Inference of tree-structured auto-regressive models of gene expression parameters from generation-snapshot data

Emrys Reginato, Aline Marguet, Eugenio Cinquemani

Inference of tree-structured auto-regressive models of gene expression parameters from generation-snapshot data*

Emrys Reginato, Aline Marguet[1] and Eugenio Cinquemani[1,†]

***Abstract*— In previous work, we proposed Auto-Regressive (AR) modelling on population trees for the stochastic transmission of individual-cell kinetic gene expression parameters at cell division. We addressed inference of the AR model parameters from individual gene expression profiles in a growing population, under the assumption of known parental relationships. In this paper, we explore the same inference problem in the case where only the generation that cells belong to is known, while parental relationships are unknown. First assuming that individual-cell parameters are measured directly with known degree of uncertainty, we develop a likelihood-based method that is applicable beyond the specific case of gene expression. Then, for data consisting of gene expression profiles, we extend the method into a pipeline for the identification of the AR model parameters via preliminary reconstruction of individual-cell parameters and their uncertainty. Performance of all methods is demonstrated via simulations inspired from real data.**

I. Introduction

Variability of gene expression kinetics and other phenotypic traits is obiquitously observed in single-cell experiments over isogenic cell populations. It is at the roots of important phenomena such as bet-hedging and adaptation, and in shaping biochemical networks via evolution [15]. Several modelling and inference methods have been developed to explore stochasticity within and across individual cells and obtain mathematical models explaining the variability in the data [19], [7]. Most methods simplify the description of population growth by treating single cells as independent individuals, while the parental relationships are not taken into account [20], [21], [14], [9], [1]. This can be a limitation, introduce bias in the analysis and overlook important phenomena observed in the data [3], [17], [5], [18].

In previous work [12], we considered an AR model for the stochastic transmission of kinetic parameters of gene expression from a mother cell to each of the two daughter cells of a growing cellular population, thus generalizing Mixed-Effects (ME) modelling to populations of individuals with tree-structured correlations. We further developed an inference method to reconstruct the AR model parameters from single-cell gene expression time profiles, under the assumption of known parental relationships among the observed cells. The biological relevance of the approach, which allows one to characterize the onset of phenotypic variability over generations in terms of the estimated AR model parameters, was demonstrated on the subset of fluorescence microscopy gene expression profiles from [11] for which parental relationships are available.

In this paper, we consider the same AR modelling framework, but we address inference in the case where parental relationships are unknown. Assuming that the generation an individual belongs to (that is its depth in the population tree) is known, we pursue estimation of the AR model parameters from individual-cell observations within generations, which we call generation-snapshot data. The first motivation for the work is that, as [11] witnesses, parental relationships among cells may be unavailable or nontrivial to obtain even from microscopy experiments (a counterexample being the use of mother-machines [16]). The broader motivation for the work is that the study of inference of tree-structured population models from snapshot data, as routinely collected *e.g.* via flow-cytometry experiments, currently seems lacking. Our work is a first step in this direction.

First assuming direct measurements of individual-cell traits of interest, we develop an exact maximum-likelihood method for inference of the AR model parameters from empirical means within generations, and an approximate generalization of the method incorporating empirical variances. Focusing on the case of a full binary tree with a single ancestor, we show by simulations the important role played by correlations across different generations and by the dynamics stemming from the single ancestor on estimation performance. We then extend the method to cope with indirect measurements of single-cell traits, focusing on the case where the traits of interest are kinetic gene expression rates and observations are gene expression profiles. We show that the method is capable of AR model inference on simulations of the experiments of [11]. While focused on cellular populations, results can be of interest for inference of tree-structured models in any application field (multiresolution analysis, phylogeny, image processing, ...).

In Sec. II, we discuss the modelling framework and set the stage for the inference problem. In Sec. III, we develop and assess performance of inference from direct measurements of individual traits. Sec. IV addresses inference from single-cell gene expression profiles. Conclusions and foreseen developments are discussed in Sec. V. Proofs, reported in Appendix in the interest of reviewing, will be removed to fit space constraints in a final paper version.

II. Modelling evolution of individual cell traits in a growing cell population

Borrowing from [12], we consider the following model for the evolution of single-cell traits in a population of dividing

*Work supported in part by project AnaComBa, Equipe-Action Persyval of the Université Grenoble Alpes. All authors are with Université Grenoble Alpes, Inria, 38000, Grenoble, France

[1] These authors contributed equally to the work.

[†]Corresponding author, eugenio.cinquemani@inria.fr

cells. Let $\boldsymbol{\varphi}^v$ be a vector of size m of real parameters quantifying one or several traits of an individual cell v. It is assumed that cell traits are constant over the life span of the cell. If v^- denotes the mother of cell v, we let $\boldsymbol{\varphi}^v$ evolve according to the AR model

$$\boldsymbol{\varphi}^v = A\boldsymbol{\varphi}^{v^-} + (I - A)\mathbf{b} + \boldsymbol{\eta}^v, \tag{1}$$

where A is a matrix of size $m \times m$, $\mathbf{b}$ a vector of size m, and $\boldsymbol{\eta}^v$ is a Gaussian random vector, independent across v, with variance $\Omega \in \mathcal{M}_m(\mathbb{R})$. By this model, daughter cell parameters are the result of a balance between mother cell parameters $\boldsymbol{\varphi}^{v^-}$ and reference parameters $\mathbf{b}$, plus a noise term that reflects randomness of the newborn cell. Matrix A is assumed diagonal with elements between zero and one. This matrix quantifies persistence: The closer A to the identity, the stronger the influence of the mother cell. More formally, since $\mathrm{Cov}(\boldsymbol{\varphi}^v, \boldsymbol{\varphi}^{v-}) = A\mathrm{Var}(\boldsymbol{\varphi}^{v-})$, A can be seen as the (normalized) covariance between mother and daugther cell parameters. We refer to A, $\mathbf{b}$ and Ω as the population parameters.

Provided a suitable transformation of $\boldsymbol{\varphi}^v$, this model may easily account for quantitative constraints and/or non-Gaussian distribution of individual cell traits. In [12], the model was developed to describe stochastic inheritance of kinetic gene expression parameters $\boldsymbol{\phi}^v$. Via the transformation $\boldsymbol{\phi}^v = \exp \boldsymbol{\varphi}^v$, the model suitably describes gene expression parameters as non-negative, log-normally distributed random variables. Here, we will not restrict ourselves to a specific (set of) trait(s), but we will come back to the specific case of single-cell gene expression kinetics in Sec. IV.

The above model takes explicitly into account the lineage cell tree, that is, tree-structured parental relations over subsequent cell generations. For simplicity we refer to the case of a complete binary tree, though several results that follow can be generalized. Let $v = 0$ be the index of the common ancestor cell at generation 0, and let S_n be the set of indices of cells of generation n (with reference to Eq. (1), if $v^- \in S_{n-1}$, then $v \in S_n$).

For every $v \in S_n$, with $n = 0, 1, \ldots$, we assume that a noisy version of the cell trait $\boldsymbol{\varphi}^v$,

$$\tilde{\boldsymbol{\varphi}}^v = \boldsymbol{\varphi}^v + \boldsymbol{\epsilon}^v \tag{2}$$

is available, where $\boldsymbol{\epsilon}^v \sim \mathcal{N}(0, R(n))$. Eq. (2) may equally represent the noisy measurement of a directly observable trait (*e.g.* cell size at birth), or the result of an estimation of individual-cell parameters from indirect measurements (*e.g.* kinetic gene expression rates from single-cell gene expression profiles, see Sec. IV).

In [12], under stationarity assumptions, we built inference algorithms for A, $\mathbf{b}$ and Ω assuming that the parental relationships among the observed cells were known. Here, instead, we investigate inference in absence of this information, while we still assume that the generation of a cell is known. That is, our measurements are given by the collections $\{\tilde{\boldsymbol{\varphi}}^v : v \in S_n\}$, with $n = 0, 1, \ldots$, which we call generation-snapshot data.

We aim at developing inference methods based on matching the dynamics of model-predicted statistics with measurement statistics calculated within generations. With the generation index in place of a time index, this is resemblant of existing moment-matching procedures for identification of reaction networks (see *e.g.* [20]), with the major difference that these procedures do not account for population lineages. In our case, due to the tree structure of the data-generating process, how to define and relate model and data statistics is not obvious and requires exploration. We focus the analysis on statistics up to second order (which are sufficient statistics under Gaussian assumptions). We consider that the parameters $\boldsymbol{\varphi}^0$ of the common ancestor are fixed, though generally unknown.

First consider the statistics of the individual cell parameters $\boldsymbol{\varphi}^v$. Since the AR model is the same along all branches of the lineage tree (all rooted in the same ancestor), we can define the same mean $\boldsymbol{\mu}(n) = \mathbb{E}(\boldsymbol{\varphi}^v)$ and variance $\Sigma(n) = \mathrm{Var}(\boldsymbol{\varphi}^v)$ for all $v \in S_n$. As for linear (as opposed to tree-structured) AR processes,

$$\boldsymbol{\mu}(n) = A\boldsymbol{\mu}(n-1) + (I - A)\mathbf{b} = A^n\boldsymbol{\varphi}^0 + (I - A^n)\mathbf{b}, \tag{3}$$

$$\Sigma(n) = A\Sigma(n-1)A^{\mathbf{T}} + \Omega = \sum_{i=0}^{n-1} A^i \Omega A^{i^{\mathbf{T}}}. \tag{4}$$

For large n, $\boldsymbol{\mu}(n)$ converges to $\mathbf{b}$ and $\Sigma(n)$ to the (unique) solution of $\Sigma = A\Sigma A^{\mathbf{T}} + \Omega$. Next consider empirical statistics of the measurements $\tilde{\boldsymbol{\varphi}}^{v_n}$ within different generations, namely the sample means and variances defined by

$$\tilde{\boldsymbol{\mu}}(n) = \frac{1}{|S_n|} \sum_{v \in S_n} \tilde{\boldsymbol{\varphi}}^v, \tag{5}$$

$$\tilde{\Sigma}(n) = \frac{1}{|S_n| - 1} \sum_{v \in S_n} (\tilde{\boldsymbol{\varphi}}^v - \tilde{\boldsymbol{\mu}}(n)) (\tilde{\boldsymbol{\varphi}}^v - \tilde{\boldsymbol{\mu}}(n))^{\mathbf{T}}. \tag{6}$$

How do (5)–(6) relate with (3)–(4) and with the model parameters? Precise relationships will be developed in the next section. Here, to provide an intuition behind the inference methods and results that will follow, we illustrate the properties of model and empirical statistics by numerical simulation in Julia [2].

Fig. 1 reports example profiles of (3)–(4) and (5)–(6) over 16 generations for a scalar model ($m = 1$, with matrices A, Σ, Ω replaced by scalars a, σ^2, ω^2, and non-bold notation for scalars replacing vectors). Empirical statistics are obtained from the random simulation of model (1) with $\omega^2 = 0.1$, and noiseless measurements (2). For the two cases of weak ($a = 0.1$) and strong inheritance ($a = 0.9$), for illustration purposes, we distinguish two scenarios: non-stationary ($\varphi^0 = \log(5) \neq b$) and stationary process mean ($\varphi^0 = b$), with $b = \log(32)$. We refer to them as the "atypical" and "typical" ancestor scenarios, in the same order.

In general, the profiles of the empirical statistics resemble the model statistics, with fluctuations that depend on the random terms η^v. This similarity is a first indication of viability for a moment matching approach to inference, the

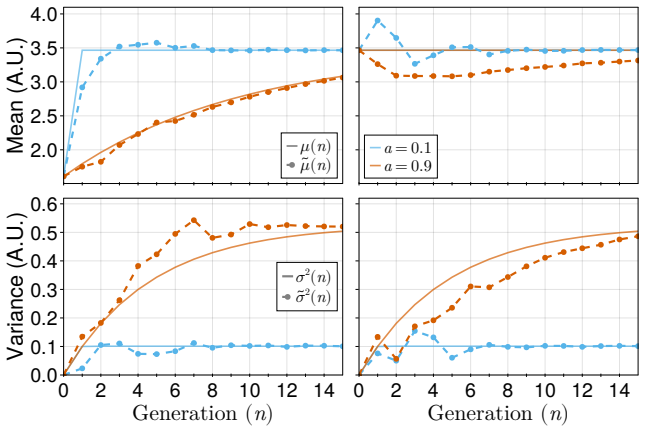

Fig. 1. Moment dynamics in the atypical (left) and typical (right) ancestor scenarios for two values of a. Solid lines: Equations (3)–(4); Dash-dotted lines: Empirical moments (5)–(6) from one simulation.

performance of which will depend on the informativeness of the moments about the model parameters.

In the case of an atypical ancestor, the mean displays transient dynamics that converge to the steady-state value b at (exponential) rate a (see Eq. (3)). For small enough a, convergence is attained within the experimental period, thus mean data suffice in principle for estimation of both a and b. For large a, instead, the steady-state value b is not apparent.

In the case of a typical ancestor, the model mean provides the value of b, but it does not convey any information about a. In contrast, due to the small number of individuals in the early generations, the empirical mean may significally depart from b (to an extent dependent on ω^2). Value b is recovered within the experimental period for small a only. This happens because the observed individuals are related through their lineage tree, whence the empirical statistics are correlated across generations. If this is not taken into account, interpretation of mean data is deceptive and may lead to estimation bias. Instead, since the timescale of fluctuations depends on a, one may be able to reconstruct a even if the model mean dynamics are insensitive to it. This fact will be formalized and exploited in the next section.

Concerning variance profiles, given the single ancestor, transient dynamics are present in either the typical and the atypical scenario. Indeed, by Eq.(3), model variance converges from zero to $\omega^2/(1-a^2)$ at rate a^2 (*i.e.* faster than convergence of means). Thus, provided a formal relation between model and empirical variance is established (see next section), variance dynamics convey information about a and also about ω, irrespective of the ancestor.

In conclusion, we saw that such simple model of trait evolution over a tree of dividing cells gives rise to dynamics of generation snapshot statistics that are more complex to interpret, but also potentially more informative, than population snapshot data collected on independent individuals [7]. In the next section we address the question of how to appropriately process this data for model inference.

III. Reconstruction of Trait Evolution Dynamics from Direct Measurements

In this section we shall develop methods to infer parameters $\theta = (A, \mathbf{b}, \Omega, \boldsymbol{\varphi}_0)$ from snapshot statistics $\tilde{\boldsymbol{\mu}}(n)$, $\tilde{\Sigma}(n)$, over $N_g + 1$ generations, *i.e.* $n = 0, 1, \ldots, N_g$. We will develop an exact maximum likelihood approach for inference from mean data, and an extension for the joint use of mean and variance data, and demonstrate their performance on simulated data. For all n, we assume $R(n)$ to be known (or estimated in a preliminary step, Sec. IV), and invertible.

A. Estimation from empirical means only

Let $\mathbf{y} = (\tilde{\boldsymbol{\mu}}(0)^{\mathbf{T}}, \tilde{\boldsymbol{\mu}}(1)^{\mathbf{T}}, \cdots, \tilde{\boldsymbol{\mu}}(N_g)^{\mathbf{T}})^{\mathbf{T}} \in \mathbb{R}^{m\times(N_g+1)}$. This is a Gaussian random vector whose mean $\bar{\mathbf{y}}_\theta = \mathbb{E}_\theta(\mathbf{y})$ and covariance matrix $\Gamma_\theta = \mathrm{Var}_\theta(\mathbf{y})$ have structure

$$\bar{\mathbf{y}}_\theta = \begin{pmatrix} \bar{\mathbf{y}}_\theta(0) \\ \vdots \\ \bar{\mathbf{y}}_\theta(N_g) \end{pmatrix}, \quad \begin{pmatrix} \Gamma_\theta(0,0) & \cdots & \Gamma_\theta(N_g,0)^{\mathbf{T}} \\ \vdots & \ddots & \vdots \\ \Gamma_\theta(N_g,0) & \cdots & \Gamma_\theta(N_g,N_g) \end{pmatrix}, \tag{7}$$

with $\bar{\mathbf{y}}_\theta(i) = \mathbb{E}_\theta(\tilde{\boldsymbol{\mu}}(i))$ and $\Gamma_\theta(i,j) = \mathrm{Cov}_\theta(\tilde{\boldsymbol{\mu}}(i), \tilde{\boldsymbol{\mu}}(j))$ for $0 \le i, j \le N_g$.

Proposition 1: For a a complete binary tree, for $0 \le j \le i \le N_g$, it holds that

$$\bar{\mathbf{y}}_\theta(i) = A^i\boldsymbol{\varphi}^0 + (I - A^i)\mathbf{b}\,, \tag{8}$$

$$\Gamma_\theta(i,j) = \frac{1}{2^{i\wedge j}}\left(A^{|i-j|}\sum_{k=0}^{i\wedge j-1} 2^k A^k \Omega A^{k^{\mathbf{T}}} + \delta_{i,j}R(i)\right), \tag{9}$$

with $\delta_{i,j}$ the Kronecker delta and $i\wedge j$ the minimum between i and j.

Notice that the expression of $\bar{\mathbf{y}}_\theta(i)$ is equal to (3). The off-diagonal blocks $\Gamma_\theta(i,j)$, with $i \neq j$, are nonzero as a result of correlation across generations. Factor $A^{|i-j|}$ in (9) represents the correlation decay of empirical means, *i.e.* the timescale of their fluctuations along generations: As observed in the previous section based on simulation, the smaller the A, the faster the fluctuations.

Given $\mathbf{y}$, for a suitable parameter space Θ, we define $\hat{\theta}_{\mathbf{y}} \in \Theta$ as the maximum likelihood estimator of θ. Equivalently,

$$\hat{\theta}_{\mathbf{y}} = \arg\min_{\theta\in\Theta} \ell(\theta|\mathbf{y}) \tag{10}$$

where, up to an additive constant independent of θ, $\ell(\theta|\mathbf{y})$ is the negative log-likelihood function, given by

$$\ell(\theta|\mathbf{y}) = \frac{1}{2}\left(\log(|\Gamma_\theta|) + (\mathbf{y}-\bar{\mathbf{y}}_\theta)^{\mathbf{T}}\Gamma_\theta^{-1}(\mathbf{y}-\bar{\mathbf{y}}_\theta)\right). \tag{11}$$

Using (7)–(9), this function can be evaluated explicitly for any $\theta \in \Theta$. Then, the solution of the (generally nonconvex) problem (10) can be sought by numerical optimization. In practice, in view of the form of (11), estimates $\hat{\theta}_{\mathbf{y}}$ result from the matching of mean dynamics (differences between data $\mathbf{y}$ and model predictions $\bar{\mathbf{y}}_\theta$ must be small) and of correlations (differences $\mathbf{y} - \bar{\mathbf{y}}_\theta$ must agree with the structure of matrix Γ_θ). We thus fully profit from the information carried by empirical means and illustrated in the previous section.

B. Estimation from empirical means and variances

In principle, by extending the approach of the previous section, one could tackle inference from empirical mean and variance data by maximization of their joint likelihood. Unfortunately, deriving formulas for this joint likelihood is a formidable task, due to the non-Gaussian nature of the random variables $\tilde{\Sigma}(n)$ and the intricate interplay of individual observations across different generations. We therefore introduce a fitting cost function combining the negative log-likelihood $\ell(\theta|\mathbf{y})$ with a suitable fitting term for the dynamics of the empirical variance over generations. For $n = 1, \dots, N_g$ let $\mathbf{v}(n)$ be a column vector defined as $\mathbf{v}(n) = \mathscr{S}\big(\tilde{\Sigma}(n)\big)$, where $\mathscr{S}$ is a linear operator extracting a suitable set of elements of the (symmetric) matrices $\tilde{\Sigma}(n)$, and let $\bar{\mathbf{v}}_\theta(n) = \mathbb{E}_\theta\left(\mathbf{v}(n)\right)$. Indicating with $\mathbf{v}$ the (vector) collection of the $\mathbf{v}(n)$, we define the fitting cost

$$c(\theta|\mathbf{y},\mathbf{v}) = \frac{1}{2}\left(\log(|\Gamma_\theta|) + (\mathbf{y}-\bar{\mathbf{y}}_\theta)^{\mathbf{T}}\Gamma_\theta^{-1}(\mathbf{y}-\bar{\mathbf{y}}_\theta)\right) + \frac{1}{2}\sum_{n=1}^{N_g}\big(\mathbf{v}(n)-\bar{\mathbf{v}}_\theta(n)\big)^{\mathbf{T}}W_\theta^{-1}(n)\big(\mathbf{v}(n)-\bar{\mathbf{v}}_\theta(n)\big) \tag{12}$$

and the estimator of θ from empirical mean and variance data as

$$\hat{\theta}_{\mathbf{y},\mathbf{v}} = \arg\min_{\theta\in\Theta} c(\theta|\mathbf{y},\mathbf{v}). \tag{13}$$

The first term of Eq. (12) is the expression of $\ell(\theta|\mathbf{y})$ from Eq. (11). In analogy with this term, the second term amounts to squared residuals between empirical (variance) statistics and their expected value, weighted by suitable invertible matrices $W_\theta(n)$ that must ensure an appropriate tradeoff between the fitting cost for means (first term) and variances (second term). We will come back on the definition of the $W_\theta(n)$ soon. Different from $\ell(\theta|\mathbf{y})$, in the second term, contributions from different generations are treated separately.

Proposition 2: For a complete binary tree, for $n = 0, \dots, N_g$, it holds that $\bar{\mathbf{v}}_\theta(n) = \mathscr{S}\left(\mathbb{E}_\theta\left(\tilde{\Sigma}(n)\right)\right)$, with

$$\mathbb{E}_\theta\left(\tilde{\Sigma}(n)\right) = \Sigma_\theta(n) + R(n) + \frac{1}{2^n-1}\sum_{i=0}^{n-1}(1-2^i)A^i\Omega{A^i}^{\mathbf{T}}, \tag{14}$$

where $\Sigma_\theta(n)$ is given by (4).

Eq. (14) shows that the empirical variance $\tilde{\Sigma}(n)$ has expected value equal to the model variance $\Sigma(n)$ plus ($R(n)$ and) a nontrivial term that results from the correlation of individuals across generations. Failing to account for this term in a naive match between empirical and model-predicted variances would introduce fitting bias. Instead, by virtue of Eq. (14), the second term in (12) duly quantifies deviations between empirical and model-predicted variance dynamics. As discussed in Sec. II, this term is expected to provide the necessary complement to empirical mean data for the estimation of all population parameters.

Inspired from $\ell(\theta|\mathbf{y})$, provided a suitable definition of $\mathscr{S}$ guaranteeing invertibility, one may define $W_\theta(n) = \mathrm{Cov}_\theta(\mathbf{v}(n))$. Calculating this covariance matrix is generally complex. We instead propose to define this matrix under the simplifying assumption that individuals within one generation are uncorrelated (as if $A = 0$). In this case, for indices k (resp. k') such that $\mathscr{S}$ maps element (i,j) (resp (i',j')) of its input matrix into element k (resp. k') of its output vector, the element in position (k,k') of matrix $W_\theta(n)$ is given by (see [6, Thm. 3.3.15])

$$(|S_n|-1)^{-1}(\varsigma_{i,i'}\varsigma_{j,j'} + \varsigma_{i,j'}\varsigma_{j,i'}), \tag{15}$$

where $\varsigma_{i,j}$ is the element of position (i,j) of matrix $\Sigma_\theta(n) + R(n)$. Using (14) and (15), the fitting cost (12) can be evaluated efficiently as an explicit function of θ. The solution to the (generally nonconvex) problem (13) can then be sought by numerical optimization. The effectiveness of our approach is demonstrated by simulation in the next section.

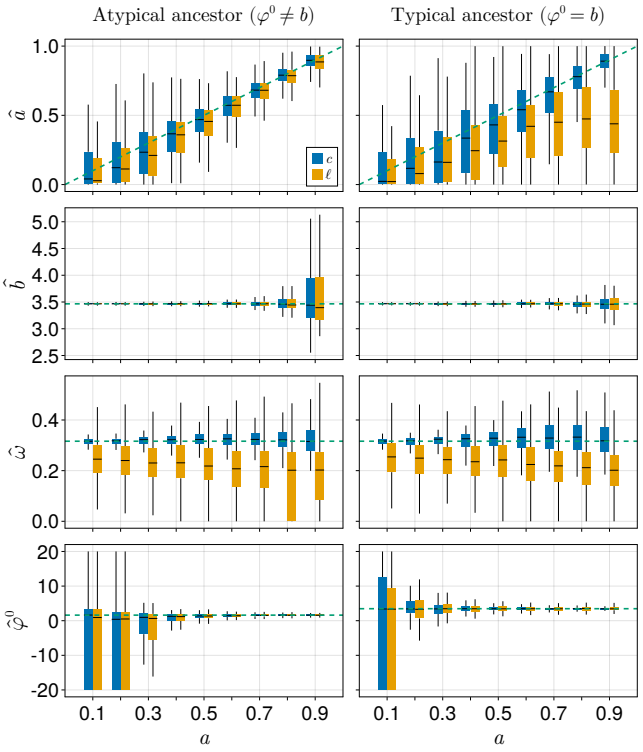

Fig. 2. Estimation performance for different simulated values of a from direct measurements of individual parameters in the atypical (left) and typical (right) ancestor scenarios, using means only (Eq. (10), yellow), and also using variances (Eq. (13), blue). True values: Dashed green line

C. Numerical performance assessment

We now evaluate and compare performance of the estimation methods developed above. For illustration purposes, we distinguish the atypical and a typical ancestor scenario, though this information is not used in estimation. Fig. 2 reports the results from Monte-Carlo analysis of estimation performance based on simulated datasets. One dataset consists in one random realization of Eq. (1) over a complete binary tree with 10 generations ($N_g = 9$). Noisy measurements of the simulated parameters are generated in accordance with

Eq. (2). Simulation parameters are $b = \log(32)$, $\omega^2 = 0.1$ and $R(n)$ identical for all n and equal to $r^2 = 0.1$. For every value $a = 0.1, 0.2, \ldots, 0.9$, we generated 500 such datasets for the typical ancestor scenario $\boldsymbol{\varphi}_0 = b$ and 500 datasets for the atypical ancestor scenario $\boldsymbol{\varphi}_0 = \log(5)$. For each simulated case, boxplots summarize the estimation results obtained from the application of the methods of Sec. III-A (usage of empirical means only) and of Sec. III-B (usage of empirical means and variances, with $\mathscr{S}$ set to the identity map) to each of the 500 datasets. In our implementation in Julia, every estimation run takes between 1 and 2 seconds on a modern laptop.

Results reflect the expectations illustrated in Sec. II. In the atypical ancestor scenario, using empirical means only (Problem (10)) yields estimates of a that appear unbiased, are more accurate for larger values of a. This reflects the fact that the transient mean dynamics are not exhausted within the very first generations, where empirical means are noisier due to the smaller number of individuals. The same is true for the estimates of φ_0, which are essentially undefined for small a. Estimates of b instead worsen with increasing a because of the limited number of observed generations. Estimates of ω, although biased, are also of the right order of magnitude. This is an interesting consequence of the fact that ω enters Problem (10) via the covariance matrix Γ_θ. The additional use of empirical variances (Problem (13)) leads to analogous results, the only exception being the improved estimates of ω. This is explained by the role of ω in Eq. (14).

In the more challenging, typical ancestor scenario, the interest of exploiting empirical variances becomes apparent. Performance in the estimation of a remains essentially unbiased and it is clearly better than using empirical means only. Nonetheless, it is interesting to observe that using means only, estimates of a overall follow the increasing pattern of values of a tested. This is quite remarkable and entirely due to the incorporation of the covariance matrix Γ_θ in Problem (10). For the two methods, performance in the estimation of the remaining parameters (b, φ_0, and ω), is qualitatively comparable to the atypical ancestor scenario.

In summary, two things were shown in this section. First, accounting for correlations among empirical means, as per the exact maximum likelihood formulation of Problem (10), enables estimation of the model parameters, although estimates of ω and a are biased in some cases. This finding is conceptually interesting and it may have practical relevance for applications where only empirical means are available. Second, additionally exploiting empirical variances as in Problem (13) enables estimation of all model parameters in all conditions. While departing from an exact maximum-likelihood approach, this finding qualifies optimization (13) as an effective parameter estimation method, and leads us to focus on this method in the sequel.

IV. Reconstruction from indirect parameter measurements: Gene expression case study

In the previous section, we have developed methods to reconstruct parameters $\theta = (A, \mathbf{b}, \Omega, \boldsymbol{\varphi}_0)$ based on noisy measurements $\tilde{\boldsymbol{\varphi}}^v$ of the individual-cell parameters $\boldsymbol{\varphi}^v$ and known measurement uncertainties $R(n)$. We now discuss application of the previously developed methods to scenarios where measurements of individual-cell parameters are not directly available, and measurement uncertainties $R(n)$ are a priori unknown.

Motivated by [11], we do so for the case study of gene expression dynamics. In [11], expression of an osmosensitive gene in response to repeated osmotic shocks is monitored over time in individual yeast cells by the use of a fluorescent reporter protein and videomicroscopy. Data is used to investigate variability of kinetic rate parameters of gene expression across cells. In [11], correlation of the estimated parameters across mother and daughter cells is only evaluated *a posteriori* based on the few cells for which parental relationships are available. In our more recent work [12], the correlation among mother and daughter cells expressed by model (1) is estimated directly, again limited to the cells with known parental relationships. The methods presented in Sec. III enable estimation of mother-daughter correlations without requiring the knowledge of parental relationships. To achieve this, however, individual-cell parameter estimates $\tilde{\boldsymbol{\varphi}}^{v_n}$ and uncertainty $R(n)$ must be obtained for every generation n in a preliminary step.

In Sec. IV-A we elaborate on this preliminary step. We propose robust methods to compute (noisy) individual-cell parameters $\tilde{\boldsymbol{\varphi}}^v$ along with their uncertainty $R(n)$ from gene expression time profiles. In conjunction with the method of Section III-B, this establishes a pipeline to obtain estimates of parameters θ from gene expression time profiles in absence of lineage information. In Sec. IV-B, we show performance on simulations of the experiments of [11]. While focused on the gene expression case study, the methods proposed can be readily generalized.

A. Inference from single-cell gene expression profiles

The procedure to calculate individual-cell parameters $\tilde{\boldsymbol{\varphi}}^v$, with $v \in S_n$, and $R(n)$ from gene expression time profiles applies separately to every generation n. We assume that the generation that cells belong to is known (for comments on this point see Sec. V). For a fixed n, let $t_0, \ldots, t_{N_v - 1}$ be (increasing) measurement times for cell $v \in S_n$. Let $\tilde{p}^v_1, \ldots, \tilde{p}^v_{N_v}$ be measurements of the cellular concentration of a (fluorescent) reporter protein,

$$\tilde{p}^v_\ell = p^v_\ell + \varepsilon_\ell, \tag{16}$$

with p^v_ℓ the true protein concentration at time t_ℓ and $\varepsilon_\ell \sim \mathcal{N}(0, \sigma^2_\varepsilon)$ measurement noise uncorrelated across times and cells. Variance σ^2_ε is considered unknown. We assume that $p^v_\ell = p(t_\ell | \boldsymbol{\varphi}^v)$, where $p(t|\boldsymbol{\varphi}^v)$ is the solution at time t of the simple gene expression model

$$\frac{\mathrm{d}}{\mathrm{d}t} p(t) = -\gamma p(t) + \kappa^v u(t), \quad p^v(t^v_0) = p^v_0, \tag{17}$$

with $\boldsymbol{\varphi}^v = (\kappa^v, p^v_0)$. In Eq. (17), γ is a concentration decay rate resulting from growth dilution and protein degradation, κ^v is the protein synthesis rate upon gene expression activation, and $u(t)$ represents promoter activation ($u = 1$) and

deactivation ($u = 0$) at time t in response to known exogenous stimuli, assumed identical across cells [11]. For stable reporter proteins, this simple model is a viable approximation of transcription-translation dynamics, and γ is determined by growth rate uniformly across cells (see [4] and references therein). Eventually, among the entries of $\boldsymbol{\varphi}^v$, we will focus on the evolution over generations of κ^v. That is, model (1) will be restricted to the scalar $\varphi^v = \kappa^v$. No transmission model is postulated for protein concentrations. Yet we let p_0^v be different across cells and thus part of the unknown individual-cell parameters $\boldsymbol{\varphi}^v$.

Estimation of $\boldsymbol{\varphi}^v$ from data $\mathcal{D}^v = \{(t_\ell^v, \tilde{p}_\ell^v),\ \ell = 0, \dots, N_v\}$ could be performed by least-squares fitting separately for every cell v, however, performance may be limited for sparse data. We propose instead a Mixed-Effects (ME) approach [10]. According to the basic ME paradigm, which is sufficient to our purpose, the unknown parameters $\boldsymbol{\varphi}^v$ of all individuals of a given population (for us, S_n) are treated as random outcomes from a common distribution $\mathcal{F}(\Xi)$ with parameters Ξ. Given knowledge of an individual statistical response model $p(\cdot|\boldsymbol{\varphi}^v)$, and a measurement model (16), data $\mathcal{D}^v$ from all individuals $v \in S_n$ are pooled together to calculate estimates of Ξ along with estimates $\hat{\boldsymbol{\varphi}}^v$ of individual parameters $\boldsymbol{\varphi}^v$. An estimate $\hat{\sigma}_\varepsilon^2$ of σ_ε^2 is also calculated in the procedure. Thanks to all individuals being treated as part of a same statistical population, ME inference is known to outperform individual inference especially for noisy, short individual time series [10]. We will specifically rely on the ME inference method known as SAEM [10], assuming that $\mathcal{F}(\Xi)$ is a log-normal distribution to complete the problem specification.

Assume that estimates $\{\hat{\boldsymbol{\varphi}}^v : v \in S_n\}$ and $\hat{\sigma}_\varepsilon^2$ have been obtained by ME inference. To enable application of the method of Section III-B, we set $\tilde{\boldsymbol{\varphi}}^v = \hat{\boldsymbol{\varphi}}^v$ for all v, and deduce $R(n)$ from $\hat{\sigma}_\varepsilon^2$ as follows. For every individual v, the variance Υ^v of the parameter estimate $\hat{\boldsymbol{\varphi}}^v$ is approximated locally by $\Upsilon^v = \hat{\sigma}_\varepsilon^2 (G_v^{\mathbf{T}} G_v)^{-1}$, where the ℓth row of matrix G_v is the (local) sensitivity of $p(t_\ell|\boldsymbol{\varphi}^v)$ to variations in $\boldsymbol{\varphi}^v$. These are given by $[\partial p(t_\ell|\boldsymbol{\varphi}^v)/\partial \boldsymbol{\varphi}^v]_{\boldsymbol{\varphi}^v = \hat{\boldsymbol{\varphi}}^v}$ and can be easily calculated by means of the sensitivity equations [8]. Finally, $R(n)$ is defined as the mean of Υ^v across all $v \in S_n$.

In sums, the whole pipeline goes as follows.

- For $n = 0, \dots, N_g$:
  1) Given data $\mathcal{D}^v$ for all $v \in S_n$, calculate estimates $\{\hat{\boldsymbol{\varphi}}^v : v \in S_n\}$ and $\hat{\sigma}_\varepsilon^2$ by SAEM;
  2) For every individual $v \in S_n$ run sensitivity equations to calculate G_v and $\Upsilon^v = \hat{\sigma}_\varepsilon^2 (G_v^{\mathbf{T}} G_v)^{-1}$;
  3) Calculate $R(n)$ as the mean of Υ^v across $v \in S_n$, and set $\tilde{\boldsymbol{\varphi}}^v = \hat{\boldsymbol{\varphi}}^v$ for all $v \in S_n$;
  4) Calculate empirical statistics (5)-(6)
- Find $\hat{\theta}$ by solving Problem (13).

The last step can be easily modified to focus on a subset of the entries of $\boldsymbol{\varphi}^v$, as it is the case in next section. Note that estimates of $\tilde{\mu}(n)$ and $\tilde{\Sigma}(n)$ are also inherently calculated by SAEM in the reconstruction of Ξ. However, their construction in SAEM may not fulfill Eq. (14).

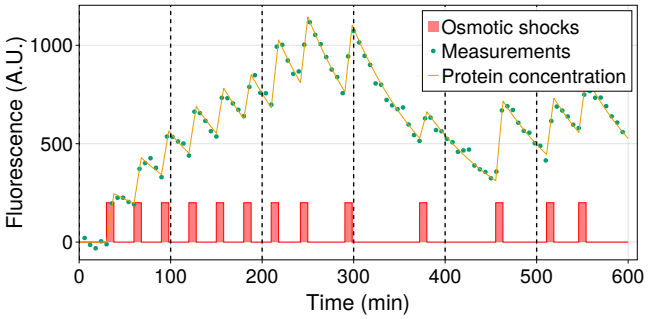

Fig. 3. Example simulation of gene expression dynamics along one branch of the population tree. Vertical dashed lines: Cell division times.

B. Simulation results

We now show the performance of the estimation pipeline described above by a numerical Monte-Carlo study. We consider the gene expression model (17), with parameter $\varphi^v = k^v$ obeying the AR model (1) and γ fixed to 0.01. We consider cells v over $N_g + 1 = 10$ generations and assume that all cells of generation n are born at time $100n$ and divide at time $100(n+1)$ (minutes). For parameters b and ω fixed as in Sec. III-C, in the typical ancestor scenario ($\boldsymbol{\varphi}^0 = b$), we generated 50 gene expression datasets for each value $a = 0.1, \dots, 0.9$. Every dataset is obtained in three steps. In the first step, we simulate a complete binary tree of parameter values κ^v over $N_g + 1 = 10$ generations, as per Eq. (1). In a second step, we simulate (17) for every κ^v over the time period of the corresponding generation, with initial conditions p_0^v fixed to $p^{v^-}(t_0^v)$, *i.e.* the concentration level in the mother cell at division (for the ancestor the initial condition is set to zero). From the simulated values p_ℓ^v, we finally obtain noisy measurements $\tilde{p}_\ell^v$ as per Eq. (17), with realistic measurement noise strength set to $\sigma_\varepsilon = 20$. Fig. 3 illustrates the chosen gene expression input profile $u(t)$ and example simulated data over one branch of one population tree. For every value of a, we then run the estimation pipeline of Sec. IV-A on each of the 50 simulated datasets. Due to difficulties encountered with the SAEM implementations in Julia, we rather implemented and run the first part of the pipeline in Matlab [13] based on function `nlmefitsa`. One complete run of the pipeline for one population tree takes about 5 minutes.

Estimation statistics are reported in the form of boxplots in Fig. 4. It can be appreciated that results are qualitatively similar to the results of Sec. III-C, which were based on direct parameter measurements. In particular, the estimates obtained here from noisy gene expression data remain unbiased and reasonably concentrated around the true values in all cases. Quantitative comparison with the results in Fig. 2 is not appropriate, since the error variance matrices $R(n)$ that enter the results of Fig. 4 are estimated from the simulated data and generation-dependent. A detailed study of how the preliminary step of estimation of single-cell parameters and matrices $R(n)$ contributes to overall estimation uncertainty is part of future work.

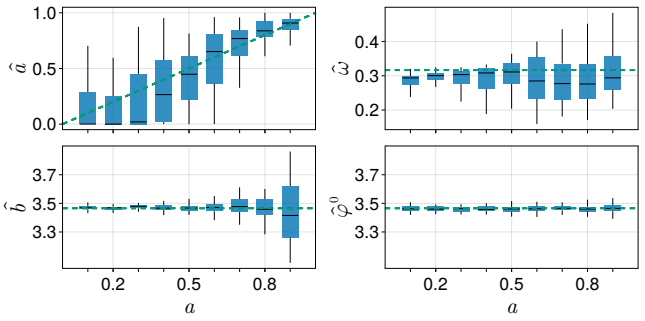

Fig. 4. Estimation performance from gene expression profiles, for different simulated values of a. True values: Dashed green line

V. CONCLUSIONS AND FUTURE WORK

We have presented methods for estimation of an AR model of individual-cell trait evolution on a population tree from trait statistics within generations. We developed general methods for arbitrary traits measured directly, and a specific yet generalizable extension for kinetic rate parameters measured indirectly in terms of single-cell gene expression profiles. We showed that the tree-structured correlation of the parameters plays a crucial role in the definition of appropriate estimators, and demonstrated performance in simulations directly related with real experiments from [11]. Our contributions are of direct interest to single-cell microscopy data, and they provide a step into the broader problem of treating snapshot measurements from tree-structured populations.

While our methods have been showcased relative to evolution of single (scalar) parameters (a scenario one may always get back to in case of uncorrelated parameters), they have been presented for vectors of correlated parameters. Simulation-based performance assessment of such scenarios is in progress. Application of the methods developed on the data in [11] and comparison of results from methods assuming lineage information is also in progress. This study will also address in practice the strongest assumption made, namely, that the generation that cells belong to is known. Unreported efforts show that the experimental data can be reconciled with the assumptions made, and first results are positive.

Beyond the biological case study, our contribution is of potential interest to any application with tree-structured data. Future directions of research include theoretical assessment of identifiability and estimation performance, extension of the methods to partially observed trees, and developments toward snapshot data not aligned with generations.

VI. ACKNOWLEDGMENTS

The authors thank Jakob Ruess for useful discussions.

APPENDIX

Proof: (Proposition 1) For every $v \in S_n$ and every $k \in \{0,\dots,n\}$, let $v^k \in S_k$ be the index of the only ancestor of v at generation k.

$$\Gamma_\theta(n+\ell,n) = \mathbb{E}\left[\frac{1}{|S_{n+\ell}|}\sum_{v\in S_{n+\ell}}\left(\sum_{k=1}^{n+\ell} A^{n+\ell-k}\boldsymbol{\eta}^{v^k}+\boldsymbol{\epsilon}^v\right)\right] \times \mathbb{E}\left[\frac{1}{|S_n|}\sum_{v\in S_n}\left(\sum_{k=1}^{n} A^{n-k}\boldsymbol{\eta}^{v^k}+\boldsymbol{\epsilon}^v\right)\right]^{\mathrm{T}}.$$

Using that $\boldsymbol{\epsilon}^v$ and $\boldsymbol{\eta}^v$ are of mean 0, independent across v, and independent from each other, and that the sets $(S_n, n \geq 0)$ are fixed, we obtain

$$\Gamma_\theta(n+\ell,n) = \frac{\delta_{0,\ell}}{|S_n|}R(n) + \frac{1}{|S_{n+\ell}||S_n|}\sum_{\substack{u\in S_{n+\ell}\\ v\in S_n}}\sum_{i=1}^{n+\ell}\sum_{j=1}^{n} A^{n+\ell-i}\mathbb{E}\left[\boldsymbol{\eta}^{u^i}\boldsymbol{\eta}^{v^{j\mathrm{T}}}\right]A^{n-j^{\mathrm{T}}}$$

Next, as $\boldsymbol{\eta}^v$ are independent across v and of mean 0, if $u^i \neq v^j$, we have $\mathbb{E}\left[\boldsymbol{\eta}^{u^i}\boldsymbol{\eta}^{v^{j\mathrm{T}}}\right] = 0$. Therefore,

$$\Gamma_\theta(n+\ell,n) = \frac{\delta_{0,\ell}}{|S_n|}R(n) + \frac{1}{|S_{n+\ell}||S_n|}\sum_{i=1}^{n}\sum_{\substack{u\in S_{n+\ell}\\ v\in S_n}} A^{n+\ell-i}\mathbb{E}\left[\boldsymbol{\eta}^{u^i}\boldsymbol{\eta}^{v^{i\mathrm{T}}}\right]A^{n-i^{\mathrm{T}}}.$$

To conclude, for all $i \in \{1,\dots,n\}$, we need to compute

$$D_i := \#\left\{(u,v)\in S_{n+\ell}\times S_n, \text{ such that } u^i = v^i\right\}.$$

For $i \in \{1,\dots,n\}$, as we consider a complete binary tree, we have 2^i choices for the common ancestor w at generation i, and 2^{n-i} (resp. $2^{n+\ell-i}$) choices for descendant of w at generation n (resp. $n+\ell$). Finally, $D_i = 2^{2n+\ell-i} = |S_n||S_{n+\ell}|2^{-i}$. Then, as for all v, $\mathbb{E}\left[\boldsymbol{\eta}^v\boldsymbol{\eta}^{v^{\mathrm{T}}}\right] = \Omega$, changing the indices in the sum, we get

$$\Gamma_\theta(n+\ell,n) = \frac{\delta_{0,\ell}}{|S_n|}R(n) + \frac{1}{2^{n+\ell}}\sum_{i=0}^{n-1} 2^i A^{i+\ell}\Omega A^{i^{\mathrm{T}}}.$$

■

Proof: (Proposition 2) We have

$$(|S_n|-1)\tilde{\boldsymbol{\Sigma}}(n) = \sum_{v\in S_n}\left(\tilde{\boldsymbol{\varphi}}^v - \tilde{\boldsymbol{\mu}}(n)\right)\left(\tilde{\boldsymbol{\varphi}}^v - \tilde{\boldsymbol{\mu}}(n)\right)^{\mathrm{T}}$$
$$= \frac{1}{|S_n|^2}\sum_{v\in S_n}\left(\sum_{u\in S_n}(\tilde{\boldsymbol{\varphi}}^v - \tilde{\boldsymbol{\varphi}}^u)\right)\left(\sum_{w\in S_n}(\tilde{\boldsymbol{\varphi}}^v - \tilde{\boldsymbol{\varphi}}^w)\right)^{\mathrm{T}}$$
$$= \frac{|S_n|-1}{|S_n|}\sum_{v\in S_n}\tilde{\boldsymbol{\varphi}}^v\tilde{\boldsymbol{\varphi}}^{v^{\mathrm{T}}} - \frac{1}{|S_n|}\sum_{\substack{u,v\in S_n\\ u\neq v}}\tilde{\boldsymbol{\varphi}}^v\tilde{\boldsymbol{\varphi}}^{u^{\mathrm{T}}}.$$

Next, for $u, v \in S_n$,

$$\mathbb{E}_\theta\left[\tilde{\boldsymbol{\varphi}}^v\tilde{\boldsymbol{\varphi}}^{u^{\mathrm{T}}}\right] = \mathbb{E}_\theta\left[\boldsymbol{\varphi}^v\boldsymbol{\varphi}^{u^{\mathrm{T}}}\right] + \delta_{u,v}R(n),$$

so that

$$|S_n|\mathbb{E}_\theta\left[\tilde{\boldsymbol{\Sigma}}(n)\right] = \sum_{v\in S_n}\mathbb{E}_\theta\left[\boldsymbol{\varphi}^v\boldsymbol{\varphi}^{v^{\mathrm{T}}}\right] + R(n) - \frac{1}{(|S_n|-1)}\sum_{\substack{u,v\in S_n\\ u\neq v}}\mathbb{E}_\theta\left[\boldsymbol{\varphi}^v\boldsymbol{\varphi}^{u^{\mathrm{T}}}\right]. \tag{18}$$

Recall that for every $v \in S_n$ and every $k \in \{0,\dots,n\}$, $v^k \in S_k$ denotes the index of the only ancestor of v at generation k. Then, combining

$$\boldsymbol{\varphi}^v = \boldsymbol{\mu}(n) + \sum_{k=1}^{n} A^{n-k}\boldsymbol{\eta}^{v^k}$$

with the fact that the $\boldsymbol{\eta}^v$ are of mean 0 and independent across v, we obtain

$$\mathbb{E}_\theta\left[\boldsymbol{\varphi}^v\boldsymbol{\varphi}^{u^{\mathrm{T}}}\right] = \boldsymbol{\mu}(n)\boldsymbol{\mu}(n)^{\mathrm{T}} + \sum_{k=1}^{|u\wedge v|} A^{n-k}\Omega A^{n-k^{\mathrm{T}}},$$

where $u\wedge v$ denotes the most recent common ancestor of u and v, and $|u\wedge v|$ its corresponding generation. Then,

$$\frac{1}{(|S_n|-1)|S_n|}\sum_{\substack{u,v\in S_n\\ u\neq v}}\mathbb{E}_\theta\left[\boldsymbol{\varphi}^v\boldsymbol{\varphi}^{u^{\mathrm{T}}}\right] = \boldsymbol{\mu}(n)\boldsymbol{\mu}(n)^{\mathrm{T}} + \frac{1}{(|S_n|-1)|S_n|}\sum_{\substack{u,v\in S_n\\ u\neq v}}\sum_{k=1}^{|u\wedge v|} A^{n-k}\Omega A^{n-k^{\mathrm{T}}}, \tag{19}$$

Next, for $j \in \{0,\dots,n-1\}$, as we consider a complete binary tree, we have

$$\sum_{\substack{u,v\in S_n\\ u\neq v}}\mathbf{1}_{|u\wedge v|=j} = |S_n|2^{n-j-1},$$

so that

$$\begin{aligned}
&\sum_{\substack{u,v\in S_n\\ u\neq v}}\sum_{k=1}^{|u\wedge v|} A^{n-k}\Omega A^{n-k^{\mathrm{T}}}\\
&= \sum_{j=0}^{n-1}\sum_{\substack{u,v\in S_n\\ u\neq v}}\mathbf{1}_{|u\wedge v|=j}\sum_{k=1}^{j} A^{n-k}\Omega A^{n-k^{\mathrm{T}}}\\
&= |S_n|\sum_{j=0}^{n-1}\sum_{k=1}^{j} 2^{n-j-1}A^{n-k}\Omega A^{n-k^{\mathrm{T}}}\\
&= |S_n|\sum_{k=1}^{n-1}(2^k-1)A^k\Omega A^{k^{\mathrm{T}}}.
\end{aligned} \tag{20}$$

Finally, combining (18), (19) and (20), we obtain

$$\mathbb{E}_\theta\left[\tilde{\boldsymbol{\Sigma}}(n)\right] = \Sigma_\theta(n) + R(n) - \frac{1}{(|S_n|-1)}\sum_{k=1}^{n-1}(2^k-1)A^k\Omega A^{k^{\mathrm{T}}}.$$

■

REFERENCES

[1] C. Aditya, F. Bertaux, G. Batt, and J. Ruess. Using single-cell models to predict the functionality of synthetic circuits at the population scale. *PNAS*, 119(11):e2114438119, March 2022.

[2] J. Bezanson, A. Edelman, S. Karpinski, and V. B. Shah. Julia: A Fresh Approach to Numerical Computing. *SIAM Review*, 59(1):65–98, 2017. _eprint: https://doi.org/10.1137/141000671.

[3] E. Y. Bijman, H.-M. Kaltenbach, and J. Stelling. Experimental analysis and modeling of single-cell time-course data. *Curr. Opin. Syst. Biol.*, 28:100359, December 2021.

[4] H. de Jong, C. Ranquet, D. Ropers, C. Pinel, and J. Geiselmann. Experimental and computational validation of models of fluorescent and luminescent reporter genes in bacteria. *BMC Syst. Biol.*, 2010.

[5] L. Duso and C. Zechner. Stochastic reaction networks in dynamic compartment populations. *PNAS*, 117(37):22674–22683, 2020.

[6] A. K. Gupta and D. K. Nagar. *Matrix Variate Distributions*. Chapman and Hall/CRC, 1999.

[7] J. Hasenauer, S. Waldherr, M. Doszczak, N. Radde, P. Scheurich, and F. Allgower. Identification of models of heterogeneous cell populations from population snapshot data. *BMC Bioinform.*, 12(1):125, 2011.

[8] H.K. Khalil. *Nonlinear Systems*. Pearson Education. Prentice Hall, 2002.

[9] M. Komorowski, B. Finkenstädt, C. Harper, and D. Rand. Bayesian inference of biochemical kinetic parameters using the linear noise approximation. *BMC Bioinform.*, 10(1):343, 2009.

[10] Marc Lavielle. *Mixed Effects Models for the Population Approach Models, Tasks, Methods and Tools*. Chapman and Hall/CRC, New York, 2014.

[11] A. Llamosi, A. M. Gonzalez-Vargas, C. Versari, E. Cinquemani, G. Ferrari-Trecate, P. Hersen, and G. Batt. What Population Reveals about Individual Cell Identity: Single-Cell Parameter Estimation of Models of Gene Expression in Yeast. *PLOS Comput. Biol.*, 12(2):e1004706, February 2016.

[12] A. Marguet, M. Lavielle, and E. Cinquemani. Inheritance and variability of kinetic gene expression parameters in microbial cells: modeling and inference from lineage tree data. *Bioinformatics*, 35(14):i586–i595, July 2019.

[13] MATLAB. *version 9.13 (R2022b)*. The MathWorks Inc., Natick, Massachusetts, 2010.

[14] B. Munsky, B. Trinh, and M. Khammash. Listening to the noise: Random fluctuations reveal gene network parameters. *Mol. Syst. Biol.*, 5(318), 2009.

[15] A. Raj and A. van Oudenaarden. Nature, Nurture, or Chance: Stochastic Gene Expression and Its Consequences. *Cell*, 135(2):216–226, 2008.

[16] S. Taheri-Araghi, S. Bradde, J. T. Sauls, N. S. Hill, P. A. Levin, J. Paulsson, M. Vergassola, and S. Jun. Cell-Size Control and Homeostasis in Bacteria. *Curr. Biol.*, 25(3):385–391, February 2015.

[17] P. Thomas. Making sense of snapshot data: ergodic principle for clonal cell populations. *J. R. Soc. Interface*, 14(136):20170467, November 2017.

[18] N. Totis, C. Nieto, A. Kuper, C. Vargas-Garcia, A. Singh, and S. Waldherr. A Population-Based Approach to Study the Effects of Growth and Division Rates on the Dynamics of Cell Size Statistics. *IEEE Control Syst. Lett.*, 5(2):725–730, April 2021.

[19] S. Waldherr. Estimation methods for heterogeneous cell population models in systems biology. *J. R. Soc. Interface*, 15(147):20180530, October 2018.

[20] C. Zechner, J. Ruess, P. Krenn, S. Pelet, M. Peter, J. Lygeros, and H. Koeppl. Moment-based inference predicts bimodality in transient gene expression. *PNAS*, 109(21):8340–8345, May 2012.

[21] C. Zechner, M. Unger, S. Pelet, M. Peter, and H. Koeppl. Scalable inference of heterogeneous reaction kinetics from pooled single-cell recordings. *Nat. Methods*, 11:197–202, 2014.