THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Efficient low-order approximation of first-passage time distributions

Citation for published version:
Schnoerr, D, Cseke, B, Grima, R & Sanguinetti, G 2017, 'Efficient low-order approximation of first-passage time distributions' Physical Review Letters.

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Physical Review Letters

OPEN ACCESS

Efficient low-order approximation of first-passage time distributions

David Schnoerr[1], Botond Cseke[3], Ramon Grima[2], and Guido Sanguinetti[1,*]

[1]*School of Informatics, University of Edinburgh, Edinburgh, UK*
[2]*School of Biological Sciences, University of Edinburgh, Edinburgh, UK*
[3]*Microsoft Research, Cambridge, UK*

We consider the problem of computing first-passage time distributions for reaction processes modelled by master equations. We show that this generally intractable class of problems is equivalent to a sequential Bayesian inference problem for an auxiliary observation process. The solution can be approximated efficiently by solving a closed set of coupled ordinary differential equations (for the low-order moments of the process) whose size scales with the number of species. We apply it to an epidemic model and a trimerisation process, and show good agreement with stochastic simulations.

Many systems in nature consist of stochastically interacting agents or particles. Such systems are frequently modelled as reaction processes whose dynamics are described by master equations [1]. There are several examples of stochastic modelling of reaction processes in the fields of systems biology [2, 3], ecology [4], epidemiology [5], social sciences [6] and neuroscience [7]. The mathematical analysis of such stochastic processes, however, is highly non-trivial.

A particularly important quantity of interest is the *first-passage time* (FPT), that is, the random time it takes the process to first cross a certain threshold [8, 9]. FPT distributions play a crucial role both in the theory of stochastic processes and in their applications across various disciplines as they allow us to investigate quantitatively the uncertainty in the emergence of system properties within a finite time horizon. For example, the time it takes cells to respond to external signals by expressing certain genes may be modelled as a FPT problem. Different characteristics of this first time distribution, for example the variance of the FPT, may represent evolutionarily different strategies that organisms adopt to filter fluctuations in the environment [10–12]. Examples from other disciplines include the extinction time of diseases in epidemic models, or the time it takes to form a certain number of polymers in polymerisation processes.

FPTs for stochastic processes have been of interest in statistical physics for many decades [13]. For certain random walk or spatial diffusion processes analytic solutions have been derived [14–16]. Recently, analytic results have been found for effective one-dimensional diffusion processes to a target [17–19]. For multi-dimensional diffusion processes to small targets approximate solution have been derived using singular perturbation methods and matched asymptotic expansions [20–23].

The problem of computing FPT distributions for reaction processes modelled by master equations, however, is much less explored. Generally, no tractable evolution equations exist except for one variable, one step processes [1, 13], or certain linear and/or catalytic processes [24–26]. For single-time properties of the underlying master equation efficient approximation methods exist relying on continuous state spaces [27], but it is not clear how to extend them for the computation of FPTs. Spectral methods constitute efficient approximations for small systems [28, 29]. Since these methods typically scale with the size of the state space, they are not applicable to large systems. Some existing FPT approaches for master equations consider rare events in single-species systems and/or mean FPTs only [30–33].

In this article, we approach the problem of computing FPTs from a novel perspective. We show that the FPT problem can be formulated exactly as a Bayesian inference problem. We achieve this by introducing an auxiliary observation process that determines whether the process has crossed the threshold up to a given time. This novel formulation allows us to derive an efficient approximation scheme that relies on the solution of a small set of ordinary differential equations. We will use this approximation to analyse the FPT distributions in several non-trivial examples. We focus on reaction networks with discrete state spaces modelled by master equations, but the developed method can also be applied to Fokker-Planck equations.

The standard approach to compute the FPT of a process $\boldsymbol{x}_t$ to leave a certain region C is to compute the *survival probability* $Z_{[0,t]}$, that is, the probability that the process remains in C on the time interval $[0, t]$ [13]. The FPT distribution is then given by the negative time-derivative of $Z_{[0,t]}$. The latter can be written as a path integral over the process with absorbing boundary of C [13]. Equivalently, one can reweigh the unconstrained process by an indicator function $p(C_{[0,t]}|\boldsymbol{x}_{[0,t]})$ on the paths $\boldsymbol{x}_{[0,t]}$ such that $p(C_{[0,t]}|\boldsymbol{x}_{[0,t]}) = 1$ if $\boldsymbol{x}_\tau \in C$ for $\tau \in [0, t]$ and zero otherwise. One can then write the survival probability $Z_{[0,t]}$ up till time t as a path integral over the density $p(\boldsymbol{x}_{[0,t]})$ of the unconstrained process as

$$Z_{[0,t]} = \int \mathcal{D}\boldsymbol{x}_{[0,t]}\, p(\boldsymbol{x}_{[0,t]})p(C_{[0,t]}|\boldsymbol{x}_{[0,t]}). \qquad (1)$$

At the heart of our method lies the interpretation of $p(C_{[0,t]}|\boldsymbol{x}_{[0,t]})$ as a *binary observation process*: an observer external to the system assesses if the process has left the region of interest or not. In this interpretation, the survival probability $Z_{[0,t]}$ constitutes the *marginal*

* gsanguin@inf.ed.ac.uk

likelihood of this auxiliary observation process. The problem of computing $Z_{[0,t]}$ and hence the FPT distribution is thus formally equivalent to a *Bayesian inference problem*. Note however, that there is no experimental data involved and no data is being simulated.

Moreover, note that so far no approximations have been made and (1) is exact. However, it is not obvious how to compute the path integral in (1). To make progress, we approximate the continuous process with paths $\boldsymbol{x}_{[0,t]}$ by a discrete-time version $(\boldsymbol{x}_{t_0}, \ldots, \boldsymbol{x}_{t_N})$ at points $t_0 = 0, \ldots, t_N = t$ with spacing $\Delta t = t/N$. The effects of such a discretisation of time on certain survival probabilities has recently been studied in [34]. We will later take the continuum limit $\Delta t \to 0$ and are hence not concerned with such effects.

This means that the *global observation process* $p(C_{[0,t]}|\boldsymbol{x}_{[0,t]})$ can be written as a product of *local observation processes* $p(C_{t_i}|\boldsymbol{x}_{t_i})$ as

$$p(C_{[0,t]}|\boldsymbol{x}_{t_0}, \ldots, \boldsymbol{x}_{t_N}) = \prod_{i=0}^{N} p(C_{t_i}|\boldsymbol{x}_{t_i}), \qquad (2)$$

where $p(C_{t_i}|\boldsymbol{x}_{t_i}) = 1$ if $\boldsymbol{x}_{t_i} \in C$ and zero otherwise. This gives the model a Markovian structure and allows us to cast it into a sequential Bayesian inference problem, as follows. First, we approximate the binary observation factors in (2) by a smooth approximation of the form

$$p(C_{t_i}|\boldsymbol{x}_{t_i}) \approx \exp\left(-\Delta t \, U(\boldsymbol{x}_{t_i}, t_i)\right), \qquad (3)$$

where $U(\boldsymbol{x}_{t_i}, t_i)$ is a smooth function that is large for $\boldsymbol{x}_{t_i} \notin C$ and close to zero for $\boldsymbol{x}_{t_i} \in C$, with a sharp slope at the boundary. Moreover, we require $U(\boldsymbol{x}_{t_i}, t_i)$ to have a tractable expectation w.r.t. a normal distribution. The smooth approximation in (3) proves computationally expedient in the algorithm below and will allow us to take the continuum limit $\Delta t \to 0$. Note that this approximation is equivalent to approximating the global binary constraint with the global soft (that is, continuous) constraint

$$p(C_{[0,t]}|\boldsymbol{x}_{[0,t]}) = \exp\left(-\int_0^t d\tau \, U(\boldsymbol{x}_\tau, \tau)\right). \qquad (4)$$

The survival probability $Z_{[0,t]}$ in (1) for the discrete-time system factorises as

$$Z_{[0,t]} \approx p(C_{t_0}) \prod_{i=1}^{N} p(C_{t_{i+1}}|C_{\leq t_i}), \qquad (5)$$

where $p(C_{t_0})$ is the probability of being in C at time t_0 and $p(C_{t_{i+1}}|C_{\leq t_i}) \equiv p(C_{t_{i+1}}|C_{t_i}, C_{t_{i-1}}, \ldots, C_{t_0}) = \int d\boldsymbol{x}_{t_{i+1}} p(C_{t_{i+1}}|\boldsymbol{x}_{t_{i+1}}) p(\boldsymbol{x}_{t_{i+1}}|C_{\leq t_i})$ is the probability that the process is found to be in C at time t_{i+1}, given that it was in C for all previous time points. The computation of these factors corresponds to a sequential Bayesian inference problem which can be solved by iteratively (i) solving the master equation forward between

measurement points and (ii) updating the distribution using the observation model. More specifically, the two steps comprise

(i) Suppose we know $p(\boldsymbol{x}_{t_i}|C_{\leq t_i}) \equiv p(\boldsymbol{x}_{t_i}|C_{t_i}, C_{t_{i-1}}, \ldots, C_{t_0})$ at time t_i, that is, the marginal distribution of the process at time t_i conditioned on the current and all previous observations. Suppose further that using this as the initial distribution, we can solve the system (the master equation) forward in time until time point t_{i+1} to obtain $p(\boldsymbol{x}_{t_{i+1}}|C_{\leq t_i})$, that is, the marginal distribution of the process at time t_{i+1} conditioned on previous observations (note that $p(\boldsymbol{x}_{t_{i+1}}|C_{\leq t_i})$ does not include the observation $C_{t_{i+1}}$ at time t_{i+1}).

(ii) To obtain $p(\boldsymbol{x}_{t_{i+1}}|C_{\leq t_{i+1}})$ we need to take the observation $p(C_{t_{i+1}}|\boldsymbol{x}_{t_{i+1}})$ at time point t_{i+1} into account. This is achieved by means of Bayes' theorem as

$$p(\boldsymbol{x}_{t_{i+1}}|C_{\leq t_{i+1}}) = \frac{p(C_{t_{i+1}}|\boldsymbol{x}_{t_{i+1}})p(\boldsymbol{x}_{t_{i+1}}|C_{\leq t_i})}{Z_{t_{i+1}}}, \qquad (6)$$

where we defined the normalisation $Z_{t_{i+1}} = p(C_{t_{i+1}}|C_{\leq t_i})$. Note that the latter is just a factor in (5).

Performing steps (i) and (ii) iteratively from t_0 to t_N and keeping track of the normalisation factors in (6) one can thus, in principle, compute the survival probability according to (5).

However, steps (i) and (ii) are generally intractable, and we propose an approximation method in the following. For step (i), we need to solve the system forward in time. We do this approximately by means of the *normal moment closure* [35–37], which approximates the discrete process by a continuous one and assumes the single-time probability distribution to be a multivariate normal distribution $\mathcal{N}(\boldsymbol{x}_t; \boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t)$ with mean $\boldsymbol{\mu}_t$ and covariance $\boldsymbol{\Sigma}_t$. Using this assumption on the master equation leads to a closed set of ordinary differential equations for $\boldsymbol{\mu}_t$ and $\boldsymbol{\Sigma}_t$ which can be solved numerically [27].

Now suppose that we have solved the system forward from time t to $t + \Delta t$ using normal moment closure to obtain $\hat{\boldsymbol{\mu}}_{t+\Delta t}$ and $\hat{\boldsymbol{\Sigma}}_{t+\Delta t}$ and hence the distribution $p(\boldsymbol{x}_{t+\Delta t}|C_{\leq t}) = \mathcal{N}(\boldsymbol{x}_{t+\Delta t}; \hat{\boldsymbol{\mu}}_{t+\Delta t}, \hat{\boldsymbol{\Sigma}}_{t+\Delta t})$ (step (i)). We next have to perform the observation update in (6) in step (ii) to obtain $p(\boldsymbol{x}_{t+\Delta t}|C_{\leq t+\Delta t})$. In order to be able to use normal moment closure again in the next (i) step, we approximate $p(\boldsymbol{x}_{t+\Delta t}|C_{\leq t+\Delta t})$ by a multivariate normal distribution with mean $\boldsymbol{\mu}_{t+\Delta t}$ and covariance $\boldsymbol{\Sigma}_{t+\Delta t}$ of the r.h.s. in (6). This approach is known as *assumed-density filtering* in the statistics literature [38]. In summary, with the described approximations, steps (i) and (ii) comprise

(i) Solve normal moment closure equations for $\boldsymbol{\mu}_t$ and $\boldsymbol{\Sigma}_t$ from time t to $t+\Delta t$ to obtain $\hat{\boldsymbol{\mu}}_{t+\Delta t}$ and $\hat{\boldsymbol{\Sigma}}_{t+\Delta t}$,

where $\boldsymbol{\mu}_t$ and $\boldsymbol{\Sigma}_t$ are respectively the mean and co-variance of the approximating normal distribution $\mathcal{N}(\boldsymbol{x}_t; \boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t)$.

(ii) Compute the mean $\boldsymbol{\mu}_{t+\Delta t}$ and covariance $\boldsymbol{\Sigma}_{t+\Delta t}$ of the r.h.s. of (6) and approximate $p(\boldsymbol{x}_{t+\Delta t}|C_{\leq t+\Delta t})$ in (6) with a corresponding normal distribution $\mathcal{N}(\boldsymbol{x}_{t+\Delta t}; \boldsymbol{\mu}_{t+\Delta t}, \boldsymbol{\Sigma}_{t+\Delta t})$.

Next, we derive a continuous time description combining step (i) and (ii). This is achieved by first expanding the update in step (i) leading from $\boldsymbol{\mu}_t$ and $\boldsymbol{\Sigma}_t$ to $\hat{\boldsymbol{\mu}}_{t+\Delta t}$ and $\hat{\boldsymbol{\Sigma}}_{t+\Delta t}$ in Δt which gives a single Euler step update of the moment closure equations. Similarly, we expand step (ii) which leads from $\hat{\boldsymbol{\mu}}_{t+\Delta t}$ and $\hat{\boldsymbol{\Sigma}}_{t+\Delta t}$ to $\boldsymbol{\mu}_{t+\Delta t}$ and $\boldsymbol{\Sigma}_{t+\Delta t}$, as follows. Note first that by definition, the normalisation $Z_{t+\Delta t}$ in (6) can be written as

$$Z_{t+\Delta t} \approx \int d\boldsymbol{x}\, \mathcal{N}(\boldsymbol{x}; \hat{\boldsymbol{\mu}}_{t+\Delta t}, \hat{\boldsymbol{\Sigma}}_{t+\Delta t}) \exp\left(-\Delta t\, U(\boldsymbol{x}, t + \Delta t)\right). \tag{7}$$

Taking the logarithm of both sides, expanding in Δt and taking derivatives w.r.t. to $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$ one can derive the desired expansion of the update in step (ii). The resulting expansions of steps (i) and (ii) can then be combined to give unifying update equations for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ (see Supplemental Material [39] for a derivation). Finally, taking the continuum limit $\Delta t \to 0$ gives the following closed set of differential equations

$$\frac{\partial}{\partial t}\boldsymbol{\mu}_t = \left(\frac{\partial}{\partial t}\boldsymbol{\mu}_t\right)^{\mathrm{MC}} - \boldsymbol{\Sigma}_t \frac{\partial}{\partial \boldsymbol{\mu}_t}\langle U(\boldsymbol{x}_t, t)\rangle_{\mathcal{N}(\boldsymbol{x}_t; \boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t)}, \tag{8}$$

$$\frac{\partial}{\partial t}\boldsymbol{\Sigma}_t = \left(\frac{\partial}{\partial t}\boldsymbol{\Sigma}_t\right)^{\mathrm{MC}} - 2\boldsymbol{\Sigma}_t \left(\frac{\partial}{\partial \boldsymbol{\Sigma}_t}\langle U(\boldsymbol{x}_t, t)\rangle_{\mathcal{N}(\boldsymbol{x}_t; \boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t)}\right)\boldsymbol{\Sigma}_t, \tag{9}$$

$$\frac{\partial}{\partial t}\log Z_{[0,t]} = -\langle U(\boldsymbol{x}_t, t)\rangle_{\mathcal{N}(\boldsymbol{x}_t; \boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t)}. \tag{10}$$

Here, the first terms on the r.h.s. of (8) and (9) are respectively the equations for the mean and covariance as obtained from the normal moment closure (MC) for the unconstrained system, while the second terms incorporate the auxiliary observation. Equation (10) gives the desired survival probability for the process. We term this method for computing FPT distributions *Bayesian First-Passage Times* (BFPT).

Equations (8)-(10) are the central result of this article. They constitute closed form ordinary differential equations for the mean, covariance and log-survival probability of the process, for which efficient numerical integrators exist. Solving these equations forward in time on an interval $[0, t]$ provides an approximation of the survival probability $Z_{[0,\tau]}$ for all $\tau \in [0, t]$ (on the time grid of the numerical integrator), from which the FPT distribution $p(\tau; C)$ can be derived for all $\tau \in [0, t]$ by taking the negative derivative of $Z_{[0,\tau]}$, that is, $p(\tau; C) = -\partial/\partial\tau Z_{[0,\tau]}$. The number of equations scales with the square of the number of species and the method hence is applicable to large systems. Crucially, and in contrast to stochastic simulations and spectral methods, the complexity of
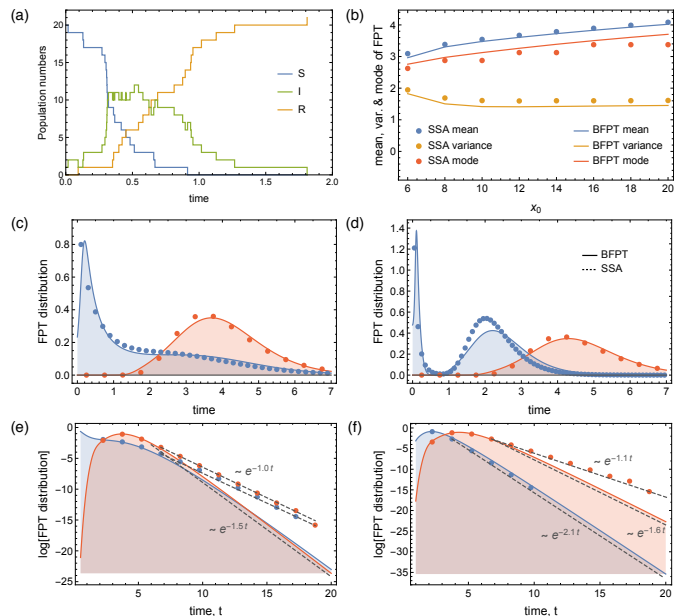

FIG. 1. Results for the epidemic system (12). (a) simulated path of the process. (b) mean, variance and mode of the FPT distribution of species I becoming extinct as a function of the initial populations x_0 of species S, from the stochastic simulation algorithm (SSA, dots, 10^6 samples per point) and from BFPT (lines). The rate constants are set to $k_1 = 0.5$ and $k_2 = 1$ and the initial value for species I is set to $y_0 = 2x_0$. (c) and (d) FPT distributions as obtained from the SSA (dots, 10^7 samples per parameter set) and BFPT (lines) for the parameter set (x_0, y_0, k_1, k_2) chosen as $(6, 1, 0.25, 1)$ (blue, (c)), $(20, 10, 0.5, 1)$ (red, (c)), $(20, 1, 0.5, 2)$ (blue, (d)), $(40, 10, 0.25, 1)$ (red, (d)). The parameter a in (11) was chosen as $a = -3$ for the blue curve in (c) and $a = -1.5$ for all other figures. (e) and (f) same results as (c) and (d) but logarithmic scale.

the method is independent of the population size and the size of the state space. Similar equations were obtained in a different context in [40, 41] by means of a variational approximation.

In the derivation of (8)-(10) we utilised three approximations: after discretising time, we approximated the unconstrained process using normal moment closure and the observation updates by projections onto a normal distribution. We then approximated the binary observation model by a soft loss function which allowed us to take the continuum limit in time. Depending on the problem, the relative contribution of the three sources to the overall error may vary.

The choice of loss function $U(\boldsymbol{x}, t)$ depends on the problem at hand. In general, for computational convenience one needs to be able to compute analytically the expectation of the loss function w.r.t. a multivariate normal distribution. In our examples, we use an exponential loss function to constrain the ith component of the state vector $\boldsymbol{x}$ about a threshold c

$$U(\boldsymbol{x}, t) = \exp(a(x^i - c)), \quad a \in \mathbb{R}, c \in \mathbb{R}. \tag{11}$$

The absolute value of the parameter a determines the steepness of the loss function. In principle, we choose a as large as numerically feasible. For a detailed discussion on the choice of loss function see Supplemental Material [39].

We now examine the performance of BFPT on three examples. For the analytically tractable Poisson birth process we find that BFPT captures the low order moments and the bulk mass of the distribution accurately whilst giving the correct scaling law for the tail of the distribution (see Supplemental Information for details).

Next, we consider an epidemic system consisting of a susceptible population S, an infected population I and a recovered population R and interactions

$$S + I \xrightarrow{k_1} 2I, \quad I \xrightarrow{k_2} R. \tag{12}$$
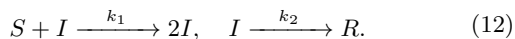
This system is frequently modelled as a continuous-time Markov-jump process to model a disease spreading through a population. k_1 and k_2 in (12) are the corresponding rate constants. Let $\boldsymbol{x}_t = (z_t, y_t, z_t)$, where x_t, y_t and z_t denote the populations of S, I and R, respectively. We are interested in the probability distribution of time for the disease to be permanently eradicated, that is, the time it takes the process to reach a state with $y_t = 0$.

Fig. 1 (b) shows the mean, variance and mode of the FPT to extinction as obtained from our method and the stochastic simulation algorithm [42]. We find that BFPT accurately captures the mean, variance and mode of the FPT over a wide range of varying initial values for S and I.

Fig. 1 (c) and (d) show the FPT distributions for four different parameter sets. The modality, mode and overall shape of the FPT are well captured, even for highly skewed and bimodal distributions (c.f. blue curve in Fig. 1 (c) and (d), respectively). In some cases the method predicts less peaked distributions than actual (not shown here). Fig. 1 (e) and (f) show the same results on logarithmic scale. We observe that our method correctly predicts an exponential scaling (straight lines in logarithmic scale), although the scaling is not always accurate, indicating a worse approximation in the tails of the distribution.

The value of the approach is borne out by considering its computational efficiency: for the results shown in Fig. 1, BFPT is several orders of magnitude faster than stochastic simulations. For example, simulating 10^7 paths to obtain the results shown in Fig. 1 (c)-(f) takes about $10^3 - 10^4$ seconds in our implementation of the direct stochastic simulation algorithm [42], whereas BFPT takes less than a second.

Finally, we apply BFPT to a polymerisation system of monomers X, dimers XX and trimers XXX with interactions

$$X + X \xrightarrow{k_1} XX, \quad XX + X \xrightarrow{k_2} XXX. \tag{13}$$
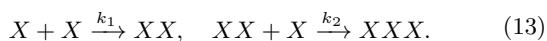
Starting from a fixed number of 10^3 of monomers, zero dimers and zero trimers, we are interested in the FPT it
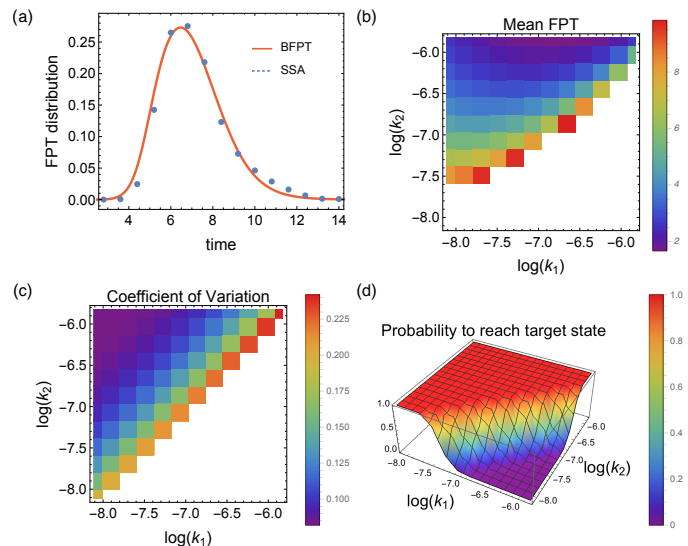

FIG. 2. Results for the polymerisation system in (13). (a) FPT distribution for the parameters $k_1 = k_2 = 10^{-3}$ obtained from the stochastic simulation algorithm (SSA, dots, 10^4 samples). (b) and (c) heat plots of the mean and the coefficient of variation (defined as standard deviation divided by mean) of the FPT to produce 200 trimers starting with 10^3 monomers, as a function of k_1 and k_2 on logarithmic scale. (d) corresponding 3D-plot for the normalisation of the FPT distribution, that is, the probability with which at least 200 trimers are being produced. The white areas in (b) and (c) indicate that either the value is larger than the plotted range or that the target state is reached with such small probability that an estimation of moments is not sensible. The parameter a in (11) was fixed to $a = 0.2$ for all figures.

takes to produce 200 trimers. We are interested in exploring the dependence of this FPT distribution on the parameters of the system (dimerisation and trimerisation rate k_1 and k_2 respectively); such parameter exploration is computationally too demanding to be performed by brute force simulation without access to dedicated hardware since the FPT distribution needs to be estimated for a large number of parameter sets.

Fig. 2 shows the results for this process. We observe excellent agreement between BFPT and simulations for a particular value of the parameters (Fig. 2 (a)). The heat plot for the mean as a function of k_1 and k_2 indicates that for a given trimerisation rate k_2 a minimal mean FPT is achieved for an intermediate value of dimerisation rate k_1 (Fig. 2 (b)). We find a linear relationship $k_2 \approx 2.3k_1$ between the two rates for the location of these minima. The variance of FPT behaves quantitatively similarly (not shown in the figure). The coefficient of variation (Fig. 2 (c)), however, becomes minimal for small values of k_1 for a given k_2. This reveals an unexpected trade-off between an optimal mean FPT and optimal noise-to-mean ratio (coefficient of variation). Fig. 2 (d) shows the probability that the target state is reached, that is, the probability that at least 200 trimers are being produced. We find that there are two parameter regions, one with

probability close to one and one with probability close to zero, and a small transition range between these two with boundary $k_2 \approx 0.55k_1$.

In conclusion, we have shown that the problem of computing survival probabilities and FPT distributions for stochastic processes can be formulated as a sequential Bayesian inference problem. This novel formulation opens the way for a new class of efficient approximation methods from machine learning and computational statistics to address this classical intractable problem. Here, we derived an approximation for FPT distributions which relies on solving a small set of ordinary differential equations. This results in considerable efficiency gains; empirically, we found the approximation to be highly ac-

curate in several examples. However, we do not have at present systematic error estimates for the method; we leave the investigation of such bounds and possible correction methods for future work. In particular, it will be interesting to study the tail behaviour of FPT distributions with our method, as these were not always captured well in our examples. We notice that, while we applied our method to processes with discrete state spaces modelled by master equations, in principle it can equally easily be applied to Fokker-Planck processes.

[1] C. W. Gardiner, *Handbook of Stochastic Methods*, Vol. 3 (Springer Berlin, 1985).
[2] A. Eldar and M. B. Elowitz, Nature **467**, 167 (2010).
[3] R. Grima, Phys. Rev. Lett. **102**, 218103 (2009).
[4] A. J. McKane and T. J. Newman, Phys. Rev. Lett. **94**, 218102 (2005).
[5] M. I. Dykman, I. B. Schwartz, and A. S. Landsman, Phys. Rev. Lett. **101**, 078101 (2008).
[6] J. Fernández-Gracia, K. Suchecki, J. J. Ramasco, M. San Miguel, and V. M. Eguíluz, Phys. Rev. Lett. **112**, 158701 (2014).
[7] T. Betz, D. Lim, and J. A. Käs, Phys. Rev. Lett. **96**, 098103 (2006).
[8] M. R. D'Orsogna and T. Chou, Phys. Rev. Lett. **95**, 170603 (2005).
[9] S. Condamin, O. Bénichou, V. Tejedor, R. Voituriez, and J. Klafter, Nature **450**, 77 (2007).
[10] E. Kussell and S. Leibler, Science **309**, 2075 (2005).
[11] S. Tănase-Nicola, P. B. Warren, and P. R. ten Wolde, Phys. Rev. Lett. **97**, 068102 (2006).
[12] T. J. Kobayashi, Phys. Rev. Lett. **104**, 228104 (2010).
[13] S. Redner, *A guide to first-passage processes* (Cambridge University Press, 2001).
[14] O. Bénichou, C. Loverdo, M. Moreau, and R. Voituriez, Reviews of Modern Physics **83**, 81 (2011).
[15] A. J. Bray, S. N. Majumdar, and G. Schehr, Advances in Physics **62**, 225 (2013).
[16] F. Aurzada and T. Simon, in *Lévy Matters V* (Springer, 2015) pp. 183–224.
[17] A. Godec and R. Metzler, Phys. Rev. X **6**, 041037 (2016).
[18] A. Godec and R. Metzler, Scientific reports **6**, 20349 (2016).
[19] A. Godec and R. Metzler, Journal of Physics A: Mathematical and Theoretical **50**, 084001 (2017).
[20] A. Singer, Z. Schuss, D. Holcman, and R. Eisenberg, Journal of Statistical Physics **122**, 437 (2006).
[21] D. Holcman and Z. Schuss, SIAM Review **56**, 213 (2014).
[22] S. A. Isaacson, A. J. Mauro, and J. Newby, Physical Review E **94**, 042414 (2016).
[23] P. C. Bressloff and S. D. Lawley, Journal of Physics A: Mathematical and Theoretical **50**, 195001 (2017).
[24] G. Bel, B. Munsky, and I. Nemenman, Physical biology **7**, 016003 (2009).
[25] B. Munsky, I. Nemenman, and G. Bel, The Journal of chemical physics **131**, 12B616 (2009).
[26] R. Grima and A. Leier, J. Phys. Chem. B **121**, 13 (2016).
[27] D. Schnoerr, G. Sanguinetti, and R. Grima, J. Phys. A **50**, 093001 (2017).
[28] P. Deuflhard, W. Huisinga, T. Jahnke, and M. Wulkow, SIAM Journal on Scientific Computing **30**, 2990 (2008).
[29] V. Kazeev, M. Khammash, M. Nip, and C. Schwab, PLoS computational biology **10**, e1003359 (2014).
[30] M. Assaf, A. Kamenev, and B. Meerson, Physical Review E **78**, 041123 (2008).
[31] M. Assaf and B. Meerson, Physical Review E **81**, 021116 (2010).
[32] S. Be'er and M. Assaf, Journal of Statistical Mechanics: Theory and Experiment **2016**, 113501 (2016).
[33] M. F. Weber and E. Frey, Rep. Prog. Phys. **80**, 046601 (2017).
[34] S. N. Majumdar, A. J. Bray, and G. C. Ehrhardt, Physical Review E **64**, 015101 (2001).
[35] L. A. Goodman, Biometrics **9**, 212 (1953).
[36] D. Schnoerr, G. Sanguinetti, and R. Grima, The Journal of chemical physics **141**, 08B616_1 (2014).
[37] D. Schnoerr, G. Sanguinetti, and R. Grima, The Journal of Chemical Physics **143**, 11B610_1 (2015).
[38] P. S. Maybeck, *Stochastic models, estimation, and control*, Vol. 3 (Academic press, 1982).
[39] See Supplemental Material at http:// for a derivation of (8)-(10), a detailed discussion of the choice of the loss function in (11), as well as a derivation of the exact results for the Poisson birth process , .
[40] B. Cseke, M. Opper, and G. Sanguinetti, in *Advances in Neural Information Processing Systems* (2013) pp. 971–979.
[41] B. Cseke, D. Schnoerr, M. Opper, and G. Sanguinetti, J. Phys. A. **49** (2016).
[42] D. T. Gillespie, J. Comput. Phys. **22**, 403 (1976).