Comparing diffusion and weak noise approximations for inference in reaction models

Andreas Ruttor, Florian Stimberg, and Manfred Opper

Computer Science, TU Berlin, Germany

1 Inference for reaction systems

The problem of probabilistic inference for stochastic reaction models in systems biology has attracted considerable interest, see e.g. [1]. A well studied problem is the limiting case, where the number of molecules in the system is sufficiently large to allow for a deterministic description of the dynamics by a set of (usually nonlinear) ordinary differential equations.

Inference becomes far more complicated when fluctuations are relevant. The dynamics is modelled by a continuous time Markov jump process, which describes stochastic changes of the number of molecules of a given type due to the reactions in the system. A simplified modelling of the stochastic dynamics is possible, when molecule numbers are large enough to be approximated by continuous random variables. A common computational technique for this limit is the replacement of the discrete jump process by a Markov process with continuous sample paths, i.e. by a diffusion process. Despite this simplification, statistical inference using Markov chain Monte Carlo (MCMC) methods is still computationally demanding [2].

A different approach to probabilistic inference designed for the same range of problems, where molecule numbers are not too small, has been termed weak noise approximation in [3] and was motivated by van Kampen's system size expansion [4]. It is based on the idea that relative fluctuations of molecule numbers may not be large in such cases and could—to lowest order—be well approximated by Gaussian random variables. This method leads to the solution of ordinary differential equations for the moments of the Gaussians, which can be solved in times that are usually much smaller than the ones required for Markov chain Monte Carlo approaches.

Hence, one might ask the question, whether the use of the simpler approach may lead to dramatically different results. To address this question we compare the results of the two methods on the well-known Lotka-Volterra model (for a definition of the reactions and rate constants see [5]). Other models of reaction systems are currently under investigation.

2 Diffusion approximation and MCMC

The state of reaction models is described by a vector $\mathbf{x} = (x_1, \ldots, x_M)^\top$, where x_i is the number of molecules of species i. The stochastic dynamics is a Markov jump

process (MJP) defined by a rate function $f(\mathbf{x}'|\mathbf{x})$ which determines the temporal change of transition probabilities via $P(\mathbf{x}', t+\Delta t|\mathbf{x}, t) \simeq \delta_{\mathbf{x}',\mathbf{x}} + \Delta t\, f(\mathbf{x}'|\mathbf{x})$ for $\Delta t \to 0$.

The diffusion approximation to this process is defined by the stochastic differential equation $d\mathbf{x}(t) = \mathbf{f}(\mathbf{x})dt + \mathbf{D}^{1/2}(\mathbf{x})d\mathbf{w}(t)$, for which the drift vector $\mathbf{f}$ and the diffusion matrix $\mathbf{D}$ agree with the first and second jump moments of the jump process, i.e.

$$\mathbf{f}(\mathbf{x}) = \sum_{\mathbf{x}' \neq \mathbf{x}} f(\mathbf{x}'|\mathbf{x})(\mathbf{x}' - \mathbf{x}) \qquad \mathbf{D}(\mathbf{x}) = \sum_{\mathbf{x}' \neq \mathbf{x}} (\mathbf{x}' - \mathbf{x}) f(\mathbf{x}'|\mathbf{x})(\mathbf{x}' - \mathbf{x})^\top. \tag{1}$$

The probability density of time discretised sample paths of the diffusion process conditioned on noisy observations is given by

$$p(\mathbf{x}_{0:T}|D) \approx \frac{p(\mathbf{x}_0)}{Z} \left[\prod_{t=0,\Delta t,\ldots}^{T-\Delta t} \mathcal{N}(\mathbf{x}_{t+\Delta t}; \mathbf{x}_t + \mathbf{f}(\mathbf{x}_t)\Delta t, \mathbf{D}(\mathbf{x}_t)) \right] \left[\prod_{i=1}^{n} \mathcal{N}(y_i; \mathbf{x}_{t_i}, \sigma^2) \right],$$

which is obtained from an Euler approximation to the SDE and where $\mathcal{N}(x; m, v)$ is the density at x of a Gaussian with mean m and variance v, $D = (y_1, \ldots, y_n)$ are the observations, and σ^2 is the variance of the observation noise. Samples from this density can be obtained by Metropolis-Hastings steps (see [2] for details).

3 Weak noise

The starting point of the weak noise approximation is an exact expression of the conditional marginal density of the state vector $p_t(\mathbf{x}|D) \propto p_t(\mathbf{x}|D_{<t})\, r_t(\mathbf{x})$, which is well known from the theory of hidden Markov models. The first factor $p_t(\mathbf{x}|D_{<t})$ is the conditional distribution of the state based only on the observations $D_{<t} \equiv \{\mathbf{y}_i\}_{t_i<t}$ *before* time t. For times between observations this probability fulfils the *forward* Fokker-Planck equation with jump conditions at the observations. And $r_t(\mathbf{x}) \equiv p(D_{\geq t}|\theta, \mathbf{x}_t = \mathbf{x})$ is the likelihood of *future observations* $D_{\geq t} = \{\mathbf{y}_i\}_{t_i \geq t}$ conditioned on the present state $\mathbf{x}(t) = \mathbf{x}$. The likelihood of all data is then $p(D|\theta) = \sum_{\mathbf{x}} p_0(\mathbf{x}) r_0(\mathbf{x})$, where $p_0(\mathbf{x})$ is the distribution of the initial state. r_t fulfils the Kolmogorov backward equation

$$\left[\frac{\partial}{\partial t} + \mathbf{f}(\mathbf{x}, t)^\top \nabla + \frac{1}{2} \mathrm{Tr}(\mathbf{D}(\mathbf{x}, t) \nabla^\top \nabla) \right] r_t(\mathbf{x}) = 0\,. \tag{2}$$

The weak noise expansion is based on the assumption that typical state vectors are close to a nonrandom time dependent state $\mathbf{b}(t)$. Therefore one sets $\mathbf{x} = \mathbf{b}(t) + \epsilon\mathbf{u}$ with an expansion parameter ϵ (which is set later to 1) and also rescales the *noise* $\mathbf{D} \to \epsilon^2\mathbf{D}$. An expansion of the backward equation up to order ϵ^2 yields

$$r_t(\mathbf{x}) \propto \exp\left[-\frac{1}{2} (\mathbf{x} - \mathbf{b})^\top \mathbf{S}^{-1} (\mathbf{x} - \mathbf{b}) \right], \tag{3}$$

where the macroscopic state $\mathbf{b}$ and the matrix $\mathbf{S}$ satisfy the differential equations

$$\dot{\mathbf{b}} = \mathbf{f}(\mathbf{b}) \qquad \dot{\mathbf{S}} = \mathbf{A}\mathbf{S} + \mathbf{S}\mathbf{A}^\top - \mathbf{D}(\mathbf{b}) \tag{4}$$

with $A_{ij}(t) = \partial_{x_j} f_i(\mathbf{x})\big|_{\mathbf{x}=\mathbf{b}(t)}$.

Using a similar expansion for the forward Fokker Planck equation a Gaussian approximation for $p_t(\mathbf{x}|D)$ is obtained, where the mean state vector $\mathbf{m}$ and the covariance matrix $\mathbf{C}$ evolve according to

$$\dot{\mathbf{m}} = \mathbf{g}(\mathbf{m}) \qquad \dot{\mathbf{C}} = \mathbf{H}\mathbf{C} + \mathbf{C}\mathbf{H}^\top + \mathbf{D}(\mathbf{m}) \tag{5}$$

with $H_{ij}(t) = \partial_{x_j} g_i(\mathbf{x})\big|_{\mathbf{x}=\mathbf{m}(t)}$ and

$$\mathbf{g}(\mathbf{x}, t) \approx \mathbf{f}(\mathbf{x}) - \mathbf{D}(\mathbf{b}(t))\mathbf{S}^{-1}(t)(\mathbf{x} - \mathbf{b}(t))\,. \tag{6}$$

Parameter estimation is based on the total likelihood $p(D|\theta)$ of all observations, which is the result of the backward integration. The expected order of magnitude of the parameters and other prior knowledge can be described in form of a prior distribution $p(\theta)$. Then approximate marginal posteriors are calculated from a Laplace approximation of the posterior density $p(\theta|D) \propto p(D|\theta)\ p(\theta)$. Setting $F(\theta) \equiv -\log\left(p(D|\theta)\ p(\theta)\right)$, Laplace's approximation is given by

$$-\log p(\theta_i|D) \approx F(\theta_i, \theta^*_{\setminus i}) + C + \frac{1}{2}\log\left|\frac{\partial^2 F(\theta_i, \theta_{\setminus i})}{\partial\theta^2}\right|_{\theta=\theta^*}, \tag{7}$$

where θ^* denotes the most likely parameters, i.e. $\theta^* = \arg\min_\theta F(\theta)$, and $\theta_{\setminus i}$ all parameters *without* θ_i.

4 Comparison of Weak Noise and MCMC

Figure 1 compares the results of parameter estimation for MCMC sampling based on the approach of [2] and our weak noise approximation [3, 5]. Both methods have been implemented in Matlab. Obtaining 500,000 samples (50,000 discarded as burn-in, thinning factor 100) from MCMC took roughly 80.5 hours on a Intel Core 2 processor, while the approximate inference algorithm ran only for one hour. It is clearly visible, that it produces results comparable to those obtained by MCMC sampling, although it is vastly faster.

Results of state inference using both algorithms are shown in figure 2. Here the parameters of the model have been fixed to their true values $\alpha = 0.05$, $\beta = 0.01$, $\gamma = 0.05$, and $\delta = 0.01$. Obtaining 500,000 samples took 44.5 hours, while it was possible to calculate the marginal posterior using approximate inference in less than one minute. However, both results are nearly identical. Consequently, using the weak noise approximation enables very fast parameter estimation and state inference without loosing much accuracy.

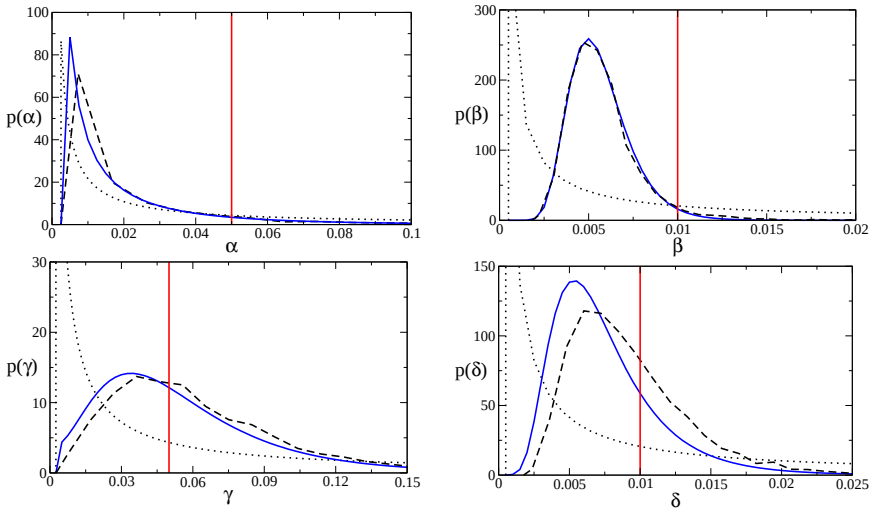

Fig. 1. Posterior distributions of the rate constants. Solid blue lines show the results of the approximation, while the histograms obtained from MCMC are plotted as dashed black lines. The prior used in both algorithms is denoted by dotted lines.

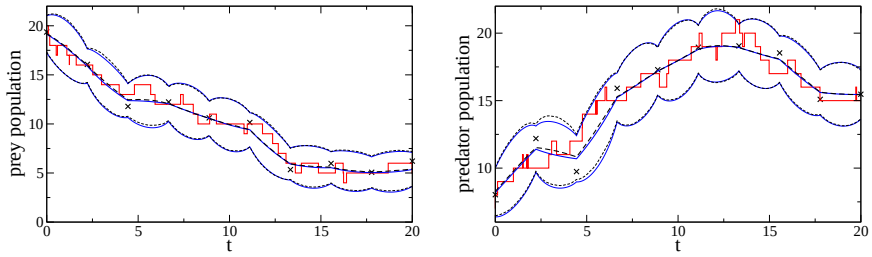

Fig. 2. Posterior distributions of the path. Solid blue lines show the results of the approximation, those obtained from MCMC are plotted as dashed black lines. The mean is denoted by thick lines, while thin lines surround the 95%-confidence interval. The true process and the observations taken from it are drawn as a red line and red crosses, respectively.

References

1. Neil D. Lawrence, Mark Girolami, Magnus Rattray, and Guido Sanguinetti, editors. *Learning and Inference in Computational Systems Biology.* MIT Press, 2010.
2. A. Golightly and D. J. Wilkinson. Markov chain monte carlo algorithms for sde parameter estimation. In Lawrence et al. [1], chapter 9, pages 253–275.
3. Andreas Ruttor, Guido Sanguinetti, and Manfred Opper. Approximate inference for stochastic reaction processes. In Lawrence et al. [1], chapter 10, pages 189–205.
4. N. G. van Kampen. *Stochastic Processes in Physics and Chemistry.* North-Holland, Amsterdam, 1981.
5. Andreas Ruttor and Manfred Opper. Efficient statistical inference for stochastic reaction processes. *Phys. Rev. Lett.*, 103(23):230601, 2009.