2013 European Control Conference (ECC)
July 17-19, 2013, Zürich, Switzerland.

Marginal Dynamics of Stochastic Biochemical Networks in Random Environments

C. Zechner, S. Deb, and H. Koeppl

Abstract— **Stochastic simulation algorithms provide a powerful means to understand complex biochemical processes as well as to solve the inverse problem of reconstructing hidden states and parameters from experimental single-cell data. At presence, a repertoire of efficient algorithms for simulating and calibrating stochastic reaction networks is available. However, most of these approaches do not account for the fact that each cell of a clonal population is exposed to a random extrinsic environment, i.e., the agglomerate of so-called extrinsic factors such as cell size, shape or cell cycle stage. We recently proposed a dynamic description of stochastic chemical kinetics in random but unknown extrinsic environments, reflected by a stochastic process where uncertain parameters are marginalized out. In this work we further investigate that process and provide additional analytical results. We demonstrate the marginalization using several biologically relevant parameter distributions and derive exact waiting-time distributions. We further show that the marginalized process model can achieve a variance reduction in the context of parameter inference.**

I. Introduction

Single-cell technologies in molecular biology offer an unprecedented view on biochemical processes happening inside the living cell. In particular, readouts based on fluorescent proteins or bioluminescence in conjunction with time-lapse microscopy provide rich data for building quantitative single-cell models. At this resolution the heterogeneity within a clonal population of cells becomes apparent [1], [2], [3]. On the one hand the random character of chemical reactions contributes to this heterogeneity, on the other hand the clonal cells are not perfectly identical due to so-called extrinsic factors [4], [5]. More generally, one may say that the particular biochemical process to be studied operates within a molecular environment that is different for every cell [6], [7].

A natural question to ask is whether the biochemical process can be marginalized with respect to this random molecular environment – giving rise to a dynamic model which inherently accounts for that extrinsic uncertainty. The innovation theorem for counting processes provides means to construct a marginalized counting process [8], [9], where the uncertainty is directly encoded into the dynamics that now turns out to be time-inhomogeneous. This surprising result is well known in traditional sampling models. For instance, a Bernoulli trial process with a Beta-distributed random success rate can be shown to be an equivalent to a Pólya urn process where with each draw, a new ball of the same color is added to the urn together with the draw. It is immediate that the latter process is time-inhomogeneous in resemblance to the desired general description of a marginal process.

We recently developed a marginal counterpart of traditional stochastic chemical kinetics that – in addition to the molecular stochasticity – implicitly accounts for the random extrinsic environment [10]. In that work, such processes were applied to the practical problem of reconstructing hidden states and parameters from experimental microscopy data. The main idea of this paper is to follow up on [10] and provide a more detailed perspective on the marginal process framework. More specifically, we derive explicit process constructions for several distributions of random extrinsic factors. We also demonstrate how such a process description can be used to improve the statistical inference of model parameters in the context of applied stochastic modeling.

The remaining part of the paper is organized as follows. In Section II we introduce the traditional continuous-time Markov chain (CTMC) framework for modeling chemical kinetics and demonstrate how random extrinsic environments can be included into the description. In Section III we derive the marginal process and provide necessary proofs and several analytical examples. Finally, we demonstrate correctness and usefulness of the approach within a few simulation studies in Section IV.

II. Stochastic Chemical Kinetics

A. Notation

Before we introduce the mathematical formalism, we briefly discuss the basic notation used throughout the paper. All random variables and their realizations are denoted by upper-case and lower-case symbols, respectively. Furthermore, we use different symbols to indicate densities (i.e., p) and probabilities (i.e., P). Conditional expectations are expressed as $\mathbb{E}[A \mid b, c] = \int_{\mathcal{A}} ap(a \mid b, c)\mathrm{d}a$, with $\mathcal{A}$ as the support of random variable A, while b and c denote realizations of random variables B and C, respectively. Moreover, we use bold symbols and subscripts for time-dependent quantities on intervals $[a, b]$, whereas the subscript is dropped if $a = 0$ and $b = T$, e.g., $\mathbf{f}_{[a,b]} = \{f(t) \mid t \in [a, b]\}$ and $\mathbf{f} \equiv \mathbf{f}_{[0,T]}$.

B. Continuous-Time Markov Chains

We describe the temporal dynamics of a reaction network through a continuous-time Markov chain (CTMC), accounting for the discrete randomness of molecular binding and modification events. We assume that the system under study involves d chemical species, whose abundances at time t are collected in $X(t) \in \mathbb{R}^d$. Moreover, the system is associated with a set of L reaction channels with corresponding rate

C. Zechner, S. Deb and H. Koeppl are with the Automatic Control Lab., Department of Information Technology and Electrical Engineering, ETH Zurich, Switzerland. H. Koeppl is also with IBM Zurich Research Laboratory, Rueschlikon, Switzerland. {zechner,koeppl}@control.ee.ethz.ch

constants $C = \{C_i \mid i = 1, \ldots, L\}$. Then, the probability that reaction i fires within a sufficiently small $\mathrm{d}t$ is given by

$$\begin{aligned} &P(X(t+\mathrm{d}t) = x + \nu_i \mid X(t) = x, C_i = c_i) \\ &\quad = h(x, c_i)\mathrm{d}t = c_i g_i(x)\mathrm{d}t \end{aligned} \tag{1}$$

with $\nu_i \in \mathbb{R}^d$ as the stoichiometric change vector for reaction i and g_i as a polynomial in x given by the law of mass-action, for instance. Furthermore, we denote by $\mathbf{X}$ the process $X(t)$ on the time interval $[0, T]$.

C. Path Likelihood Functions

Assume now that we have full access to a realization $\mathbf{x}$ of $\mathbf{X}$. Estimating the kinetic parameters C given $\mathbf{x}$ requires knowledge of the path likelihood function, which is given by [11], [12]

$$p(\mathbf{x} \mid c) = \prod_{i=1}^{L} c_i^{r_i} \exp\left\{-c_i \int_0^T g_i(x(s))\mathrm{d}s\right\}, \tag{2}$$

with r_i as the number of occurrences of reaction i in $\mathbf{x}$. According to a Bayesian inference procedure, C is associated with a prior distribution $p(c)$. In case of independent Gamma-distributed priors, i.e., $p(c) = \prod_{i=1}^{L} \mathcal{G}(\alpha_i, \beta_i)$, also the posterior distribution will be a product of independent Gamma distributions, i.e.,

$$p(c \mid \mathbf{x}) = \prod_{i=1}^{L} \mathcal{G}\left(\alpha_i + r_i, \beta_i + \int_0^T g_i(x(s))\mathrm{d}s\right). \tag{3}$$

D. Modeling Heterogeneous Populations

We model the influence of certain extrinsic factors (e.g., cell size) by means of certain kinetic parameters that randomly vary across a population of cells. Without loss of generality, we assume that the first I parameters in C (denoted $S = \{C_i \mid i = 1, \ldots, I\}$) are fixed, i.e., shared among cells, while the subsequent $J = L - I$ parameters $E = \{C_i \mid i = I+1, \ldots, L\}$ randomly vary accross the cell population. The individual set of parameters for the m-th cell is then denoted $C^m = \{S, E^m\}$, with $E^m \mid (A = a) \sim p(e \mid a)$ and A as a set of shape parameters or hyperparameters.

III. A MARGINAL PROCESS FRAMEWORK

In the context of simulation, a process description revealing the dynamics of a single cell with unknown extrinsic environment (i.e., E^m) appears natural. Mathematically, this can be accomplished by a marginal jump process, where the individual extrinsic parameters have been integrated out. Such constructions can be obtained via the *innovation theorem* for counting processes [8]. While a rigorous treatment of the latter involves more sophisticated martingale theory, an intuitive derivation is straight-forward. Although similar derivations can be found in [10] or [9], the key steps are outlined in the following for the sake of completeness.

A. The Innovation Theorem For Chemical Kinetics

Theorem 3.1: The marginal reaction hazards corresponding to the extrinsic variables E are given by

$$h_i(\mathbf{x}_{[0,t]}, t) = \mathbb{E}\left[E_j \mid \mathbf{x}_{[0,t]}, a\right] g_i(x(t)) \tag{4}$$

for $j = i - I$ and $i = I+1, \ldots, L$.

Proof: For simplicity assume $E \in \mathcal{E}^J$ is a continuous random variable. Then, the marginal jump probability for reaction i becomes

$$\begin{aligned} &P\left(X(t+\mathrm{d}t) = x(t) + \nu_i \mid \mathbf{x}_{[0,t]}, a\right) \\ &\quad = \int_{\mathcal{E}} P(X(t+\mathrm{d}t) = x + \nu_i, e_j \mid \mathbf{x}_{[0,t]}, a)\mathrm{d}e_j \\ &\quad = \int_{\mathcal{E}} P(X(t+\mathrm{d}t) = x + \nu_i \mid x(t), e_j) p(e_j \mid \mathbf{x}_{[0,t]}, a)\mathrm{d}e_j \\ &\quad = \left[\int_{\mathcal{E}} e_j p(e_j \mid \mathbf{x}_{[0,t]}, a)\mathrm{d}e_j\right] g_i(x(t))\mathrm{d}t \\ &\quad = \mathbb{E}\left[E_j \mid \mathbf{x}_{[0,t]}, a\right] g_i(x(t))\mathrm{d}t \\ &\quad = h_i(\mathbf{x}_{[0,t]}, t)\mathrm{d}t. \end{aligned}$$

■

Corollary 3.2: Following [9], the marginal reaction hazards can be rewritten as

$$\begin{aligned} &h_i(\mathbf{x}_{[0,t]}, t) = \mathbb{E}\left[E_j \mid \mathbf{x}_{[0,t]}, a\right] g_i(x(t)) \\ &\quad = \left(\int_{\mathcal{E}} e_j p(e_j \mid \mathbf{x}_{[0,t]}, a)\mathrm{d}e_j\right) g_i(x(t)) \\ &\quad = \left(\frac{\int_{\mathcal{E}} e_j p(\mathbf{x}_{[0,t]} \mid e_j) p(e_j \mid a)\mathrm{d}e_j}{p(\mathbf{x}_{[0,t]} \mid a)}\right) g_i(x(t)) \\ &\quad = \frac{\int_{\mathcal{E}} e_j \int_{\mathcal{E}} \cdots \int_{\mathcal{E}} p(\mathbf{x}_{[0,t]} \mid e) p(e \mid a)\mathrm{d}e_1 \cdots \mathrm{d}e_J}{\int_{\mathcal{E}^J} p(\mathbf{x}_{[0,t]} \mid e) p(e \mid a)\mathrm{d}e} g_i(x(t)) \\ &\quad = \frac{\mathbb{E}\left[e_j p(\mathbf{x}_{[0,t]} \mid e) \mid a\right]}{\mathbb{E}\left[p(\mathbf{x}_{[0,t]} \mid e) \mid a\right]} g_i(x(t)). \end{aligned} \tag{5}$$

Together with (2) we recognize the last expression as a function of the Laplace transform of $p(e \mid a)$, i.e.,

$$h_i(\mathbf{x}_{[0,t]}, t) = -\frac{\mathcal{L}^{r_{I+1}, \ldots, r_i+1, \ldots, r_L}(\eta)}{\mathcal{L}^{r_{I+1}, \ldots, r_L}(\eta)} g_i(x(t)) \tag{6}$$

with

$$\begin{aligned} \eta &= [\eta_1, \ldots, \eta_J] \\ &= \left[\int_0^t g_{I+1}(x(s))\mathrm{d}s, \ldots, \int_0^t g_L(x(s))\mathrm{d}s\right], \end{aligned}$$

r_i as the number of reactions of type i and $\mathcal{L}^{q_1, \ldots, q_J}$ as the partial derivative of the Laplace transform $\mathcal{L}$ w.r.t. $\eta_1, \ldots, \eta_J$ taken $q_1, \ldots, q_J$ times, respectively. Note that it is straight forward to show that (4) also applies if E is a discrete random variable. In that case, (6) becomes a function of the moment generating function (MGF) of E, i.e.,

$$h_i(\mathbf{x}_{[0,t]}, t) = \frac{\mathcal{M}^{r_{I+1}, \ldots, r_i+1, \ldots, r_L}(\zeta)}{\mathcal{M}^{r_{I+1}, \ldots, r_L}(\zeta)} g_i(x(t)) \tag{7}$$

with

$$\zeta = \left[-\int_0^t g_{I+1}(x(s))\mathrm{d}s, \ldots, -\int_0^t g_L(x(s))\mathrm{d}s\right].$$

Example 3.1 (Univariate Gamma Distribution): Besides its convenient analytical properties, the Gamma distribution was shown to be widely applicable in molecular biology, especially in the context of stochastic gene expression [13].

Hence, we consider the case where the extrinsic variable E corresponding to reaction i is a univariate random variable with $p(e \mid a) = \mathcal{G}(\alpha, \beta)$. We know that for the Gamma distribution, the Laplace transform is given by

$$\mathcal{L}(\eta) = \frac{\beta^\alpha}{(\beta+\eta)^\alpha} \tag{8}$$

and its r_i-th derivative turns out to be

$$\mathcal{L}^{r_i}(\eta) = (-1)^{r_i}\frac{(\alpha+r_i-1)!}{(\alpha-1)!}\frac{\beta^\alpha}{(\beta+\eta)^{\alpha+r_i}}. \tag{9}$$

Then, using (6) the marginal hazard function is found to be

$$h_i(\mathbf{x}_{[0,t]}, t) = \frac{\alpha+r_i}{\beta+\int_0^t g_i(x(s))\mathrm{d}s} g_i(x(t)). \tag{10}$$

Alternatively, according to (4) we note that (10) is the conditional expectation of E given a complete sample path $\mathbf{x}_{[0,t]}$. In a similar way, explicit formulas can be calculated for multivariate generalizations of the Gamma distribution [14].

Example 3.2 (Poisson Distribution): In many practical scenarios it might be the case that the reaction hazards are modulated by a discrete-valued extrinsic variable, e.g., due to a variability in copy numbers of certain species. For instance, if E follows a Poisson distribution with rate parameter λ, the MGF is given by

$$\mathcal{M}(\zeta) = exp\{\lambda(e^\zeta - 1)\} \tag{11}$$

and its j-th derivative is found to be

$$\mathcal{M}^j(\zeta) = \mathcal{M}(\zeta)(\sum_{k=1}^{j} d_{j,k}(\lambda e^\zeta)^j) \tag{12}$$

with the coefficients $d_{j,k}$ following the recursive relation

$$\begin{aligned} d_{j+1,1} &= 1 \\ d_{j+1,k} &= d_{j,k-1} + k d_{j,k} \;\forall\, k = 2,\ldots,j \\ d_{j+1,j+1} &= 1 \end{aligned} \tag{13}$$

The marginal hazard function can then be written as

$$h_i(\mathbf{x}_{[0,t]}, t) = \frac{\sum_{k=1}^{r_i+1} d_{r_i+1,k}(\lambda e^\zeta)^k}{\sum_{k=1}^{r_i} d_{r_i,k}(\lambda e^\zeta)^k} g_i(x(t)). \tag{14}$$

Example 3.3 (Bernoulli Distribution): Consider a cell population that is randomly partitioned into two subgroups. For some reason (maybe due to different cell-cycle stages), one group can express a certain gene efficiently while the other group cannot. This can be modeled by assuming that the corresponding kinetic rate constants are drawn from a Bernoulli distribution such that each cell's rate constant is $E = \kappa_0$ with probability p or $E = \kappa_1$ with probability $q = 1 - p$. In this case, the MGF is given by

$$\mathcal{M}(\zeta) = pe^{\zeta\kappa_0} + qe^{\zeta\kappa_1}. \tag{15}$$

Taking the r_i-th and the (r_i+1)-th derivative yields the marginal reaction hazard

$$h_i(\mathbf{x}_{[0,t]}, t) = \frac{pe^{\zeta\kappa_0}\kappa_0^{r_i+1} + qe^{\zeta\kappa_1}\kappa_1^{r_i+1}}{pe^{\zeta\kappa_0}\kappa_0^{r_i} + qe^{\zeta\kappa_1}\kappa_1^{r_i}} g_i(x(t)). \tag{16}$$

Note that the above derivations are straight-forward to extend for discrete probability distributions with more than two outcomes.

B. Waiting-Time Distributions

It is well know that for homogeneous reaction channels, the waiting-times until the next event occurs are exponentially distributed, i.e., if t_k denotes the time of the last reaction and $t_{k+1} > t_k$ the next time point when reaction i fires, we have

$$t_{k+1} - t_k \sim \mathrm{Exp}\left(c_i g_i(x(t_k))\right). \tag{17}$$

It has been shown that non-exponential waiting-time distributions can be observed in biochemical processes involving multi-step events (e.g., in mRNA transcription or reporter maturation) [15]. However, this can also be the case for single reactions that are exposed to a random extrinsic environment. This can be understood from the fact that the marginal process describing the heterogeneous population is non-Markovian and that its propensity functions are explicitly time-varying. The cumulative waiting-time distribution can be found such as for a non-homogeneous Markov process by analyzing

$$\begin{aligned} &P(t_{k+1} - t_k > t) \\ &\quad = 1 - \exp\left\{-\int_0^t h_i(\mathbf{x}_{[0,t_k+s]}, t_k+s)\mathrm{d}s\right\}. \end{aligned} \tag{18}$$

Example 3.4 (Waiting-Time Distribution): The waiting-time distribution at time t_k for the i-th reaction with Gamma-type heterogeneity is given by

$$\begin{aligned} &P(t_{k+1} - t_k > t) \\ &\quad = 1 - \exp\left\{-\int_0^t h_i(\mathbf{x}_{[0,t_k+s]}, t_k+s)\mathrm{d}s\right\} \\ &\quad = 1 - \exp\left\{-(\alpha+r_i)\left(\ln\left[\beta + \int_0^{t_k+t} g_i(x(s))\mathrm{d}s\right]\right.\right. \\ &\qquad \left.\left. - \ln\left[\beta + \int_0^{t_k} g_i(x(s))\mathrm{d}s\right]\right)\right\} \end{aligned} \tag{19}$$

and with $G_i(t_k) := \int_0^{t_k} g_i(x(s))\mathrm{d}s$ we arrive at

$$\begin{aligned} &P(t_{k+1} - t_k > t) \\ &\quad = 1 - \exp\left\{-(\alpha+r_i)\Big(\ln\left[\beta + G_i(t_k) + t g_i(x(t_k))\right]\right. \\ &\quad \left.\left. - \ln\left[\beta + G_i(t_k)\right]\Big)\right\} = 1 - \left[\frac{\frac{\beta+G_i(t_k)}{g_i(x(t_k))}}{\frac{\beta+G_i(t_k)}{g_i(x(t_k))} + t}\right]^{(\alpha+r_i)}, \end{aligned} \tag{20}$$

which we recognize as the CDF of a Pareto Type II - or "Lomax" distribution such that

$$t_{k+1} - t_k \sim \text{Lomax}\left(\frac{\beta + G_i(t_k)}{g_i(x(t_k))}, \alpha + r_i\right). \tag{21}$$

The mean waiting-time is given by

$$\mathbb{E}\left[t_{k+1} - t_k\right] = \frac{\beta + G_i(t_k)}{g_i(x(t_k))(\alpha + r_i - 1)}, \tag{22}$$

which for large α and β, converges to $\frac{\beta}{g_i(x(t_k))\alpha}$, which corresponds to the mean waiting-time obtained for a homogeneous reaction with rate constant $\frac{\alpha}{\beta}$.

C. Variance Reduction Through Marginalization

Stochastic biochemical models are often governed by non-linearities such that analytical characterization of the system dynamics becomes tedious. In those cases, stochastic simulation algorithms provide a powerful means to both analyze and reconstruct such systems without relying on systematic approximations. In the context of system analysis, one is typically interested in assessing expectations of certain functions of the process $X(t)$ such as its moments or sensitivity with respect to another quantity. In contrast, reverse-engineering problems often deal with estimating a-posteriori statistics of the process parameters conditional on measurements. In both cases, many Monte Carlo samples might be necessary to obtain estimates with reasonably low variance. Based on a simple parameter inference scenario, we demonstrate in the following how the proposed process marginalization can be exploited to achieve variance reduction.

In order to avoid the numerous technicalities that arise when dealing with realistic measurement models, we assume a simplified complete-observation inference scenario. However, the reader should note that the considerations below also extend to the realistic case of noisy and incomplete measurements [10]. More specifically, we assume that the measurements consist of M complete sample paths $\mathbf{x}^1, \dots, \mathbf{x}^M$. Remember that in this case, the posterior distributions over the fixed kinetic parameters are simply given by (3). Hence, we exclude these parameters from the following analysis and focus on the inference of $E^1, \dots, E^M$ and A. Hence, the desired posterior distribution is given by

$$p(e^1, \dots, e^M, a \mid \mathbf{x}^1, \dots, \mathbf{x}^M) \propto \left(\prod_{m=1}^{M} p(\mathbf{x}^m \mid e^m) p(e^m \mid a)\right) p(a). \tag{23}$$

Note that due to analytical intractability, a suitable sampling or approximation algorithm needs to be used in order to numerically determine (23). However, in case of large M, such approaches will typically suffer from the increasing dimensionality of the latent parameter space. We now pose the question, whether the posterior distribution itself can be marginalized over the extrinsic variables $E^1, \dots, E^M$. Formally we can write

$$\begin{aligned} p(a \mid \mathbf{x}^1, \dots, \mathbf{x}^M) &= \\ &= \int_{\mathcal{E}^J} \cdots \int_{\mathcal{E}^J} p(e^1, \dots, e^M, a \mid \mathbf{x}^1, \dots, \mathbf{x}^M) \mathrm{d}e^1 \cdots \mathrm{d}e^M \\ &\propto \left(\prod_{m=1}^{M} p(\mathbf{x}^m \mid a)\right) p(a), \end{aligned} \tag{24}$$

with $p(\mathbf{x}^m \mid a)$ as the path likelihood function of the marginal process. It was shown in [10] that if the kinetic parameter corresponding to reaction i is subject to a Gamma-type heterogeneity, i.e., $E \mid (A = a) \sim p(e \mid a) = \mathcal{G}(\alpha, \beta)$ with $a = \{\alpha, \beta\}$, this path likelihood function has the form

$$p(\mathbf{x} \mid \alpha, \beta) = \frac{\beta^\alpha \Gamma(\alpha + r_i)}{\Gamma(\alpha)} \exp\left\{-\int_0^t h_i(\mathbf{x}_{[0,s]}, s) \mathrm{d}s\right\}, \tag{25}$$

with $h_i(\mathbf{x}_{[0,t]}, t)$ given by (10). In conjunction with (25), (24) can be evaluated analytically and thus, samples can be efficiently drawn, e.g., using a standard Metropolis-Hastings sampler. We further note that

$$\begin{aligned} &p(e^1, \dots, e^M, a \mid \mathbf{x}^1, \dots, \mathbf{x}^M) \\ &\quad = p(e^1, \dots, e^M \mid \mathbf{x}^1, \dots, \mathbf{x}^M, a) p(a \mid \mathbf{x}^1, \dots, \mathbf{x}^M) \\ &\quad = \left(\prod_{m=1}^{M} p(e^m \mid \mathbf{x}^m, a)\right) p(a \mid \mathbf{x}^1, \dots, \mathbf{x}^M) \end{aligned} \tag{26}$$

and hence, the posterior distribution over the extrinsic variables could be obtained by first running a marginal inference scheme in order to draw samples $a^{(k)} \sim p(a \mid \mathbf{x}^1, \dots, \mathbf{x}^M)$ and subsequently determining the full conditional distributions $p(e^m \mid \mathbf{x}^m, a^{(k)})$[1].

Assume now we are interested in estimating the posterior statistic $\theta = \mathbb{E}\left[f \mid \mathbf{x}^1, \dots, \mathbf{x}^M\right]$, with $f : \mathcal{A} \times \mathcal{E}^{J \times M} \to \mathbb{R}$ as a function of A and $E^1, \dots, E^M$. Drawing N i.i.d. samples from the posterior distribution $p(a, e^1, \dots, e^M \mid \mathbf{x}^1, \dots, \mathbf{x}^M)$, such an estimate can be computed as $\hat{\theta} = N^{-1} \sum_{k=1}^{N} f^{(k)}$. We are interested in the estimators variance

$$\text{Var}\left[\hat{\theta}\right] = \frac{1}{N} \text{Var}\left[f\right], \tag{27}$$

which for a fixed N only depends on the variance of f. Applying the law of total variance we further obtain

$$\text{Var}\left[f\right] = \text{Var}\left[\mathbb{E}\left[f \mid a\right]\right] + \mathbb{E}\left[\text{Var}\left[f \mid a\right]\right]. \tag{28}$$

The term on the l.h.s. is obtained in case the inference is performed on the full latent space $\mathcal{A} \times \mathcal{E}^{J \times M}$. In contrast, when samples are drawn from the marginal distribution $p(a \mid \mathbf{x}^1, \dots, \mathbf{x}^M)$ and the conditional expectation $\mathbb{E}\left[f \mid a\right]$ can be determined, the variance is given by the first term on the r.h.s. of (28). Given that $\mathbb{E}\left[\text{Var}\left[f \mid a\right]\right] \geq 0$ we expect a variance reduction in $\hat{\theta}$.

[1] We know from eq. (3) that in case of a Gamma-type heterogeneity, the full conditional distributions are Gamma as well.

IV. Results

In the following we provide simulation results for the forward problem of simulating biochemical networks in random environments, as well as the inverse problem of estimating heterogeneous parameters from complete sample-path measurements. All studies built upon the simple mass-action model depicted in Fig.1.

$$\varnothing \underset{s_1}{\overset{e^m}{\rightleftarrows}} \mathsf{A} \qquad \mathsf{A} \xrightarrow{s_2} \mathsf{A} + \mathsf{B} \qquad \mathsf{B} \xrightarrow{s_3} \varnothing$$

Fig. 1. Heterogeneous mass-action model. We assume a Gamma-type heterogeneity in the production rate of species *A*, i.e., $e^m \sim \mathcal{G}(\alpha, \beta)$ for the m-th cell. All other reactions, i.e., those corresponding to s_{1-3} are assumed to be shared among cells.

A. Simulation of Marginal Dynamics

Stochastic simulation of the marginal process dynamics turns out to be straight forward using available methods that can cope with the time-inhomogeneity of the marginal hazard functions, such as the *first reaction method* [16]. We simulated the marginal dynamics of the model from Fig.1 with $\alpha = 4$, $\beta = 20$, $s_1 = 0.03s^{-1}$, $s_2 = 0.008s^{-1}$ and $s_3 = 0.001s^{-1}$ for $M = 2$ cells. Fig.2A shows the levels of species B and the corresponding marginal hazard functions $h(\mathbf{x}^m_{[0,t]}, t)$[2] that were computed during simulation. The dashed lines indicate the homogeneous hazard functions under knowledge of the extrinsic environments E^1 and E^2. Both marginal hazard functions start at $h(x^m_0, 0) = \frac{\alpha}{\beta}$ and subsequently converge to the homogeneous hazards, i.e., $h(\mathbf{x}^m_{[0,t]}, t) \to e^m g(x^m(t))$. Intuitively, this can be understood as the process continuously estimating its own intensity from the past. Consequently, due to the decreasing uncertainty of those estimates, each process will adapt to a different steady state intensity.

Next, we evaluated the waiting-time distribution derived in Example 3.4. In particular, we computed the waiting-time CDF for Gamma-type heterogeneities (i.e., Lomax distributions) with different coefficients of variation (CV) (see Fig.2B). The dashed line corresponds to the case without extrinsic variability (i.e., a CV of zero) such that the Lomax distribution collapses to an exponential distribution. Interestingly, we observe that the waiting-time distribution tends to have heavier tails under presence of heterogeneity.

B. Marginal Inference

Again, we consider the model from Fig.1 with $\alpha = 4$, $\beta = 50$, $s_1 = 0.03s^{-1}$, $s_2 = 0.1s^{-1}$ and $s_3 = 0.001s^{-1}$. We assume knowledge of the fixed kinetic parameters s_{1-3}, whereas α and β are to be estimated from $M = 100$ heterogeneous sample paths. We used a standard Metropolis-Hastings algorithm with lognormal proposal distributions in order to sample from the marginal posterior distribution $p(\alpha, \beta \mid \mathbf{x}^1, \ldots, \mathbf{x}^M)$. In particular, with $[\tilde{\alpha}, \tilde{\beta}]$ as the current

[2]Note that we drop the reaction index for h, g as well as the reaction counter r since there is only one heterogeneous reaction.

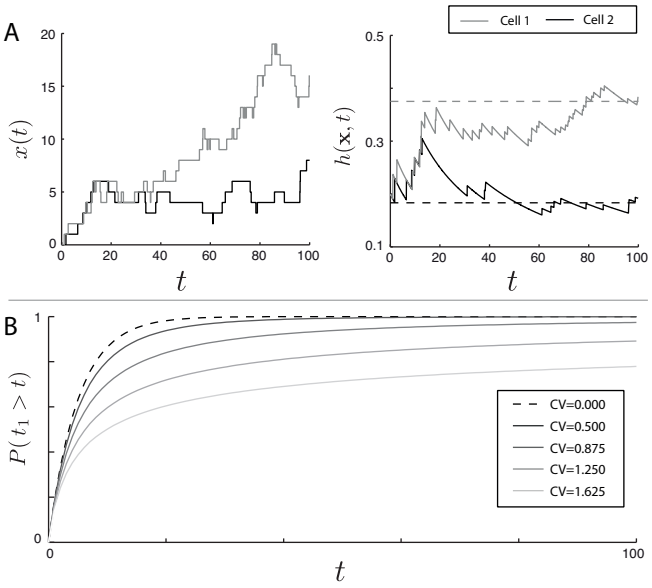

Fig. 2. Marginal dynamics. (A) Process dynamics and marginal hazard functions for two different cells. The state $x(t)$ refers to the abundance of species *B*. The dashed lines correspond to hazard functions that would have been obtained under knowledge of the individual extrinsic variables E^1 and E^2. Note that as the true values of the extrinsic variables cannot be revealed for finite t, the dashed lines were set to their MMSE values which were extracted from the sample paths. (B) Cumulative waiting-time distributions. The plot depicts the Lomax waiting-time distributions for several different CVs at time $t = 0$. The dashed line corresponds to an exponential distribution, which is obtained in the case of homogeneous kinetics (i.e., a CV of zero).

state of the Markov chain, we propose new samples as $[\alpha, \beta] \sim \mathcal{LN}\left([\ln\tilde{\alpha}, \ln\tilde{\beta}], \sigma^2 I\right)$ with $\sigma^2 = 0.04$. Fig.3A shows the Markov chain over $5e4$ update iterations for α and β separately. After discarding the samples corresponding to the initial burn-in, we computed density plots for α and β and also transformations thereof, i.e., the mean $\mu_E = \frac{\alpha}{\beta}$ and the variance $\sigma^2_E = \frac{\alpha}{\beta^2}$ of $\mathcal{G}(\alpha, \beta)$ (see Fig.3B).

C. Variance Reduction

According to (28), we expect a variance reduction when estimating $E^1, \ldots, E^M$ and A (or functions thereof) using the marginal inference framework. In order to verify that claim, we numerically estimated the expectation of a test function f using the full, as well as the marginal inference procedure for the model from Fig.1. As a test function we used

$$f(E^1, \ldots, E^M) = \frac{\sum_{m=1}^M E^m}{M}, \tag{29}$$

for which the conditional expectation becomes

$$\begin{aligned}
&\mathbb{E}\left[f \mid \alpha, \beta\right] \\
&= \int_{\mathcal{E}} \cdots \int_{\mathcal{E}} f p(e^1, \ldots, e^M \mid \mathbf{x}^1, \ldots, \mathbf{x}^M, \alpha, \beta) \mathrm{d}e^1 \cdots \mathrm{d}e^M \\
&= \frac{1}{M} \sum_{m=1}^M \int_{\mathcal{E}} e^m p(e^m \mid \mathbf{x}^m, \alpha, \beta) \\
&= \frac{1}{M} \sum_{m=1}^M \frac{\alpha + r^m}{\beta + \int_0^t g(x^m(s)) \mathrm{d}s},
\end{aligned} \tag{30}$$

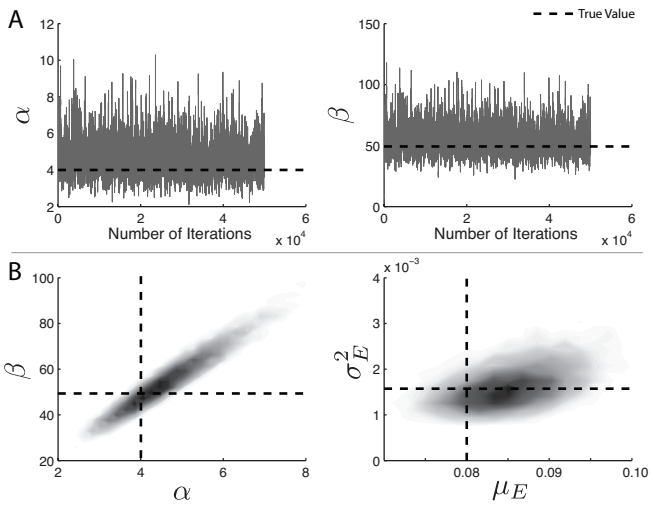

Fig. 3. Marginal inference of the hyperparameters α and β. (A) Realized Markov chains delivered by the Metropolis-Hastings algorithm for $5e4$ update iterations. (B) Joint density plots for α and β as well as μ_E and σ_E^2 (see text for further details). The dashed lines indicate the corresponding true values.

with r^m as the number of birth reactions of species A observed for the m-th cell. Hence, we computed the improved estimate $\tilde{\theta} = \sum_{k=1}^{K} \mathbb{E}\left[f \mid \alpha^{(k)}, \beta^{(k)}\right]$, whereas $\beta^{(k)}$ and $\alpha^{(k)}$, were obtained using the same Metropolis-Hastings[3] algorithm as described previously. Furthermore, we ran an MCMC scheme on the full latent space, in order to compute the naive estimate. For both estimators we determined the mean square error (MSE) and it's standard error using 2000 independent inference runs for different numbers of cells M. Fig.4 shows the individual MSE estimates for both inference schemes as well as the relative MSE decrease of the marginal compared to the full inference scheme.

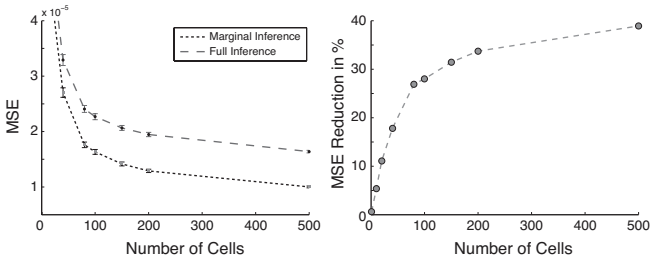

Fig. 4. MSE reduction via marginalization. The left panel shows the estimated MSEs and their standard errors as a function of M. The relative improvement of the MSE is depicted in the right panel. Each data point was computed using 2000 independent runs of the MCMC schemes.

Note that in theory, the same variance reduction could be achieved from the full inference scheme, if all dimensions of the Markov chain except those corresponding to α and β are discarded, in order to evaluate the conditional expectation from (30). However, for large M, sampling from the full parameters space becomes challenging, especially if the scheme is extended to the realistic case of incomplete and noisy measurements. Hence, the proposed marginalization not only profits from a systematic reduction in an estimators variance, but even appears central for establishing scalability.

[3]Note that since MCMC schemes deliver non-i.i.d. samples, (28) does not apply in this case. However, it does not change the fact that the marginal inference scheme yields reduced variance.

V. CONCLUSIONS

State-of-the-art single-cell techniques allow to measure several hundreds or thousands of presumably identical cells, offering massive data for quantitative modeling or reverse engineering. In this paper we argue that in order to unleash the power of such data, care must be taken with respect to the exposed variability among cells in a population. We recently proposed a marginal modeling framework for chemical kinetics in random environments and demonstrated its practical use in analyzing experimental microscopy data [10]. Here we provide further analytical results on the marginal stochastic process. In particular, we derive the resulting jump process for several practical distributions of the random environment and show how waiting-time distributions can be derived. We finally demonstrate how the marginalization scheme can be used to achieve variance reduction in the context of parameter inference.

REFERENCES

[1] M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain, "Stochastic gene expression in a single cell.," *Science (New York, N.Y.)*, vol. 297, no. 5584, pp. 1183–6, Aug. 2002.

[2] A. Colman-Lerner, A. Gordon, E. Serra, T. Chin, O. Resnekov, D. Endy, C. G. Pesce, and R. Brent, "Regulated cell-to-cell variation in a cell-fate decision system.," *Nature*, vol. 437, no. 7059, pp. 699–706, Sept. 2005.

[3] J. M. Raser and E. K. O'Shea, "Control of stochasticity in eukaryotic gene expression," *Science*, vol. 304, no. 5678, pp. 1811–1814, 2004.

[4] 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 Bioinformatics*, vol. 12, no. 1, pp. 125, 2011.

[5] C. Zechner, J. Ruess, P. Krenn, S. Pelet, M. Peter, J. Lygeros, and H. Koeppl, "Moment-based inference predicts bimodality in transient gene expression," *Proc Natl Acad Sci USA*, vol. 109, no. 21, pp. 8340–8345, 2012.

[6] A. Hilfinger and J. Paulsson, "Separating intrinsic from extrinsic fluctuations in dynamic biological systems.," *Proc Natl Acad Sci USA*, vol. 108, no. 29, pp. 12167–12172, July 2011.

[7] C. G. Bowsher and P. S. Swain, "Identifying sources of variation and the flow of information in biochemical networks," *Proc Natl Acad Sci USA*, vol. 109, pp. E1320–E1328, 2012.

[8] O. Aalen, "Nonparametric inference for a family of counting processes," *Ann Stat*, vol. 6, no. 4, pp. 701–726, 1978.

[9] O. Aalen, "Mixing distributions on a Markov chain," *Scand J Statist*, vol. 14, no. 4, pp. 281–289, 1987.

[10] C. Zechner, M. Unger, S. Pelet, M. Peter, and H. Koeppl, "Pooling single-cell recordings: Scalable inference through heterogeneous kinetics," *ArXiv e-prints*, Feb. 2013, arXiv:1302.3152.

[11] D. J. Wilkinson, *Stochastic Modelling for Systems Biology*, Chapman and Hall/CRC, 1 edition, Apr. 2006.

[12] U. Kuechler and M. Sorensen, *Exponential families of stochastic processes*, Springer (New York), 1997.

[13] Y. Taniguchi, P. J. Choi, G.-W. Li, H. Chen, M. Babu, J. Hearn, A. Emili, and X. S. Xie, "Quantifying e. coli proteome and transcriptome with Single-Molecule sensitivity in single cells," *Science*, vol. 329, no. 5991, pp. 533–538, July 2010.

[14] P. Bernardoff, "Which multivariate gamma distributions are infinitely divisible?," *Bernoulli*, vol. 12, no. 12, pp. 169–189, 2006.

[15] J. M. Pedraza and J. Paulsson, "Effects of molecular memory and bursting on fluctuations in gene expression," *Science*, vol. 319, no. 5861, pp. 339–343, 2008.

[16] D. F. Anderson, "A modified next reaction method for simulating chemical systems with time dependent propensities and delays.," *J Chem Phys*, vol. 127, no. 21, pp. 214107, Dec. 2007.