Approximating first-passage time distributions via sequential Bayesian computation

David Schnoerdl, Botond Cseke3, Ramon Grima2, and Guido Sanguinetti1,3

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

We show that computing first-passage time distributions for stochastic processes 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. The method is accurate and orders of magnitude faster than existing approaches, enabling hitherto computationally prohibitive tasks such as sensitivity analysis. We apply it to an epidemic model, an entrained oscillator and a trimerisation process, and show good agreement with exact stochastic simulations.

Many systems in nature consist of stochastically interacting agents or particles. Such systems are typically modelled as continuous-time Markov processes. There are several examples of stochastic modeling in the fields of systems biology [1,2], ecology [3], epidemiology [4], social sciences [5] and neuroscience [6]. For all but the simplest cases no analytic solutions to the corresponding master or Fokker-Planck equations are known and exact stochastic simulations of such processes are computationally expensive. Significant effort has therefore been spent on the development of approximation methods for efficient computation of systems dynamics [7].

Most methods compute the single-time marginal probability distributions of the process by developing an approximation to the underlying stochastic description. Path properties, in contrast, are not captured by these equations. An important path quantity is the first-passage time (FPT), that is, the 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 permit 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 its variance, may represent evolutionarily different strategies that organisms adopt to filter fluctuations in the environment [10–12]. Other intriguing phenomena, such as the variation in period of a stochastic oscillator, can be cast as a FPT problem. Fluctuations of such oscillators have recently been found to lead to widely varying differentiation times of genetically identical cells [13]. Examples from other disciplines include the extinction time of dis-
\( (x_{t_0}, \ldots, x_{t_N}) \) at points \( t_0 = 0, \ldots, t_N = t \) with spacing \( \Delta t = t/N \). This means that the global observation process \( p(C_{[0,t]}|x_{[0,t]}) \) can be written as a product of local observation process as

\[
p(C_{[0,t]}|x_{[t_0]}, \ldots, x_{t_N}) = \prod_{i=0}^{N} p(C_{t_i}|x_{t_i}),
\]

where \( p(C_{t_i}|x_{t_i}) = 1 \) if \( x_{t_i} \in C \) and zero otherwise. This gives the problem 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

\[
p(C_{t_i}|x_{t_i}) \approx \exp \left( -\Delta t \, U(x_{t_i}, t_i) \right),
\]

where \( U(x_{t_i}, t_i) \) is a smooth function that is large for \( x_{t_i} \notin C \) and close to zero for \( x_{t_i} \in C \), with a sharp transition at the boundary. We will give specific examples later on. For now, the only property \( U(x_{t_i}, t_i) \) has to satisfy is that we can compute its expectation analytically with respect to a normal distribution and that it allows a closed-form expression in the moments of this distribution. The smooth approximation in (3) proves computationally expedient in the algorithm below and allows us to take the continuum limit \( \Delta t \to 0 \) since it is the discretised version of the global soft constraint

\[
p(C_{[0,t]}|x_{[0,t]}) = \exp \left( -\int_0^t d\tau \, U(x_\tau, \tau) \right).
\]

Due to the Markov property, the survival probability \( Z_{[0,t]} \) in (1) for the discrete-time system factorises as

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

where \( C_{\leq t_i} = \{C_i, C_{i-1}, \ldots, C_0\} \). The computation of the factors on the r.h.s. of (5) corresponds to a sequential Bayesian inference problem which can be solved iteratively as follows:

(i) Suppose we know \( p(x_{t_i}|C_{\leq t_i}) \), and that using this as the initial distribution, are able to solve the system (the master or Fokker-Planck equation) forward in time until time point \( t_{i+1} \) to obtain \( p(x_{t_{i+1}}|C_{\leq t_i}) \).

(ii) To obtain \( p(x_{t_{i+1}}|C_{\leq t_{i+1}}) \), take the constraint in (3) for time point \( t_{i+1} \) into account by means of the Bayes’ theorem as

\[
p(x_{t_{i+1}}|C_{\leq t_{i+1}}) = p(C_{t_{i+1}}|x_{t_{i+1}})p(x_{t_{i+1}}|C_{\leq t_i})/Z_{t_{i+1}}.
\]

where we defined \( Z_{t_{i+1}} = p(C_{t_{i+1}}|C_{\leq t_i}) \). Note that the latter corresponds to the ith term in (5).

Performing steps (i) and (ii) iteratively from \( t_0 \) to \( t_N \) 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 \([19]\), which assumes the single-time probability distribution to be a multi-variate normal distribution \( N(x_t; \mu_t, \Sigma_t) \), where \( \mu_t \) and \( \Sigma_t \) are the mean and covariance of the process. This assumption leads to closed evolution equations for \( \mu_t \) and \( \Sigma_t \) which can be solved numerically \([17]\).

Now suppose that we have solved the system forward from time \( t \) to \( t + \Delta t \) to obtain \( \mu_t + \Delta \mu_t \) and \( \Sigma_t + \Delta \Sigma_t \) (step (i)). In order to perform the observation update in (1) we approximate the posterior distribution \( p(x_{t+\Delta t}|C_{\leq t+\Delta t}) \) by a normal distribution with mean \( \mu_t + \Delta \mu_t \) and covariance \( \Sigma_t + \Delta \Sigma_t \), an approach known as assumed-density filtering in the statistics literature \([20]\). The corresponding normalisation \( Z_{t+\Delta t} \) in (6) can be written as

\[
Z_{t+\Delta t} \approx \int dxN(x; \mu_{t+\Delta t}, \Sigma_{t+\Delta t}) \exp \left( -\Delta t \, U(x, t + \Delta t) \right).
\]

Taking the logarithm of both sides, expanding in \( \Delta t \) both the exponential within the integral and the logarithm, and taking derivatives w.r.t. \( \mu_t \) and \( \Sigma_t \), one obtains corresponding update equations (see Appendix A for a derivation).

\[
\frac{\partial}{\partial t} \mu_t = \left( \frac{\partial}{\partial t} U(x, t) \right)_{C(x, \mu_t, \Sigma_t)} + \Sigma_t \frac{\partial}{\partial \mu_t} \left( U(x, t) \right)_{C(x, \mu_t, \Sigma_t)} \Sigma_t,
\]

\[
\frac{\partial}{\partial t} \Sigma_t = 2\Sigma_t \frac{\partial}{\partial \mu_t} \left( U(x, t) \right)_{C(x, \mu_t, \Sigma_t)} \Sigma_t \Sigma_t,
\]

\[
\frac{\partial}{\partial t} \log Z_{[0,t]} = -(U(x_{t_i}, t))_{C(x_{t_i}, \mu_{t_i}, \Sigma_{t_i})}.
\]

Here, the first terms on the r.h.s. of (8) and (9) are respectively the prior equations for the mean and covariance as obtained from the normal moment closure for the unconstrained process 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 \( Z_{[0,t]} \) for \( \tau \in [0, t] \) (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 \( Z_{[0,t]} \). We note that similar equations were obtained in a different context in \([22]\) by means of a variational approximation without the need for time-discretisation.

Note that in the derivation of (8) - (10) we utilised three approximations: we approximated the unconstrained process using normal moment closure, the binary
observation model by a soft potential and the observation updates by projections onto a normal distribution. Depending on the circumstances, the relative contribution of the three sources to the overall error may vary.

We now examine the performance of BFPT on three numerical examples. First, we consider an epidemic system consisting of a susceptible population $S$, an infected population $I$ and a recovered population $R$ and interactions

$$S + I \xrightarrow{k_1} 2I, \quad I \xrightarrow{k_2} R. \tag{11}$$

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 (11) are the corresponding rate constants. Let $x_t = (x_t, y_t, z_t)$, 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 $y_t = 0$. To approximate this extinction region by a soft constraint as in (12) we use a potential

$$U(t, x_t) = (y_t - b)^n / a^n, \quad a \in \mathbb{R}_+, b \in \mathbb{R}, n \in \mathbb{Z} \text{ and even.} \tag{12}$$

We found a polynomial potential as in (12) numerically more stable than for example an exponential potential of the form $\exp(-n(x - b))$ and will use this form also for following examples. For more details on the choice of $U(t, x_t)$ see the Appendix B.

Figure 1 (b) shows the mean, variance and mode of the FPT to extinction as obtained from our method and the stochastic simulation algorithm (SSA). We find that BFPT captures the mean value of the FPT over a wide range of varying initial values for $S$ and $I$. The variance and mode are also captured well with errors decreasing for larger initial values.

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 $10^4$ paths to obtain the results shown in Fig. 1 (d), takes about $10^2$ seconds in our implementation of the direct stochastic simulation algorithm [22], whereas BFPT takes less than a second. Crucially, while stochastic simulations scale with the number of transitions per unit time and thus with the average population size, BFPT scales with the number of species. Therefore, for larger population sizes stochastic simulations will become computationally more and more expensive while BFPT’s cost can be expected to remain of the same order.

Next, we study a stochastic oscillator that becomes entrained by an external input. Such models are frequently used to model circadian oscillators in plants [23]. We are interested in the fluctuations of the oscillatory period of the stochastic process. To this end, we use a prey-predator system of the Lotka-Volterra type consisting of a prey species $X$ and predator species $Y$ that interact via

$$\emptyset \xrightarrow{k_0} X, \quad \emptyset \xrightarrow{k_1} Y, \quad X \xrightarrow{(a \sin(2\pi \tau / p) + 1)k_2} X + X, \quad Y + X \xrightarrow{k_3} Y + Y, \quad Y \xrightarrow{k_4} \emptyset. \tag{13}$$

Note that the rate of the third reaction, which resembles
reproduction of prey $X$, oscillates around a mean value $k_2$ with amplitude $a$ and period $p$. This could for example model the changing availability of food resources due to seasons over the period of a year. Figure 2 shows the distribution of one oscillatory period. The oscillatory period is computed by estimating the FPT for the process starting from its maximum minus half a standard deviation to first reach this value again after one period (see Appendix C for details). Again BFPT is in excellent agreement with simulation results, but in this case the difference in computational costs is even more marked: due to the time-dependence of the third reaction in (13), this system cannot be simulated by standard simulation algorithms [22], requiring a computationally intensive modified simulator [24]. In this example, simulating 1000 periods to produce the results shown in Fig. 2 took about $10^3$ hours implementation. The BFPT result in Figure 2, in contrast, took only 0.1 seconds meaning that BFPT is about 7 orders of magnitude faster than the exact simulation.

Finally, we apply BFPT to a polymerisation system of monomers $X$, dimers $XX$ and trimers $XXX$ with inter-actions

$$X + X \xrightarrow{k_1} XX, \quad XX + X \xrightarrow{k_2} XXX.$$  \hspace{1cm} (14)

Starting from a fixed number of $10^3$ of monomers, zero dimers and zero trimers, we are interested in the FPT it 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.

Figure 3 shows the results for this process. We observe excellent agreement between BFPT and simulations for a particular value of the parameters (Fig. 3(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 $k_1$. 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. 3(c)), however, becomes minimal for small values of $k_1$ for a given $k_2$. This reveals an unexpected tradeoff between an optimal mean FPT and optimal noise-to-mean ratio (coefficient of variation). Figure 3(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 one approximation which relies on solving a small set of ordinary differential equations and which was found to combine distributional accuracy with high computational efficiency for several examples. This efficiency can be leveraged to perform hitherto computationally prohibitive systematic analyses such as parameter sweeps and opens the way to quantitatively understand how physical parameters interact to determine emergent behaviours in stochastic systems. The approach can in principle also be used to compute more general path properties, such as general reachability constraints in computer science; it would be interesting to investigate extensions of this approach to address intractable problems for high dimensional and spatio-temporal processes.

This work was supported by the Leverhulme Trust [RPG-2013-171]; and the European Research Council [MLCS 306999].
Appendix A: Derivation of main result

Here, we derive the main results of this work given in Equations (8)-(10). To this end, consider step (i) of the iterative scheme presented in the main text and suppose we have solved the corresponding moment closure equations from $t$ to $t + \Delta t$ to obtain $\mu_{t+\Delta t}$ and $\Sigma_{t+\Delta t}$ which we can thus expand as

$$\hat{\mu}_{t+\Delta t} = \mu_t + (\partial_{\Delta t} \mu_t)^{MC} \Delta t + O(\Delta t^2),$$

$$\hat{\Sigma}_{t+\Delta t} = \Sigma_t + (\partial_{\Delta t} \Sigma_t)^{MC} \Delta t + O(\Delta t^2).$$

These are the moments of the distribution at time $t + \Delta t$ prior to the update in (6):

$$p(x_{t+\Delta t} \leq x) = N(x; \hat{\mu}_{t+\Delta t}, \hat{\Sigma}_{t+\Delta t}).$$

Using the update equation (6) and the exponential form of the constraint given in (3) one finds that the normalisation can be written as stated in (7). Using this one can derive the relations

$$\frac{\partial}{\partial \mu_t} \log Z_{t+\Delta t} = \left( -\frac{\partial}{\partial \mu_t} \log N(x; \hat{\mu}_{t+\Delta t}, \hat{\Sigma}_{t+\Delta t}) \right) N(x; \mu_t, \Sigma_t).$$

Taking the logarithm of (7) and expanding in $\Delta t$ we further find

$$\log Z_{t+\Delta t} = -\Delta t(U(x, t + \Delta t))_{N(x, \mu_{t+\Delta t}, \Sigma_{t+\Delta t})} + O(\Delta t^2).$$

Using (A4) and (A5) this leads to the update equations

$$\mu_{t+\Delta t} = \mu_{t+\Delta t}$$

$$\Delta t \Sigma_{t+\Delta t} \frac{\partial}{\partial \mu_t} (U(x, t + \Delta t))_{N(x, \mu_{t+\Delta t}, \Sigma_{t+\Delta t})} + O(\Delta t^2),$$

$$\Sigma_{t+\Delta t} = \Sigma_{t+\Delta t}$$

$$2\Delta t \Sigma_{t+\Delta t} \left( -\frac{\partial}{\partial \Sigma_{t+\Delta t}} (U(x, t + \Delta t))_{N(x, \mu_{t+\Delta t}, \Sigma_{t+\Delta t})} \right) \Sigma_{t+\Delta t}.$$
of trimers to be smaller than 200. Choosing $a = 200$ and $b = 0$, we confine the trimer numbers to $[-200, 200]$. However, since the probability mass for negative values is vanishingly small (zero for the true, discrete process), this is basically the same as confining it to $(−∞, 200]$.

Similar arguments apply to the other studied examples.

This leaves us with the choice of the power $n$. A larger $n$ corresponds to a steeper potential and hence to a better approximation of the binary observation process. However, a larger $n$ also leads to stronger non-linearity of the equations (8)-(10). This will typically limit $n$ for numerical feasibility. Optimally, one would hope the results to converge in the power $n$ for values for which the equations are still numerically feasible. For some of the examples studied we indeed found this to be the case. For others, we chose $n$ as large as numerically feasible. A more systematic way of choosing $n$ is left for future work.

**Appendix C: Period of entrained oscillator**

Figure 2 in the main text shows the distribution of one oscillatory period of the entrained oscillator defined in [13]. Note that there are various potential ways of how to measure the oscillatory period. Here, we use a definition that allows a consistent comparison of stochastic simulations and our BFPT method. Specifically, we first estimate the maximum of the mean of species $X$ and estimate the standard deviation in the maximum. We then take this maximal value minus half the standard deviation as a deterministic initial condition for the process (with external input shifted appropriately) and measure the time it takes the process to reach this initial value again after having dropped to less than the overall average mean value.

Note that not all trajectories of the process will reach a fixed target value at each period. This means that the FPT distribution does not integrate to one over one period. To get a statistically stronger estimate of the fluctuations in the period, we take the maximal mean value minus a half a standard deviation rather than just the maximal mean value, since this significantly increases the mass of the FPT distribution within one oscillatory period.