\epsilon}^\delta)$ and taking the empirical mean. Alternatively, one may invoke Jensen's inequality to replace the expectation term with an analytical upper bound, thereby obtaining a fully analytical lower bound on the ELBO, which can be maximized instead. We pursue this second approach in Sec. 5.2.1, after first evaluating the ELBO numerically for our illustrative example with four variants.

We calculated the ELBO (73) stochastically, by generating $M = 10^4$ random realizations per iteration step. The gradients of the ELBO were computed via automatic differentiation. At the end of each gradient update step, the means $\boldsymbol{\mu}^\lambda$ were shifted such that they summed to zero and the means $\boldsymbol{\mu}^\delta$ such that their exponential sum equaled one, thus $N(0) = 1$.

To accelerate convergence toward the optimum, we initialized the variational means with the maximum-likelihood estimates $\boldsymbol{\lambda}^{\mathrm{ML}}$ and $\boldsymbol{\delta}^{\mathrm{ML}}$. As no analogous initial estimates were available for the standard deviations, we first performed 200 optimization steps keeping the means fixed at the maximum-likelihood values and updating only the standard deviations. This preliminary phase allowed the standard deviations to relax to reasonable values without disrupting the means in the process. Subsequently, all variational parameters were optimized jointly for an additional 1000 steps. Convergence was assessed visually by verifying that the initially decreasing ELBO stabilized into stochastic fluctuations around a steady baseline (see fig. S4

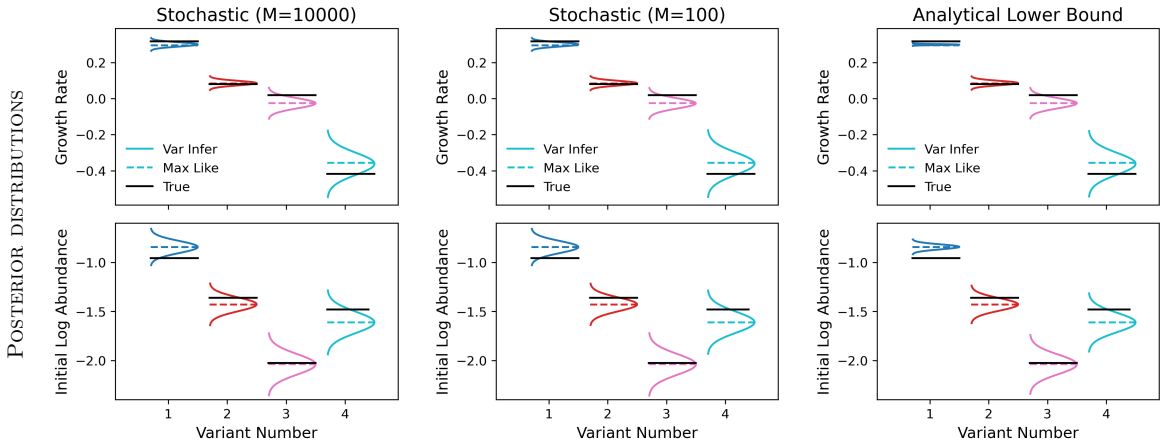

Figure 6: Gaussian posteriors from stochastic and analytical approximations to the ELBO, shown in the interval $\mu_k \pm 3\sigma_k$. True values of the growth parameters are indicated with black lines and the maximum-likelihood estimates (Max Like) with dashed lines. Decreasing the number of random realizations in the stochastic approximation, M, does not change the results. Minimizing the analytical lower bound to the ELBO underestimates the uncertainties, especially for the most abundant variant 1.

in the SI). Final parameter estimates were obtained by averaging the fluctuating values over the last 100 optimization steps.

The inferred growth-rate means and standard deviations are reported in the third row of Table 1 (VI, $T = 5$). To visualize the inferred variational parameters, on the left side of fig. 6 we show the corresponding Gaussian distributions over the intervals $\mu_k \pm 3\sigma_k$, representing three standard deviations around each mean (growth rates are on top and initial log-abundances below). Alongside the variational posterior distributions (solid colored lines), we include the maximum-likelihood estimates from Sec. 4.2 (dashed colored lines) and the true parameter values used to generate the synthetic data (solid black lines).

Visually, the uncertainties of the more abundant variants 1 and 2 are comparable and smaller than those of the less abundant variants 3 and 4, although the difference is more pronounced for the growth rates than for the initial abundances. Additionally, the growth-rate uncertainty of variant 4, which has zero counts at the last two time points, is substantially larger than that of variant 3, which is zero only at the last time point. [See also the numerical values in Table 1 (VI, $T = 5$).] These results are in line with what one would expect on the basis of the counts.

Given that a sample size of $M = 10^4$ may become prohibitively large for thousands of variants, we explored the possibility of reducing this number. The middle panels in fig. 6 demonstrate that approximating the expectation with as little as $M = 100$ random samples produces identical results. We use this smaller number in the subsequent numerical examples.

5.2.1 Analytical lower bound to the ELBO

For further analytical insight, we now analyze the lower bound to the ELBO obtained via Jensen's inequality.

Exchanging the order of expectation and logarithm in the second term in (73), as prescribed by the inequality, introduces expectations of the form $\mathbb{E}_\epsilon[e^{y_{k\tau}}]$. These can be evaluated in closed form for exponential growth with $y_{k\tau}$ given in (72):

$$\mathbb{E}_\epsilon\left[\mathrm{e}^{y_{k\tau}}\right] = \exp\left\{\mu_k^\lambda t_\tau + \frac{1}{2}(\sigma_k^\lambda)^2 t_\tau^2 + \mu_k^\delta + \frac{1}{2}(\sigma_k^\delta)^2\right\}. \tag{74}$$

Substituting this result back into (73), we obtain the analytical lower bound

$$\mathcal{F}_{\mathrm{LB}} = \sum_{\tau=0}^{T-1} \left[\sum_{k=1}^{K} n_{k\tau}(\mu_k^\lambda t_\tau + \mu_k^\delta) - n_\tau \log \left(\sum_{k=1}^{K} \mathrm{e}^{\mu_k^\lambda t_\tau + \mu_k^\delta + \frac{1}{2}(\sigma_k^\lambda)^2 t_\tau^2 + \frac{1}{2}(\sigma_k^\delta)^2} \right) \right] + \sum_{k=1}^{K} (\log \sigma_k^\lambda + \log \sigma_k^\delta), \tag{75}$$

which is to be maximized as a surrogate to the ELBO.

By setting the partial derivatives of $\mathcal{F}_{\mathrm{LB}}$ with respect to the means to zero, we obtain the optimality conditions

$$\sum_{\tau=0}^{T-1} t_\tau (n_{k\tau} - n_\tau \phi_{k\tau}^*) = 0, \qquad \sum_{\tau=0}^{T-1} (n_{k\tau} - n_\tau \phi_{k\tau}^*) = 0, \tag{76}$$

where

$$\phi_{k\tau}^* := \phi_k \left(\boldsymbol{\mu}^{\lambda *} t_\tau + \boldsymbol{\mu}^{\delta *} + \frac{1}{2}(\boldsymbol{\sigma}^{\lambda *})^2 t_\tau^2 + \frac{1}{2}(\boldsymbol{\sigma}^{\delta *})^2 \right) \tag{77}$$

and the superscript $*$ denotes values at the optimum of $\mathcal{F}_{\mathrm{LB}}$. Although these gradient conditions take the same form as those in (47), derived from the all-time log likelihood, the argument of the *softmax* function here includes means and variances.

Setting the gradients of $\mathcal{F}_{\mathrm{LB}}$ with respect to the standard deviations to zero yields

$$\frac{1}{(\sigma_k^{\lambda *})^2} = \sum_{\tau=0}^{T-1} t_\tau^2 n_\tau \phi_{k\tau}^*, \qquad \frac{1}{(\sigma_k^{\delta *})^2} = \sum_{\tau=0}^{T-1} n_\tau \phi_{k\tau}^* = \sum_{\tau=0}^{T-1} n_{k\tau}, \tag{78}$$

where the final equality follows from the second condition in (76). According to it, the inverse variance (also known as *precision*) of the initial log abundance δ_k, is obtained by adding the observed counts of that variant across all time points.

The precision of the growth rate λ_k in (78) depends on $\phi_{k\tau}^*$, which is yet to be evaluated. Nonetheless, we recognize that $n_\tau \phi_{k\tau}^*$ corresponds to the predicted counts of variant k at t_τ. Hence, the precision equals the weighted sum of these predicted counts across all times, giving more weight to the counts at later times and completely disregarding the initial counts at $t_0 = 0$.

While the proportionality between precision and number of counts is characteristic for counting processes like Poisson and multinomial, the precisions in (78) are directly for the growth parameters, and show how to aggregate the counts across time points. Admittedly, these analytical results are for the lower bound to the ELBO, and are expected to be quantitatively different from the posterior variances obtained by maximizing the ELBO itself.

To assess this difference for the illustrative example with four variants, we maximized $\mathcal{F}_{\mathrm{LB}}$ numerically using gradient ascent. The resulting posterior distributions are shown on the right of fig. 6. For the most-abundant first variant, the standard deviations obtained by maximizing the lower bound are substantially narrower that those obtained by maximizing the ELBO. This underestimation of the statistical uncertainty is in line with the single-time result for $t_0 = 0$ in Table 2. For the other three variants in fig. 6, however, the underestimation is visually undetectable. These observations are confirmed by the numerical values of the growth-rate means and standard deviations in the fourth row of Table 1 (LB, $T = 5$).

This example demonstrates that the uncertainty of highly abundant variants may be significantly underestimated when inferred from the lower bound to the ELBO. Keeping this limitation in mind, we further pursue the maximization of $\mathcal{F}_{\mathrm{LB}}$ in (75) for the special case of only to sequencing time points.

5.2.2 Sequencing at only two time points

Equations (76) and (78) admit further simplification in the case of only two sequencing time points, analogously to the situation in Sec. 4.3. Following the same logic as there, we find that for $T = 2$

$$\phi_{k0}^* = \frac{n_{k0}}{n_0}, \qquad \phi_{k1}^* = \frac{n_{k1}}{n_1}, \qquad (\sigma_k^{\lambda *})^2 t_1^2 = \frac{1}{n_1 \phi_{k1}^*} = \frac{1}{n_{k1}}, \qquad (\sigma_k^{\delta *})^2 = \frac{1}{n_{k0} + n_{k1}}. \tag{79}$$

As already observed for general T, the precision of δ_k is given by the sum of the counts of this variant at the two time points. In contrast, the precision of λ_k is proportional to the counts at the last time point only.

Using the definition of the *softmax* function, the first two equalities in (79) become

$$\frac{\mathrm{e}^{\mu_k^{\delta*}+\frac{1}{2}(\sigma_k^{\delta*})^2}}{\sum_{j=1}^K \mathrm{e}^{\mu_j^{\delta*}+\frac{1}{2}(\sigma_j^{\delta*})^2}} = \frac{n_{k0}}{n_0}, \qquad \frac{\mathrm{e}^{\mu_k^{\lambda} t_1+\mu_k^{\delta}+\frac{1}{2}(\sigma_k^{\lambda*})^2 t_1^2+\frac{1}{2}(\sigma_k^{\delta*})^2}}{\sum_{j=1}^K \mathrm{e}^{\mu_j^{\lambda} t_1+\mu_j^{\delta}+\frac{1}{2}(\sigma_j^{\lambda*})^2 t_1^2+\frac{1}{2}(\sigma_j^{\delta*})^2}} = \frac{n_{k1}}{n_1}. \tag{80}$$

When solving these for $\mu_k^{\lambda*}$ and $\mu_k^{\delta*}$, we proceed as in Sec. 4.3 and drop additive terms that are common to all variants. Taking the logarithm of the first equality, we obtain

$$\mu_k^{\delta*} = \log n_{k0} - \frac{1}{2}(\sigma_k^{\delta*})^2 = \log n_{k0} - \frac{1}{2(n_{k0}+n_{k1})}. \tag{81}$$

Compared to the maximum-likelihood estimate δ^{ML} in (50), the last term is new. It reflects the entropic effect of the variance on the mean.[c] Due the presence of the counts at t_1, this result is also different from the variational mean in (69) calculated at a single time point.

Similarly, taking the logarithm of the second equality in (80) yields

$$\mu_k^{\lambda*} = \frac{1}{t_1}\left[\log n_{k1} - \mu_k^{\delta} - \frac{1}{2}(\sigma_k^{\lambda*})^2 t_1^2 - \frac{1}{2}(\sigma_k^{\delta*})^2\right] = \frac{1}{t_1}\left[\log\frac{n_{k1}}{n_{k0}} - \frac{1}{2n_{k1}}\right]. \tag{82}$$

Again, there is an entropic correction to the maximum-likelihood estimate λ^{ML} in (52). The correction ensures that, for normally distributed λ_k, the mean of $\mathrm{e}^{\lambda_k t_1}$ equals n_{k1}/n_{k0}.

Combining the above mean and standard deviation of λ_k, we give our variational estimate of the growth rates and their uncertainties from sequencing at only two time points:

$$\mu_k^{\lambda*} \pm \sigma_k^{\lambda*} = \frac{1}{t_1}\left[\log\frac{n_{k1}}{n_{k0}} - \frac{1}{2n_{k1}} \pm \sqrt{\frac{1}{n_{k1}}}\right]. \tag{83}$$

Because this estimate is inferred from the lower bound to the ELBO, the estimated uncertainty is expected to be smaller than the one from maximizing the actual ELBO, especially for variants with high counts. The analogous result for the initial log abundances is expected to be of less practical interest.[d]

Recall that, once estimated, the means $\boldsymbol{\mu}^{\lambda}$ and $\boldsymbol{\mu}^{\delta}$ can be shifted collectively to any desired gauge.

In the fifth row of Table 1 (LB, $T = 2$) we show numerical values calculated according to (83) from the counts at only the first and third time points of the dataset in (27). It is interesting to observe how the standard deviations estimated from two time points are larger than those estimated from all five time points, shown one line above in Table 1 (LB, $T = 5$). It is hard to assess, however, whether the entropic reduction in the estimated growth rate of variant 4 is a significant improvement over the maximum-likelihood estimate in (52), shown in the second row of Table 1 (ML, $T = 2$).

Using the above expressions for $\mu_k^{\lambda*}$ and $\sigma_k^{\lambda*}$, we also calculate the means and standard deviations that correspond to the standard approach of estimating growth rates relative to a reference variant:

$$\begin{aligned}\mathbb{E}_{\mathrm{LB}}[\lambda_k - \lambda_\rho] &= \mu_k^{\lambda*} - \mu_\rho^{\lambda*} = \frac{1}{t_1}\left[\log\frac{n_{k1}}{n_{k0}} - \log\frac{n_{\rho 1}}{n_{\rho 0}} - \frac{1}{2n_{k1}} + \frac{1}{2n_{\rho 1}}\right] \\ \mathrm{Std}_{\mathrm{LB}}[\lambda_k - \lambda_\rho] &= \sqrt{(\sigma_k^{\lambda*})^2 + (\sigma_\rho^{\lambda*})^2} = \frac{1}{t_1}\sqrt{\frac{1}{n_{k1}} + \frac{1}{n_{\rho 1}}}.\end{aligned} \tag{84}$$

[c]For normally distributed log abundance δ_k, the initial abundance e^{δ_k} is log-normal. In terms of the mean and standard deviation of δ_k, the mean of e^{δ_k} is $\exp\{\mu_k^{\delta*} + \frac{1}{2}(\sigma_k^{\delta*})^2\}$, from (66). Note how the new term on the right-hand side of (81) exactly cancels the contribution from the variance, to make the mean of e^{δ_k} equal to n_{k0}.

[d]In this case, the counts at both time points contribute to the standard deviation and the entropic shift of the mean:

$$\mu_k^{\delta*} \pm \sigma_k^{\delta*} = \log n_{k0} - \frac{1}{2(n_{k0}+n_{k1})} \pm \sqrt{\frac{1}{n_{k0}+n_{k1}}}.$$

The expectation in the first line of (84) contains entropic corrections to the standard formula for growth-rate estimation, given in (53). The new terms are expected to be numerically important in the case of very few counts. The uncertainty of this corrected standard estimate is quantified by the standard deviation in the second line in (84). This uncertainty is larger than the uncertainty of our alternative estimator in (83), because it additionally contains the counts of the reference, as noted earlier at the end of Sec. 4.3. (Both of these standard deviations are expected to underestimate the uncertainty compared to maximizing the actual ELBO.)

This exhausts our analysis of exponential growth. In the next section, we apply the joint analysis of all sequencing time points—both in its maximum-likelihood and variational-inference versions—to non-exponential growth.

6 Beyond exponential growth: models with saturation

From a mathematical point of view, exponential growth is attractive because it admits simple linear solution. Similarly, linear least-squares regression is attractive because it can be solved by single matrix inversion. Such analytical tractability, however, is lost in all approaches for analyzing exponential growth that we touched on above—weighted least-squares fit to a *softmax* function, maximization of the all-time likelihood of a multinomial-*softmax* model of counting noise, and maximization of its ELBO. In these three problems, a certain scalar function had to be optimized numerically with respect to either K or $2K$ (in the case of ELBO) parameters. Very conveniently, however, the gradients of the loss function could be computed through automatic differentiation.[16] In fact, the entire gradient-update procedure is implemented with a few lines of code in modern packages like PyTorch.[26]

Once numerical gradient calculation becomes part of the estimation process, the analytical integrability of exponential growth loses its central importance. The non-linear trajectories of alternative growth models can be integrated numerically, while automatic differentiation efficiently accumulates the gradients of the loss function with respect to the model parameters. Leveraging this flexibility, in this section we apply both maximum-likelihood (Sec. 4.2) and variational-inference (Sec. 5.2) approaches to growth models with saturation. Although the strategy is illustrated for logistic and Gompertz growth, it extends readily to other growth models with only a few parameters per variant.

Handling non-exponential growth is straightforward. For the likelihood (45), we previously substituted the linear dependence of the log abundances to arrive at (46). Now we use the general growth dynamics

$$y_{k\tau} = y_k(\boldsymbol{\lambda}, \boldsymbol{\delta}; t_\tau), \tag{85}$$

which are calculated numerically by integrating the differential equations of the selected growth model. Similarly, in the ELBO (71) we now use

$$y_{k\tau} = y_k(\boldsymbol{\lambda}(\boldsymbol{\mu}^\lambda, \boldsymbol{\sigma}^\lambda; \boldsymbol{\epsilon}^\lambda), \boldsymbol{\delta}(\boldsymbol{\mu}^\delta, \boldsymbol{\sigma}^\delta; \boldsymbol{\epsilon}^\delta); t_\tau), \tag{86}$$

rather than the linear expression in (72). Everything else remains the same.

6.1 Growth dynamics in terms of variant log abundances

Here we derive the right-hand side of (85) for logistic and Gompertz growth. In these two models, the cell numbers of the variants obey the differential equations

$$\frac{\mathrm{d}N_k(t)}{\mathrm{d}t} = \lambda_k N_k(t)\, s_{\mathcal{K}}(N(t)/\mathcal{K}), \tag{87}$$

where the saturation factor $s_{\mathcal{K}}(\cdot)$ compares the total number of cells to the carrying capacity $\mathcal{K}$. The functional form of this factor for the three growth models that we consider is

$$s_{\mathcal{K}}(x) = \begin{cases} 1, & \text{exponential} \\ 1 - x, & \text{logistic} \\ -\log x, & \text{Gompertz.} \end{cases} \tag{88}$$

In logistic and Gompertz growth, the saturation factor becomes zero when its argument equals one, which happens when the total number of cells equals the carrying capacity. Away from saturation ($x \ll 1$), this factor is less than one for logistic growth, but can be arbitrarily large for Gompertz growth, thus making the instantaneous growth rate in (87) much larger than the numerical value of the parameter λ_k.[e]

When the total number of cells is constrained by the carrying capacity, the dynamics of all variants are coupled. In contrast, each variant is independent of the others in exponential growth, where there is no saturation.

From the differential equations of the cell numbers, we obtain similar equations for the log abundances:

$$\frac{\mathrm{d}y_k(t)}{\mathrm{d}t} = \lambda_k \, s_{\mathcal{K}} \left(\sum_{j=1}^{K} \mathrm{e}^{y_j(t)-\kappa} \right), \tag{89}$$

with

$$\kappa := \log \mathcal{K}. \tag{90}$$

Integrating both sides of (89) from the initial time to the current time yields

$$y_k(\boldsymbol{\lambda}, \boldsymbol{\delta}; t) = \delta_k + \lambda_k \int_0^t s_{\mathcal{K}} \left(\sum_{j=1}^{K} \mathrm{e}^{y_j(t')-\kappa} \right) \mathrm{d}t', \tag{91}$$

which provides the right-hand side of (85).

In the case of exponential growth, $s_{\mathcal{K}} = 1$ and the integral evaluates to t. For models with non-trivial saturation, like logistic and Gompertz growth, the integration in (91) needs to be performed numerically. (Analytical solutions for logistic and Gompertz growth are available only for single variant, i.e., $K = 1$.)

Another important difference from exponential growth concerns the arbitrariness of globally shifting the log abundances of all variants. In exponential growth, this freedom exists independently for both the initial log abundances, $\boldsymbol{\delta}$, and the growth rates, $\boldsymbol{\lambda}$. We fixed the gauge of the former by selecting the initial culture density as the unit for abundances, i.e., $N(0) = 1$. Note that this choice also determines the unit of the carrying capacity, $\mathcal{K}$, which is thus expressed as a multiple of the initial cell density.

The freedom to shift all growth rates in exponential growth arises because the experimental sequencing measurement captures only relative, not absolute, abundances. While this remains true in the current setting, the presence of a carrying capacity introduces an intrinsic reference for absolute abundances. Consequently, shifting all growth rates in the presence of saturation would alter the time required for the population to reach the carrying capacity. Therefore, once $\mathcal{K}$ is specified in units of $N(0)$, the growth rates obtained by maximizing either the likelihood or the ELBO correspond to absolute growth rates and can no longer be shifted arbitrarily.

6.2 Numerical examples

6.2.1 Four variants

For the illustrative example with four variants, we set the carrying capacity to be one thousand times larger than the initial cell density, i.e., $\mathcal{K} = 10^3$, so that $\kappa = \log 10^3 = 6.9$. This value of κ is indicated by horizontal dashed lines in the upper panels of fig. 1.

For the logistic growth model, we use the same growth rates as in exponential growth, given in (26). In the case of Gompertz growth, the saturation factor at the starting time evaluates to 6.9. Because this factor multiplies the parameter λ_k, the initial growth rates are substantially higher than in logistic growth, where

[e] Although it is interesting how Gompertz growth achieves saturation at $x = 1$ in a mathematically different way compared to the better known logistic growth,[27] by analyzing these two models in parallel we aim to highlight that such mathematical differences are unimportant for the implementation of the estimation approach. Alternatively, we could have considered the Richards family of models, of which logistic and Gompertz growth are special cases,[28;29] but the additional complication is not justified for our purposes.

the initial saturation factor is essentially 1. To ensure that Gompertz dynamics approach saturation on a similar timescale as logistic growth, we (arbitrarily) reduce $\boldsymbol{\lambda}$ from (26) by a factor of five for this model.

At the end of the displayed interval of 18 time units, the saturation factor equals 0.05 for logistic growth and 0.96 for Gompertz growth, showing that complete saturation ($s_{\mathcal{K}} = 0$) is reached more closely by the logistic model. This is reflected in the higher maximum log abundances observed in the upper panels of fig. 1. Because exponential growth is unconstrained by carrying capacity, its log abundances are not bounded by 6.9 and can exceed this value.

The trajectories in the top row of fig. 1 are obtained by integrating (89) using the Euler method,[30] with integration time step that is ten times smaller than the time unit of the horizontal axes, resulting in 180 steps between the initial and final times which are 18 time units apart. Decreasing the number of time steps, for example, by a factor of three, would decrease the computational time.

From the values of the Ground truth log abundances at the five selected sequencing time points, we calculate the corresponding variant fractions, and use these to generate multinomial counts as described for exponential growth. One such realization of synthetic counts is

$$\mathbf{N}_{\text{logistic}} = \begin{bmatrix} 184 & 245 & 331 & 376 & 407 \\ 97 & 77 & 51 & 20 & 12 \\ 49 & 36 & 13 & 3 & 2 \\ 82 & 17 & 3 & 0 & 0 \end{bmatrix}, \qquad \mathbf{N}_{\text{Gompertz}} = \begin{bmatrix} 184 & 263 & 338 & 368 & 400 \\ 97 & 68 & 44 & 27 & 16 \\ 49 & 36 & 13 & 3 & 5 \\ 82 & 8 & 3 & 1 & 0 \end{bmatrix}. \tag{92}$$

These are shown as symbols in the second row of fig. 1.

To be able to compare across the three models, we used the same random seed when generating the given read counts. Consequently, the counts at the initial time point—the first columns of the matrices in (27) and (92)—are identical across the three models. The counts at the subsequent time points depend both on the variant fractions determined by the Ground truth dynamics in fig. 1 and on the random draws from the multinomial distribution. Because logistic growth is practically indistinguishable from exponential growth away from saturation, the two models yield nearly identical variant fractions at early times. Indeed, in our example, their counts remain identical at the second and third time points, a consequence of having used the same random seeds in generating the multinomial samples from the variant fractions. By contrast, the counts for Gompertz growth begin to deviate immediately after the first time point (with the exception of the third variant).

The bottom row of fig. 1 displays the empirical variant fractions, $n_{k\tau}/n_\tau$. These quantities correspond to maximum-likelihood estimates at a single time point and therefore do not incorporate any temporal information from the growth model.

Temporal integration is first achieved through the all-time likelihood, after using the model of growth to describe its time-dependence. Maximizing this likelihood yields the parameter estimates $\boldsymbol{\lambda}^{\text{ML}}$ and $\boldsymbol{\delta}^{\text{ML}}$ for the two models with saturation. The variant fractions computed from these estimates are shown in the two rightmost panels of fig. 5. During the gradient-ascent optimization, the exponential sum of the $\boldsymbol{\delta}$ parameters was constrained to one, while the growth rates were left unconstrained, as their absolute magnitudes determine the timescale for approaching saturation.

The same procedure was applied to the all-time ELBO, with the only difference that, at each iteration step, $M = 100$ independent copies of the dynamics were propagated in time in order to compute the ELBO and its gradients, as described earlier. The resulting variational parameters are visualized in two different ways in fig. 7: the upper panels show the corresponding Gaussian posteriors, while the lower panels display the calculated trajectories of the variant fractions, including uncertainty bands. Results for exponential growth are also included for comparison.

When comparing the growth rates in the top row of fig. 7 across models, it is important to recall that their means are constrained to sum to zero for exponential growth, whereas no such restriction applies to logistic and Gompertz growth. In addition, both exponential and logistic growth employ the growth rates in (26), while these were scaled down by a factor of five for Gompertz growth to achieve comparable saturation timescales.

To enable direct comparison across models, Table 3 reverses these transformations: the inferred Gompertz

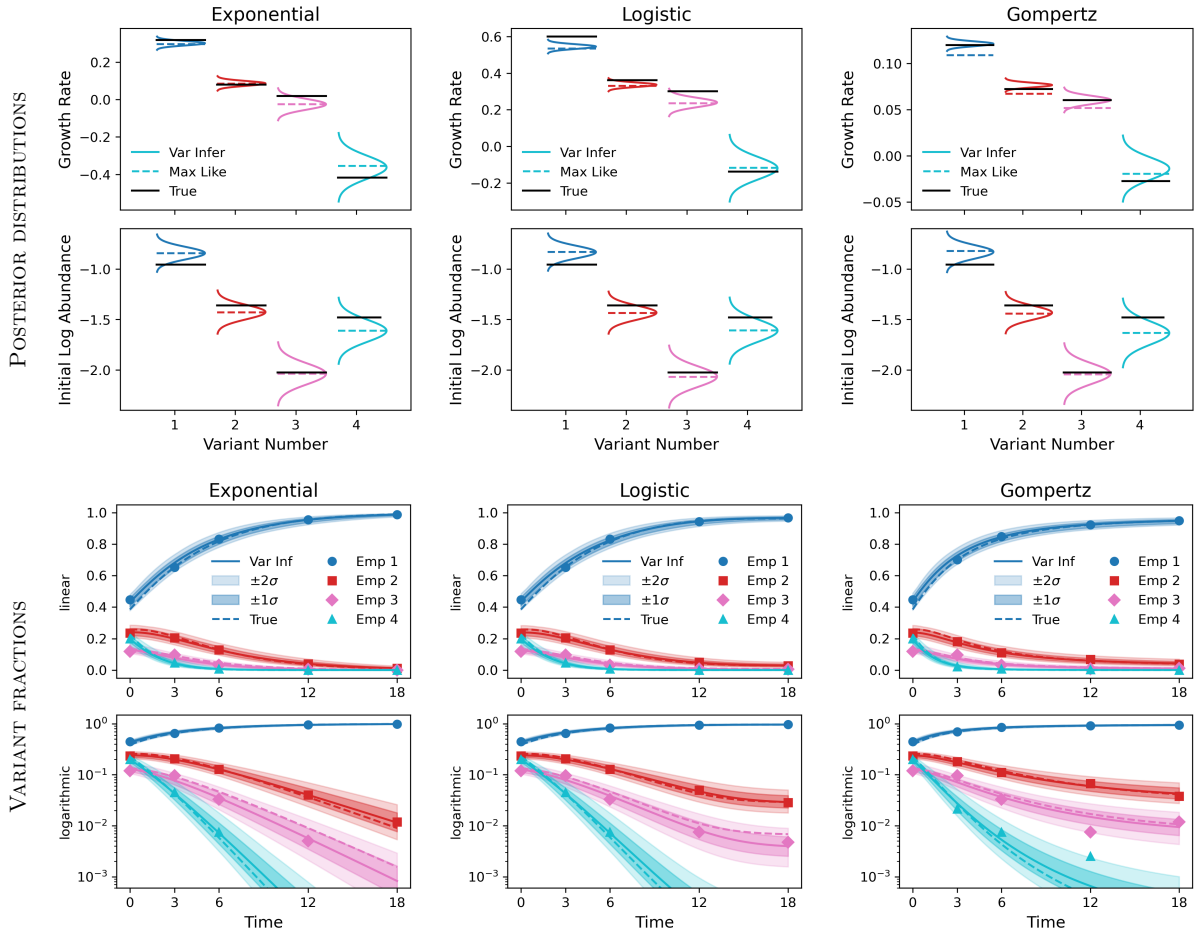

Figure 7: Variational inference of the parameters of the growth model and their uncertainties. Posterior distributions: Independent Gaussian factors shown in the interval $\mu_k \pm 3\sigma_k$ (bell curves, Var Infer), compared with maximum-likelihood estimates (dashed lines, Max Like) and true values (black lines, True). Variant fractions: Time trajectories calculated from the posterior means (colored solid lines, Var Inf) are accompanied by their uncertainty estimates (shaded bands, $\pm 1\sigma$ and $\pm 2\sigma$). The true trajectories (dashed lines, True) and the empirical ratios from the observed counts (symbols, Emp) lie within these uncertainty bands.

Table 3: Variational estimates of the growth rates and their uncertainties across the three growth models (Exp: exponential; Log: logistic; G: Gompertz). Gompertz means and standard deviations are multiplied by five; true growth rates from (26).

	$\boldsymbol{\mu}^\lambda$ (absolute values)	$\boldsymbol{\mu}^\lambda$ (zero-sum gauge)	$\boldsymbol{\sigma}^\lambda$ (uncertainties)
Exp	-	$[0.301 \quad 0.086 \quad -0.025 \quad -0.362]$	$[0.012 \quad 0.013 \quad 0.029 \quad 0.062]$
Log	$[0.545 \quad 0.337 \quad 0.240 \quad -0.119]$	$[0.295 \quad 0.086 \quad -0.011 \quad -0.370]$	$[0.012 \quad 0.012 \quad 0.025 \quad 0.061]$
G×5	$[0.606 \quad 0.383 \quad 0.302 \quad -0.069]$	$[0.301 \quad 0.077 \quad -0.003 \quad -0.375]$	$[0.013 \quad 0.012 \quad 0.022 \quad 0.060]$
True	$[0.600 \quad 0.361 \quad 0.301 \quad -0.137]$	$[0.319 \quad 0.080 \quad +0.019 \quad -0.418]$	-

means and standard deviations are multiplied by five, and the absolute means of the logistic and Gompertz rates are also shifted to a zero-sum gauge.

In contrast to the scaling and gauge adjustments applied to the growth rates, the means of the initial log abundances are normalized identically across all three models. Consequently, their Gaussian posterior distributions in fig. 7 are directly comparable.

The strong overall agreement observed in fig. 7 and Table 3 across the three growth models is highly satisfying, as it demonstrates that the subtle differences in the count data in (27) and (92) are being properly deconvolved by the respective growth models to yield consistent parameter estimates. On a more detailed level, however, it remains unclear why the variational means of the Gompertz growth rates deviate more strongly from their maximum-likelihood counterparts than do those of the exponential or logistic models.

To compute the $\pm m\sigma$ uncertainty bands ($m = 1, 2$) of the trajectories shown in the lower panels of fig. 7, for each model we generated $M = 1000$ realizations of the standard normal variables $(\boldsymbol{\epsilon}^\lambda, \boldsymbol{\epsilon}^\delta)$, and zeroed the values that were larger than m. Combining these random samples with the variational estimates of the means and standard deviations, we obtained $M = 1000$ different sets of growth parameters $(\boldsymbol{\lambda}, \boldsymbol{\delta})$, and calculated trajectories for each one of them. At each time point, the maximum and minimum variant fractions across these $M = 1000$ trajectories are shown as the upper and lower boundaries of the uncertainty bands in fig. 7. Across the three models, the observed empirical variant fractions (symbols) fall within the $\pm 2\sigma$ intervals. The true fractions used to generate the multinomial counts data (dashed lines) lie within the $\pm 1\sigma$ uncertainty bands.

6.2.2 Many variants

The minimal example of four variants discussed throughout this paper was chosen primarily for gaining intuition. From a practical perspective, however, it is essential to evaluate how the proposed inference scheme scales to a large number of variants. As a compromise between increasing the variant number while still allowing for visual inspection, we generated synthetic data for $K = 100$ variants under logistic growth.

The growth rates and initial log abundances of the variants, from which we computed the corresponding true trajectories, were drawn from Gaussian distributions. Using the true variant fractions at the five selected time points, we then generated multinomial counts with approximately $100K$ total counts per time point, following the same procedure as for $K = 4$. Prior to analysis, we verified that all variants had at least two nonzero counts across the five time points. For the realization analyzed here, the numbers of variants with 2, 3, 4, and 5 nonzero counts were 13, 37, 11, and 39, respectively (see fig. S5 in the SI).

Applying the same inference settings as in the four-variant analysis, we obtained the Gaussian posteriors shown in fig. 8A, where the variants are ordered by their true growth rates. Overall, the estimated posterior standard deviations are larger for variants with (i) lower initial abundances and (ii) smaller growth rates, as these jointly determine the total number of counts accumulated over time. In all cases, the true parameter values (black lines) lie within three standard deviations of the corresponding variational means.

Figure 8B displays the time-dependent variant fractions and their $\pm 2\sigma$ uncertainty bands for the first five variants (top panel) and the three variants before the last three (bottom panel); the former are shown on a linear scale, and the latter on a logarithmic scale. The bands were obtained from the maxima and minima across $M = 5000$ different trajectories in this case. Interestingly, even the tiny uncertainties in the growth

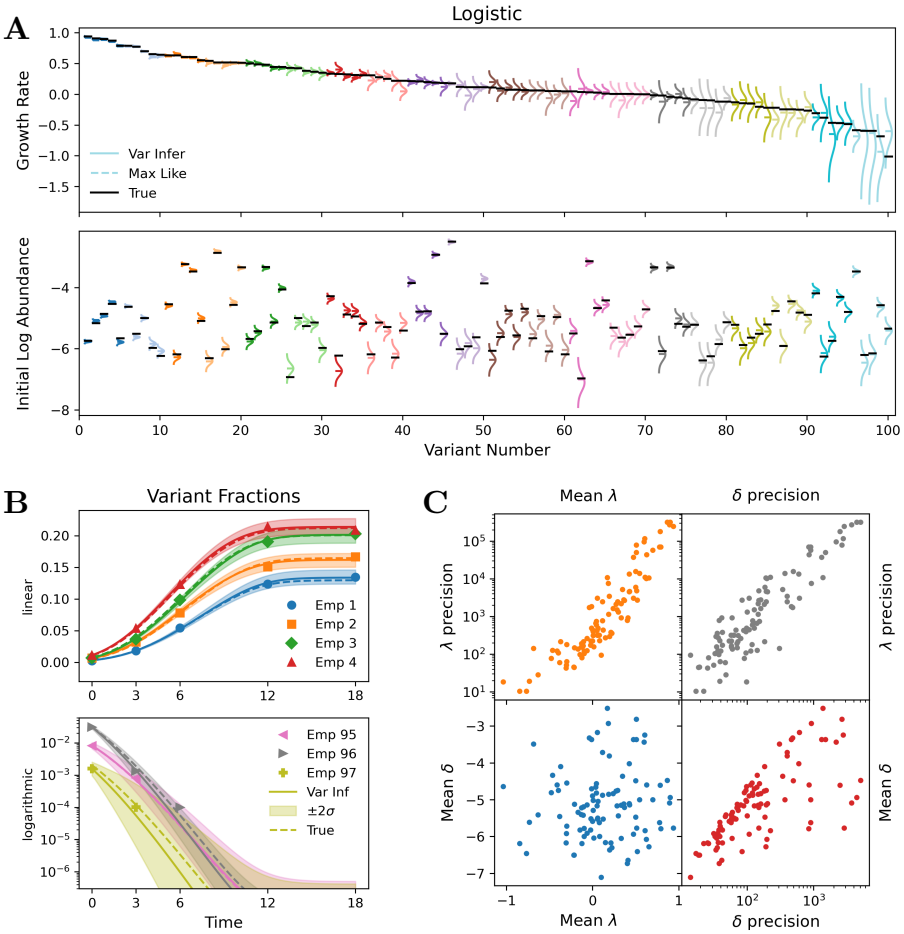

Figure 8: Variational Bayesian inference for $K = 100$ variants undergoing logistic growth. **A**: Gaussian posteriors (solid lines, Var Infer) with variants sorted according to their true growth rates. The true values (black lines) lie within the shown $\pm 3\sigma$ intervals from the estimated means. **B**: Time trajectories calculated using the variational parameters (solid lines, Var Inf) for the first four variants (top) and the three variants before the last three (bottom). The true trajectories (dashed lines, True) and the empirical count ratios (symbols, Emp) lie within the shown $\pm 2\sigma$ uncertainty bands. **C**: Scatter plots of the posterior means and precisions, where precision is the reciprocal variance (λ: growth rate; δ: initial log abundance).

rates of the first four variants in fig. 8A give rise to noticeable uncertainty bands in the upper panel of fig. 8B. This behavior illustrates that for rapidly growing variants, even small uncertainties in the growth rates can lead to noticeable differences in the predicted variant fractions. In the lower panel of fig. 8B, the uncertainty band of variant 96 (three non-zero counts) is smaller than those of variants 95 and 97 (two non-zero counts), in line with the number of non-zero counts of these variants.

In fig. 8C, we show the correlations between the inferred means and standard deviations, the latter expressed in terms of the precisions $1/\sigma_k^2$. The means of the growth rates and initial log abundances in the lower-left corner confirm that these parameters were indeed drawn from independent Gaussian distributions. The upper-left corner reveals a strong positive correlation between the means and precisions of the growth rates: faster-growing variants have more precisely estimated rates. A similar but weaker trend is observed for the initial log abundances (lower-right corner), where the variants with the highest δ-precisions have only intermediate mean δ values. The high precision of these points can be traced to the fact that they correspond to variants with high growth rates, resulting in larger total counts over time.

Finally, we note that optimizing the ELBO on a single CPU took 6.5 seconds for the logistic-growth example with $K = 4$ variants and 32 seconds for $K = 100$ variants, under identical computational settings. Further increasing the number of variants by a factor of ten to $K = 1000$, increased the runtime by the same factor to 5 minutes. Since the logistic dynamics in this example is integrated numerically, the optimization times are directly influenced by the time step of the Euler integrator. Increasing the integration time step by a factor of two, thus reducing the total number of steps in the time interval $[0, 18]$ from 180 to 90, resulted in runtimes of 17 and 128 seconds for $K = 100$ and $K = 1000$, respectively, which are smaller by a factor of two than the previous times of 32 and 300 seconds. Using higher-order integration schemes (e.g., Heun or Runge-Kutta) should therefore be considered when moving to more complex variant dynamics or when the number of variants is further increased to tens or hundreds of thousands. (This consideration is not relevant for exponential growth where the dynamics are calculated analytically.)

7 Concluding discussion

To the best of our knowledge, the *softmax* transformation has not been applied previously to the estimation of variant growth rates from sequencing counts, despite its well-established importance in the analysis of compositional data.[18;19] In this work, we revisited three established approaches for growth-rate estimation: (i) least-squares regression,[7] (ii) likelihood maximization, and (iii) variational Bayesian inference.[9] In each case, linking the probabilistic model of sequencing counts to the deterministic model of variant growth via the *softmax* transformation yielded new analytical insights and results.

In the context of weighted least-squares regression for exponential growth,[7] we argued that fitting a *softmax*-based analytical model to the data is preferable to fitting a linear model for two main reasons. First, the *softmax* fit better aligned instances of zero counts with the overall trends of the non-zero counts (*cf.* figs. 3 and 4). Second, and more importantly, it yielded estimates that were numerically invariant to the choice of reference variant. Notably, when formulating the mean-squared errors for these alternative fitting schemes, we derived the corresponding weight factors from a Bayesian treatment with a Dirichlet posterior (Sec. 3.2.1).

Despite these improvements, we contend that least-squares regression is ultimately inadequate for integrating sequencing counts across time points. Again two issues stand out: (i) the Bayesian uncertainty estimates apply to variant fractions rather than to the growth rates, which are the quantities of primary interest, and (ii) temporal integration is handled more naturally and rigorously by maximizing the likelihood under an appropriate noise model.

We modeled the noise in sequencing counts using a multinomial distribution, following the choice of Razo-Mejia and colleagues.[9] The multinomial model inherently respects the compositional constraint that variant fractions must sum to unity,[19] thereby capturing the negative covariances among them—an increase in the fraction of one variant necessarily decreases the fractions of the others. In contrast, modeling the stochasticity of individual variant counts using Poisson[6] or negative-binomial[8] distributions neglects this essential compositional structure.

Building on the intuition from least-squares fitting, we employed the *softmax* function to transform the standard parameters of the multinomial distribution—variant fractions—into log abundances. These log abundances served as the parameters of the resulting *multinomial-softmax* distribution (Sec. 4.1). Under exponential growth, they were available in closed form [eq. (10)]. Although our motivation for this reparameterization was to incorporate the growth-model parameters directly into the likelihood, the *softmax* transformation, in fact, naturally arises when the multinomial distribution is expressed as a member of the *exponential family* of distributions, where the new parameters—our log abundances—become the *natural parameters* of the distribution.[17]

In general, the maximization of the multinomial-*softmax* likelihood across multiple time points cannot be performed analytically, even under exponential growth where the log abundances are linear in time (Sec. 4.2). An exception arises in the special case of two sequencing time points (Sec. 4.3), for which we recovered the standard expression for growth-rate estimation used in the literature [eq. (53)]. The *softmax* perspective clarified the underlying role of the reference variant in this estimation procedure: because the growth rate of the reference variant relative to itself is, by definition, zero, its associated uncertainty is ill-defined. Our alternative growth-rate estimator [eq. (52)], which depends only on the counts of a single variant, does not suffer from this ambiguity.

The final methodology we examined through the lens of the *softmax* transformation was the Bayesian framework of Razo-Mejia and colleagues,[9] in which variational inference is used to estimate growth rates and quantify their associated uncertainties. Expressing variant fractions as a *softmax* transformation of the linear log abundances eliminated the need to first estimate the mean growth rate of the population, thereby simplifying the procedure of Razo-Mejia *et al.*[9]

One disconcerting aspect of variational inference is the rapid proliferation of free parameters that must be inferred from the data. When applied to a single sequencing time point, for example, the variational posterior introduced twice as many parameters as the Dirichlet posterior—and, in fact, twice as many as the number of observed counts (Sec. 5.1). In this setting of single time point, where growth dynamics are irrelevant, we examined how maximizing the ELBO determined $2K$ parameters from only K experimental observations. This analysis clarified the distinct roles of the posterior means and standard deviations: the means capture fractional information appropriate for compositional data, whereas the standard deviations encode information about absolute count magnitudes, which ultimately dictate the uncertainty of the estimated parameters (Sec. 5.1.1). This observation was further supported by the analytical lower bound to the ELBO, derived via Jensen's inequality, whose minimization yielded closed-form expressions for the optimal parameters (Sec. 5.1.3).

Just like the likelihood function on which it is based, the ELBO naturally integrates data across multiple sequencing time points [eq. (73)]. As the objective function in variational inference, the ELBO must be evaluated at each iteration for the current values of the variational parameters. However, because the expectation over the posterior distribution contained within the ELBO cannot be computed analytically, only limited analytical insight into its optimum can be obtained from examining its gradients. Even the analytical lower bound to the ELBO, derived from Jensen's inequality [eq. (75)], could not be minimized in closed form when data from several time points were integrated—except in the special case of only two time points (Sec. 5.2.2).

In this simpler case, we maximized the surrogate objective function analytically, and expressed the means and standard deviations in terms of the variant counts at the two time points [eq. (83)]. To our knowledge, the resulting growth-rate estimator has not been reported previously. Its inferred variance scales inversely with the counts at the second time point only, and introduces an entropic correction to the mean compared to the maximum-likelihood estimate [eq. (52)]. Numerical results indicated that the standard deviations obtained from this analytical ELBO surrogate substantially underestimated the true statistical uncertainty, particularly for variants with large counts (fig. 6).

The *softmax* transformation provided a natural interface between variant fractions, which parameterize the probabilistic model of counting noise, and variant log abundances, which evolved linearly with time under exponential growth. At the same time, it established a clear conceptual boundary between the probabilistic and deterministic components that are jointly needed to analyze sequencing data across multiple time points.

This separation highlighted that the log abundances—the inputs to the *softmax* function—could, in principle, originate from any underlying model of growth. Building on this observation, we extended the framework to logistic and Gompertz dynamics, which served as representative cases of non-exponential growth (Sec. 6). In this context, the use of automatic differentiation—which allowed gradients of the objective function to be accumulated seamlessly through the numerical integration of the growth model—was indispensable.[16;31]

The ability to incorporate arbitrary growth models into the proposed inference framework will open exciting opportunities for high-throughput inference of diverse biochemical parameters that influence growth. The three growth models examined here were parameterized directly in terms of variant-specific growth rates; however, more mechanistic formulations—in which the growth rates emerge as a function of underlying microscopic parameters—are equally conceivable. For instance, in studies of resistance to beta-lactam antibiotics, models have been proposed that express the growth rate of a variant in terms of the biochemical parameters $V_{\max}$ and K_M of the beta-lactamase enzyme it encodes.[14;15] Incorporating such models into the present inference framework would enable the simultaneous estimation of these parameters—and their associated uncertainties—in a highly parallel manner, through sequencing-based measurements of thousands of variants in pooled competition assays. The work presented here is a first step toward this goal.

Acknowledgement

We thank all members of the Toprak Lab—Fatma Coskun, Ilona Gaszek, Adam Lyon, and Sadık Yıldız—for fostering an inspiring research environment. DS is grateful to Kim Reynolds, Phil Brown, and Jerry Dinan for offering an excellent introductory course at UTSW on pooled competition experiments, through which he became aware of the Enrich2 software tool.[7] Their interest in applying Bayesian inference[9] to data analysis provided strong initial motivation for this work. DS also extends heartfelt gratitude to Ali Rana Atılgan for his constant encouragement and insightful feedback on this manuscript.

References

[1] Mariana Noto Guillen, Brittany Rosener, Serkan Sayin, and Amir Mitchell. Assembling stable syntrophic escherichia coli communities by comprehensively identifying beneficiaries of secreted goods. *Cell Systems*, 12(11):1064–1078.e7, 2021. doi: https://doi.org/10.1016/j.cels.2021.08.002. URL `https://www.sciencedirect.com/science/article/pii/S2405471221002908`.

[2] Ryan M. Otto, Agata Turska-Nowak, Philip M. Brown, and Kimberly A. Reynolds. A continuous epistasis model for predicting growth rate given combinatorial variation in gene expression and environment. *Cell Systems*, 15(2):134–148.e7, 2024. doi: https://doi.org/10.1016/j.cels.2024.01.003. URL `https://www.sciencedirect.com/science/article/pii/S2405471224000243`.

[3] Noam Prywes, Naiya R. Phillips, Luke M. Oltrogge, Sebastian Lindner, Leah J. Taylor-Kearney, Yi-Chin Candace Tsai, Benoit de Pins, Aidan E. Cowan, Hana A. Chang, Renée Z. Wang, Laina N. Hall, Daniel Bellieny-Rabelo, Hunter M. Nisonoff, Rachel F. Weissman, Avi I. Flamholz, David Ding, Abhishek Y. Bhatt, Oliver Mueller-Cajar, Patrick M. Shih, Ron Milo, and David F. Savage. A map of the rubisco biochemical landscape. *Nature*, 638(8051):823–828, 2025. doi: 10.1038/s41586-024-08455-0. URL `https://doi.org/10.1038/s41586-024-08455-0`.

[4] Ilona Kamila Gaszek, Muhammed Sadik Yildiz, Devin Meng, Jose Alberto de la Paz, Sophia Maria Alvarez, Faruck Morcos, Milo Lin, and Erdal Toprak. Higher-order epistasis drives evolutionary unpredictability toward novel antibiotic resistance. *bioRxiv*, Jul 2025. ISSN 2692-8205 (Electronic); 2692-8205 (Linking). doi: 10.1101/2025.07.08.663783.

[5] Thuy N. Nguyen, Christine Ingle, Samuel Thompson, and Kimberly A. Reynolds. The genetic landscape of a metabolic interaction. *Nature Communications*, 15(1):3351, 2024. doi: 10.1038/s41467-024-47671-0. URL `https://doi.org/10.1038/s41467-024-47671-0`.

[6] Sasha F. Levy, Jamie R. Blundell, Sandeep Venkataram, Dmitri A. Petrov, Daniel S. Fisher, and Gavin Sherlock. Quantitative evolutionary dynamics using high-resolution lineage tracking. *Nature*, 519(7542): 181–186, 2015. doi: 10.1038/nature14279. URL https://doi.org/10.1038/nature14279.

[7] Alan F. Rubin, Hannah Gelman, Nathan Lucas, Sandra M. Bajjalieh, Anthony T. Papenfuss, Terence P. Speed, and Douglas M. Fowler. A statistical framework for analyzing deep mutational scanning data. *Genome Biology*, 18(1):150, 2017. doi: 10.1186/s13059-017-1272-5. URL https://doi.org/10.1186/s13059-017-1272-5.

[8] Joao A. Ascensao, Kelly M. Wetmore, Benjamin H. Good, Adam P. Arkin, and Oskar Hallatschek. Quantifying the local adaptive landscape of a nascent bacterial community. *Nature Communications*, 14(1): 248, 2023. doi: 10.1038/s41467-022-35677-5. URL https://doi.org/10.1038/s41467-022-35677-5.

[9] Manuel Razo-Mejia, Madhav Mani, and Dmitri Petrov. Bayesian inference of relative fitness on high-throughput pooled competition assays. *PLOS Computational Biology*, 20(3):e1011937–, 03 2024. URL https://doi.org/10.1371/journal.pcbi.1011937.

[10] Olivia M. Ghosh, Grant Kinsler, Benjamin H. Good, and Dmitri A. Petrov. Low-dimensional genotype-fitness mapping across divergent environments suggests a limiting functions model of fitness. *bioRxiv*, page 2025.04.05.647371, 2025. URL https://www.biorxiv.org/content/biorxiv/early/2025/05/31/2025.04.05.647371.full.pdf.

[11] Daniel R. Amor and Jeff Gore. Fast growth can counteract antibiotic susceptibility in shaping microbial community resilience to antibiotics. *Proceedings of the National Academy of Sciences*, 119(15):e2116954119, 2025/08/14 2022. doi: 10.1073/pnas.2116954119. URL https://doi.org/10.1073/pnas.2116954119.

[12] Sivan Pearl Mizrahi, Akshit Goyal, and Jeff Gore. Community interactions drive the evolution of antibiotic tolerance in bacteria. *Proceedings of the National Academy of Sciences*, 120(3):e2209043119, 2025/08/14 2023. doi: 10.1073/pnas.2209043119. URL https://doi.org/10.1073/pnas.2209043119.

[13] Kyeri Kim, Andrea Weiss, Helena R. Ma, Hye-In Son, Zhengqing Zhou, and Lingchong You. Antibiotic-mediated microbial community restructuring is dictated by variability in antibiotic-induced lysis rates and population interactions. *Nature Communications*, 16(1):2299, 2025. doi: 10.1038/s41467-025-57508-z. URL https://doi.org/10.1038/s41467-025-57508-z.

[14] Eugene A Yurtsev, Hui Xiao Chao, Manoshi S Datta, Tatiana Artemova, and Jeff Gore. Bacterial cheating drives the population dynamics of cooperative antibiotic resistance plasmids. *Molecular Systems Biology*, 9(1):683, 2025/03/06 2013. doi: https://doi.org/10.1038/msb.2013.39. URL https://doi.org/10.1038/msb.2013.39.

[15] Tatiana Artemova, Ylaine Gerardin, Carmel Dudley, Nicole M Vega, and Jeff Gore. Isolated cell behavior drives the evolution of antibiotic resistance. *Molecular Systems Biology*, 11(7):822, 2025/04/17 2015. doi: https://doi.org/10.15252/msb.20145888. URL https://doi.org/10.15252/msb.20145888.

[16] Atılım Günes Baydin, Barak A. Pearlmutter, Alexey Andreyevich Radul, and Jeffrey Mark Siskind. Automatic differentiation in machine learning: a survey. *J. Mach. Learn. Res.*, 18 5595–5637, 2017.

[17] Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Springer, New York, 2006. ISBN 9780387310732.

[18] J. Aitchison and S. M. Shen. Logistic-normal distributions: Some properties and uses. *Biometrika*, 67 (2):261–272, 1980. ISSN 00063444, 14643510. URL http://www.jstor.org/stable/2335470.

[19] John Aitchison. *The Statistical Analysis of Compositional Data*. Monographs on Statistics and Applied Probability. Chapman and Hall, London, 1986. ISBN 0412280604.

[20] Anna Posfai, Juannan Zhou, David M. McCandlish, and Justin B. Kinney. Gauge fixing for sequence-function relationships. *PLOS Computational Biology*, 21(3):e1012818–, 03 2025. URL https://doi.org/10.1371/journal.pcbi.1012818.

[21] Marco J. Maier. Dirichletreg: Dirichlet regression for compositional data in r. Research Report Series/ Department of Statistics and Mathematics 125, WU Vienna University of Economics and Business, Vienna, January 2014. URL https://epub.wu.ac.at/4077/.

[22] David M. Blei, Alp Kucukelbir, and Jon D. McAuliffe. Variational inference: A review for statisticians. *Journal of the American Statistical Association*, 112(518):859–877, 04 2017. doi: 10.1080/01621459.2017.1285773. URL https://doi.org/10.1080/01621459.2017.1285773.

[23] Karl Friston, Jérémie Mattout, Nelson Trujillo-Barreto, John Ashburner, and Will Penny. Variational free energy and the laplace approximation. *NeuroImage*, 34 (1):220–234, 2007. doi: https://doi.org/10.1016/j.neuroimage.2006.08.035. URL https://www.sciencedirect.com/science/article/pii/S1053811906008822.

[24] Diederik P. Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013. URL https://arxiv.org/abs/1312.6114.

[25] Danilo Jimenez Rezende, Shakir Mohamed, and Daan Wierstra. Stochastic backpropagation and approximate inference in deep generative models. In Eric P. Xing and Tony Jebara, editors, *Proceedings of the 31st International Conference on Machine Learning*, volume 32 of *Proceedings of Machine Learning Research*, pages 1278–1286, Bejing, China, 22–24 Jun 2014. PMLR. URL https://proceedings.mlr.press/v32/rezende14.html.

[26] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems 32*, 2019.

[27] Charles P. Winsor. The gompertz curve as a growth curve. *Proceedings of the National Academy of Sciences*, 18(1):1–8, 2025/08/25 1932. doi: 10.1073/pnas.18.1.1. URL https://doi.org/10.1073/pnas.18.1.1.

[28] Even Tjørve and Kathleen M. C. Tjørve. A unified approach to the richards-model family for use in growth analyses: Why we need only two model forms. *Journal of Theoretical Biology*, 267(3):417–425, 2010. doi: https://doi.org/10.1016/j.jtbi.2010.09.008. URL https://www.sciencedirect.com/science/article/pii/S0022519310004741.

[29] Kathleen M. C. Tjørve and Even Tjørve. The use of gompertz models in growth analyses, and new gompertz-model approach: An addition to the unified-richards family. *PLOS ONE*, 12(6):e0178691–, 06 2017. URL https://doi.org/10.1371/journal.pone.0178691.

[30] William E. Boyce and Richard C. DiPrima. *Elementary Differential Equations and Boundary Value Problems*. Wiley, Hoboken, NJ, 11th edition, 2017. ISBN 9781119443767.

[31] Alp Kucukelbir, Dustin Tran, Rajesh Ranganath, Andrew Gelman, and David M. Blei. Automatic differentiation variational inference. *J. Mach. Learn. Res.*, 18 430–474, 2017.