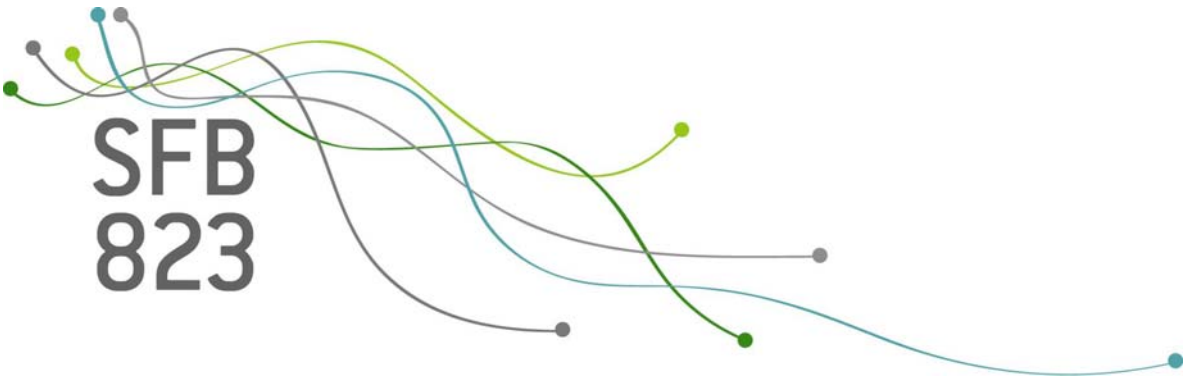
SFB 823

Bayesian prediction for stochastic processes

Simone Hermann

Nr. 27/2016

Discussion Paper



Bayesian prediction for stochastic processes

based on the Euler approximation scheme

Simone Hermann*

June 2, 2016

In many fields of statistical analysis, one is not only interested in estimation of model parameters, but in a prediction for future observations. For stochastic processes, on the one hand, one can be interested in the prediction for the further development of the current, i.e. observed, series. On the other hand, prediction for a new series can be of interest. This work presents two Bayesian prediction procedures based on the transition density of the Euler approximation, that include estimation uncertainty as well as the model variance. In a first algorithm, the pointwise predictive distribution is calculated, in a second, trajectories will be drawn. Both methods will be compared and analyzed with respect to their advantages and drawbacks and set in contrast to two commonly used prediction approaches.

Keywords: Stochastic differential equation, (jump) diffusion, predictive distribution.

1 Introduction

In many research areas, stochastic processes play an important role in stochastic modeling. There is much research done in the field of estimation of the model parameters, see, for example, Bayesian estimation in Donnet et al. (2010); Fuchs (2013); Rifo and Torres (2009).

If the process is defined by a stochastic differential equation (SDE), in many cases, an explicit solution is not available. Without an analytical representation of the process in each time point, a distribution is missing as well, and Bayesian prediction is not straightforward. Usually the SDE is approximated, for example with the Euler scheme, and an iterative transition density is available. In this case, prediction for the next observation is computable. In many areas, predictions for a far future is desired.

In the frequentist estimation, in many cases, a predictive distribution is not calculable and, therefore, trajectories are simulated from the model with point estimates plugged in, see, for example, Chiquet et al. (2009). This procedure neglects the estimation uncertainty, since the point estimation is treated as the true value. In the case of large sample sizes, the estimation uncertainty is small and this approach might work. In Vidoni (2004), a predictive density

*TU Dortmund University, Faculty of Statistics, Vogelpothsweg 87, D-44221 Dortmund, Germany, hermann@statistik.tu-dortmund.de

for time discrete stochastic processes is calculated, but prediction is only made for the next observation. In the Bayesian approach, it is common to take samples from the posterior and simulate with each of them one trajectory. All trajectories together form prediction samples, see, for example, Weinberg et al. (2007). As we will see in the remainder of this work, this procedure leads to large prediction intervals. If only point predictions are desired, this approach might work. But if prediction intervals are of interest, they are desired to be precise, i.e. small while being reliable.

In the case of an explicit solution, a distribution of the process variable in each time point is available and can be taken to calculate the Bayesian predictive distribution. This is done, for example, for the Ornstein-Uhlenbeck and the Cox-Ingersoll-Ross process in a mixed effects model in Dion et al. (2016). But in the case of approximated variables with Euler, this prediction can be imprecise for large time differences, because the approximation becomes inaccurate.

In the following, a novel approach to predict stochastic processes, based on the Euler approximation, is presented. This procedure is valuable for general diffusions as well as for jump diffusions, including the estimation uncertainty as well as possible uncertainties from latent variables. We will first present an algorithm leading to a pointwise distribution of the interesting observation variable, and afterwards, a second algorithm yields a vector of process variables, i.e. a trajectory.

In the next section, the model is presented and estimation is shortly motivated. Afterwards, the Bayesian prediction procedure is explained. In Section 4, a simulation study compares the presented methods to the two mentioned commonly used sampling methods. Afterwards, the results are summed up and a short outlook is given.

2 Model

In this work, we consider the stochastic process $\{Y_t, t \in [0,T]\}$ on the time interval $[0,T]$ defined by the SDE given by

$$dY_t = b(\phi, t, Y_t)\, dt + s(\gamma, t, Y_t)\, dW_t + h(\eta, t, Y_t)\, dN_t,\ Y_0 = y_0, \tag{1}$$

where $\{W_t,\ t \in [0,T]\}$ denotes a Brownian motion on $[0,T]$ and $\{N_t,\ t \in [0,T]\}$ denotes a non-homogeneous Poisson process (NHPP) with cumulative intensity function $\Lambda_\xi(t)$, which has to be bounded on $[0,T]$. Both, Brownian motion and Poisson process, have to be stochastically independent.

In the following, we assume to have observations in time points $0 \leq t_0 < t_1 < ... < t_n \leq T$. In some applications, the Poisson process variables are observed. In that case, the parameter to estimate are $\theta = (\phi, \gamma^2, \eta, \xi)$. If the Poisson process is unobserved, this is a latent variable, which has to be estimated too, and $\theta = (\phi, \gamma^2, \eta, \xi, N_{t_1}, ..., N_{t_n})$.

If an explicit solution of the SDE is available, a distribution $p(Y_t|\theta)$, for each time point t is accompanied and the predictive distribution

$$p(Y_t|Y_{(n)}) = \int p(Y_t|\theta) p(\theta|Y_{(n)})\, d\theta$$

comes along. But if an explicit solution it not available, the process has to be approximated. Well known and widely used is the Euler approximation, see for example Platen and Bruti-Liberati (2010) or for ordinary diffusions Fuchs (2013) or Kloeden and Platen (1992). Define

the Euler approximated variables by

$$
\begin{aligned}
&Y_0 = y_0(\phi), \qquad (2)\\
&Y_i = Y_{i-1} + b(\phi, t_{i-1}, Y_{i-1})\,\Delta_i + s(\gamma, t_{i-1}, Y_{i-1})\sqrt{\Delta_i}\,\xi_i + h(\eta, t_{i-1}, Y_{i-1})\Delta N_i,\\
&\xi_i \sim \mathcal{N}(0,1), \quad \Delta N_i \sim \text{Pois}\left(\Lambda_\xi(t_i) - \Lambda_\xi(t_{i-1})\right),\\
&\Delta_i = t_i - t_{i-1},\ i = 1, ..., n.
\end{aligned}
$$

In the following, we denote the vector of Euler approximated variables with $Y_{(n)} = (Y_0, Y_1, ..., Y_n)$. The observed variables, therefore, are assumed to be realisations of $Y_{(n)}$. Dependent on the Poisson process variables, the likelihood is the product of normal densities. In most cases, the true underlying transition density is not normal. A good overview of possibly resulting problems can be seen in Sørensen (2004) or Beskos et al. (2006) for the case $h(\eta, t, y) = 0$. However, the scope of this work is prediction. This is based on the same discretized variables as used in the estimation, which circumvents the problem of possibly biased estimations.

A good overview of Bayesian estimation for diffusion processes can be found in Fuchs (2013). For jump diffusions with an unobserved Poisson process, a filtering problem occurs. This is tackled, for example, in Roberts and Papaspiliopoulos (2004); Rifo and Torres (2009) and Hermann and Ruggeri (2016). In many cases, one does not have an explicit representation of the posterior $p(\theta|Y_{(n)})$ but simulates from it with MCMC methods and gets samples $\theta_1^*, ..., \theta_K^* \sim p(\theta|Y_{(n)})$ from the posterior distribution.

3 Prediction

For a prediction for Y^* as the Euler approximated variable for Y_{t^*} in t^*, we can approximate the predictive density by

$$
p(Y^*|Y_{(n)}) = \int p(Y^*|Y_{(n)}, \theta) \cdot p(\theta|Y_{(n)})\, d\theta \approx \frac{1}{K}\sum_{k=1}^{K} p(Y^*|Y_{(n)}, \theta_k^*).
$$

For prediction of the further development of the process, the Markov property $p(Y^*|Y_{(n)}, \theta) = p(Y^*|Y_n, \theta)$ can be used. In the case of prediction for a new series, Y^* is independent of the observations variables, and $p(Y^*|Y_{(n)}, \theta) = p(Y^*|\theta)$.

In the following, two versions of Bayesian prediction for the model in (1), are presented. The first algorithm considers the predictive distribution of one point of interest, that can lie far away from the last observed value. The problem, that the Euler approximation can become inaccurate for large time differences, will be tackled. A second algorithm is presented to sample a trajectory vector. Sampling from an arbitrary multivariate distribution is not self-evident. Both algorithms can be used equally for the prediction of a new series as well as for prediction of the further development of the series. We declare the specific points with a and b in the algorithms.

Let us assume, we want to predict the process in time point $t^* \gg t_n$, resp. $t^* \gg t_0$. As mentioned, for large time distances, the Euler approximation can be inaccurate. Therefore, the idea is to go in M steps to the interesting value and each include the predictive distribution of the last step. Therefore, we choose a sampling partition $\{t_0, t_n\} \ni \tau_0 < \tau_1 < ... < \tau_M = t^*$, for example

$$
\tau_m = m \cdot \Delta^* + t_n,\ m = 0, ..., M, \quad \Delta^* := \frac{t^* - \tau_0}{M}.
$$

Let $p(Y^*_{m+1}|Y^*_m, \theta)$ denote the transition density of the Euler approximated variables $Y^*_0 \in \{y_0, Y_n\}, Y^*_1, ..., Y^*_M = Y^*$ in $\tau_0, \tau_1, ..., \tau_M$. In the following, we denote with $\theta^*_1, ..., \theta^*_K$ samples from the posterior distribution $p(\theta|Y_{(n)})$ for the models without latent variable, which is the case for general diffusions, i.e. $h(\eta, t, y) = 0$. For the jump diffusion, prediction for the Poisson process variables $N_{\tau_1}, ..., N_{\tau_M}$ has to be made first, see for example Hermann et al. (2015), and $\theta^*_1, ..., \theta^*_K$ contain the posterior samples of the parameters as well as the prediction samples of the Poisson process variables. We sample from the pointwise predictive distribution by running the following algorithm.

Algorithm 1
Take samples $\theta^*_1, ..., \theta^*_K$ from the posterior distribution $p(\theta|Y_{(n)})$, respectively from the predictive distribution of the Poisson process.

(i) Set $m = 1$ and

(ia) $Y_0^{*(k)} := Y_n, k = 1, ..., K$, prediction for the current series,

(ib) $Y_0^{*(k)} := y_0, k = 1, ..., K$, prediction for a new series.

(ii) Draw K samples

$$Y_m^{*(1)}, ..., Y_m^{*(K)} \sim \frac{1}{K} \sum_{k=1}^{K} p\left(Y^*_m | Y^{*(k)}_{m-1}, \theta^*_k\right).$$

(iii) If $m = M$ stop,
else $m = m + 1$ and go to step (ii).

This algorithm yields samples from the predictive distribution $p(Y^*_m|Y_{(n)}), m = 1, ..., M$ for each of the points $\tau_1, ..., \tau_M$, but no trajectories. This can be seen as follows. For $M = 2$ it is

$$Y_1^{*(r)} \sim \frac{1}{K} \sum_{k=1}^{K} p\left(Y^*_1 | Y^{*(k)}_0, \theta^*_k\right) \approx \int p(Y^*_1|Y^*_0, \theta) p(\theta|Y_{(n)})\, d\theta = p(Y^*_1|Y_{(n)}), \quad r = 1, ..., K$$

for $Y^*_0 \in \{y_0, Y_n\}$. In the second step it is

$$Y_2^{*(r)} \sim \frac{1}{K} \sum_{k=1}^{K} p\left(Y^*_2 | Y^{*(k)}_1, \theta^*_k\right) \approx \int p(Y^*_2|Y^*_1, \theta) p(Y^*_1, \theta|Y_{(n)})\, d(Y^*_1, \theta) = p(Y^*_2|Y_{(n)})$$

for $r = 1, ..., K$. There are three approximations in one step as can be seen by

$$\begin{aligned}
p(Y^*_2|Y_{(n)}) &= \int p(Y^*_2|Y^*_1, \theta) p(Y^*_1, \theta|Y_{(n)})\, d(Y^*_1, \theta) \\
&\approx \int p(Y^*_2|Y^*_1, \theta) p(Y^*_1|Y_{(n)}) p(\theta|Y_{(n)})\, d(Y^*_1, \theta) \qquad (3) \\
&\approx \frac{1}{K_1 K_2} \sum_{k_1=1}^{K_1} \sum_{k_2=1}^{K_2} p\left(Y^*_2 | Y^{*(k_1)}_1, \theta^*_{k_2}\right) \qquad (4) \\
&\approx \frac{1}{K} \sum_{k=1}^{K} p\left(Y^*_2 | Y^{*(k)}_1, \theta^*_k\right). \qquad (5)
\end{aligned}$$

Approximation (3) is related to the facts that $p(Y_1^*, \theta | Y_{(n)}) = p(Y_1^* | \theta, Y_{(n)}) p(\theta | Y_{(n)})$ and $p(Y_1^* | \theta, Y_{(n)}) = p(Y_1^* | Y_0^*, \theta)$ for $Y_0^* \in \{y_0, Y_n\}$. Therefore, approximation (3) is reduced to the approximation of $p(Y_1^* | Y_0^*, \theta) \approx p(Y_1^* | Y_{(n)})$. For very large n, the posterior density gets very sharp and converges for $n \to \infty$ to a point distribution in the point that frequentists call the true parameter. Therefore, the approximation (3) is just a finite sample approximation.

Example 1
We will give an example to illustrate this by the process (1) *with* $h(\eta, t, y) = 0$, $s(\gamma, t, y) = \gamma$ *and* $b(\phi, t, y) = \phi y$. *It is approximated by*

$$Y_i = Y_{i-1}(1 + \phi\, \Delta_i) + \gamma \sqrt{\Delta_i}\, \xi_i, i = 1, ..., n$$

with $Y_0 = y_0$ *and* $\xi_i \sim \mathcal{N}(0, 1)$. *We fix* $y_0 = 0.5, \phi = 0.1$ *and* $\gamma = 0.01$ *for the simulation and assume prior distributions* $\phi \sim \mathcal{N}(0.1, 0.1^2)$ *and* $\gamma^2 \sim IG(3, 2 \cdot 0.01^2)$.

We simulate one random sample of the process in $t_0 = 0, ..., t_{10^4} = 1$ and estimate ϕ and γ^2 in a Gibbs sampler each with the first 100, the first 10^3 and the whole data series. In Figure 1 on the left side, we see in black the density of Y_1, namely the density of the $\mathcal{N}(y_0 \cdot (1 + \phi\Delta_1), \gamma^2\Delta_1)$ distribution. The colored lines each show the predictive densities $p(Y_1^* | Y_{(n)})$ with $n \in \{100, 10^3, 10^4\}$ and evolve to $p(Y_1^* | \theta)$. The blue predictive density with the sample size of 10^4 is nearly the same as the true one. For comparison, the same figure can be seen on the right side of Figure 1 for a second example.

Example 2
We consider process (1) *with* $h(\eta, t, y) = 0$, $s(\gamma, t, y) = \gamma y$ *and* $b(\phi, t, y) = \phi y$. *It is approximated by*

$$Y_i = Y_{i-1}(1 + \phi\, \Delta_i) + \gamma Y_{i-1} \sqrt{\Delta_i}\, \xi_i,\ i = 1, ..., n$$

with $Y_0 = y_0$ *and* $\xi_i \sim \mathcal{N}(0, 1)$. *We simulate random samples of the process with* $y_0 = 1, \phi = 2$, $\gamma = 1$ *and assume prior distributions* $\phi \sim \mathcal{N}(2, 2^2)$ *and* $\gamma^2 \sim IG(3, 2 \cdot 1)$.

The differences between the four curves, see the right side of Figure 1, is very small because of the small estimation uncertainty in comparison to the model variance.

Approximation (4) is explained by the density approximations $p(Y_1^* | Y_{(n)}) \approx \frac{1}{K_1} \sum_{k_1=1}^{K_1} \mathbb{1}_{\{Y_1^{*(k_1)}\}}(Y_1^*)$ and $p(\theta | Y_{(n)}) \approx \frac{1}{K_2} \sum_{k_2=1}^{K_2} \mathbb{1}_{\{\theta_{k_2}^*\}}(\theta)$, which can be arbitrary good according to the choice of K_1 and K_2.

Approximation (5) is due to the fact, that approximation (4) is computationally very costly. In Figure 2 we compare the two versions with $K = K_1 = K_2 = 1000$. There is no difference identifiable between the two curves. The calculations were made with an Intel(R) Core(TM) i5-3470 CPU (3.20GHz), 8 GB RAM, Windows 7 64bit. For comparison, version (5) took 2.62 seconds versus 1689.54 for version (4).

For arbitrary M, we can assume $Y_{M-1}^{*(k)} \sim p(Y_{M-1}^* | Y_{(n)}), k = 1, ..., K$, and with induction follows

$$Y_M^{*(r)} \sim \frac{1}{K} \sum_{k=1}^{K} p\left(Y_M^* | Y_{M-1}^{*(k)}, \theta_k^*\right) \approx \int p(Y_M^* | Y_{M-1}^*, \theta) p(Y_{M-1}^*, \theta | Y_{(n)})\, d(Y_{M-1}^*, \theta) = p(Y_M^* | Y_{(n)})$$

for $r = 1, ..., K$ with the same approximations as described before.

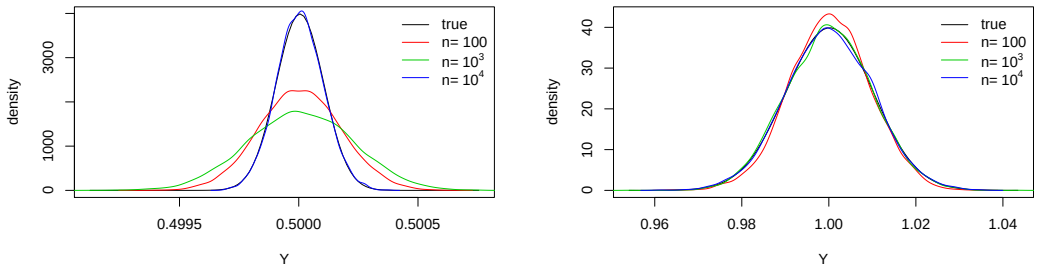


Figure 1: Comparison of the predictive densities for sample sizes $n \in \{100, 10^3, 10^4\}$ to the true density from the model definition; (left) the $\mathcal{N}(0.5 \cdot (1 + 0.1\Delta_1), 0.1\Delta_1)$ density and (right) the $\mathcal{N}(1 + 2\Delta_1, \Delta_1)$ density; posterior and predictive density each approximated by 10^4 samples.

The following algorithm is made for the case that trajectories are the value of interest. Here, it is possible not to draw exactly M samples from the trajectory, but to choose a critical value y_c and to stop if it is reached. In addition, it is possible to draw $L \neq K$ samples.

Algorithm 2
Take the samples $\theta_1^*, ..., \theta_K^* \sim p(\theta|Y_{(n)})$ from the posterior, respectively from the predictive distribution of the Poisson process variables. For one trajectory sample, repeat the following steps.

(i) Set $m = 1$ and

(ia) $Y_0^* := Y_n$, prediction for the current series,

(ib) $Y_0^* := y_0$, prediction for a new series.

(ii) Draw one sample

$$Y_m^* \sim \frac{1}{K} \sum_{k=1}^{K} p\left(Y_m^* | Y_{m-1}^*, \theta_k^*\right).$$

(iii) If $m = M$ (or $Y_m^* \geq y_c$) stop,
else $m = m + 1$ and go to step (ii).

For notation simplicity, in the following, we restrict to the case of M trajectory entries. Repeating the algorithm L times, leads to L independent samples of $p(Y_1^*, ..., Y_M^* | Y_{(n)})$. This can be seen as follows. At first, it is

$$\begin{aligned} p(Y_m^* | Y_{m-1}^*, Y_{(n)}) &= \int p(Y_m^* | Y_{m-1}^*, \theta) p(\theta | Y_{m-1}^*, Y_{(n)}) \, d\theta \\ &\approx \int p(Y_m^* | Y_{m-1}^*, \theta) p(\theta | Y_{(n)}) \, d\theta \\ &\approx \frac{1}{K} \sum_{k=1}^{K} p\left(Y_m^* | Y_{m-1}^*, \theta_k^*\right). \end{aligned} \tag{6}$$

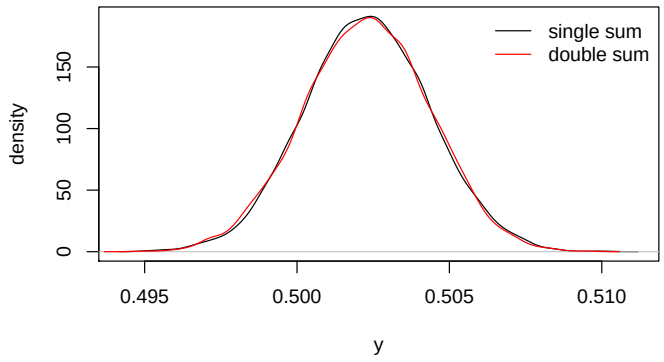


Figure 2: Investigation of approximation (5) with Example 1, approximation of density with 10^4 samples; in red: $\frac{1}{K_1 K_2}\sum_{k_1=1}^{K_1}\sum_{k_2=1}^{K_2} p\left(Y_2^*|Y_1^{*(k_1)},\theta_{k_2}^*\right)$, in black: $\frac{1}{K}\sum_{k=1}^{K} p\left(Y_2^*|Y_1^{*(k)},\theta_k^*\right)$, $K = K_1 = K_2 = 1000$.

The last approximation is the same as in (4). Approximation (6) again is a finite sample approximation. For large n, one sample Y_{m-1}^* does not effect the posterior. By the way, Y_{m-1}^* is itself drawn by the predictive distribution based on the same posterior samples. Therefore, even in finite sample sizes, the effect is negligible.

Afterwards, it is approximated

$$\begin{aligned} Y_1^* &\sim p(Y_1^*|Y_{(n)}) \\ Y_2^* &\sim p(Y_2^*|Y_1^*, Y_{(n)}) \\ &\vdots \\ Y_M^* &\sim p(Y_M^*|Y_{M-1}^*, Y_{(n)}) \end{aligned}$$

and, therefore, the joint density is

$$p(Y_1^*, ..., Y_M^*|Y_{(n)}) = p(Y_1^*|Y_{(n)}) \cdot p(Y_2^*|Y_1^*, Y_{(n)}) \cdot ... \cdot p(Y_M^*|Y_{M-1}^*, Y_{(n)})$$

and we get samples from $p(Y_1^*, ..., Y_M^*|Y_{(n)})$ with Algorithm 2.

4 Simulation study

To see the difference between Algorithm 1 and 2, we will compare them in a simulation study. 100 series are simulated from Example 2 with $n = 50$ and $t \in [0, 1]$, and the parameters are estimated. Based on the likelihood, a Metropolis within Gibbs sampler, see Robert and Casella (2004), is conducted. A Metropolis Hastings step for ϕ is made in each iteration. For γ^2, the inverse gamma distribution is conjugate, which yields an analytically available full conditional posterior. With $K = 10^4$ samples from the posterior distribution, we will compare four prediction methods. The first two are the presented ones in Algorithm 1 and 2. The third

method is simulation of K trajectories from the model based on the point estimation, here the posterior mean, as often used in the frequentist approach and will, therefore, be named frequentist sampling in the following. The fourth one is common in Bayesian analysis, where each of the K posterior samples is taken to simulate one trajectory from the model. All K trajectories also form a prediction, where pointwise mean or quantiles could be used for point prediction or prediction intervals. We will name it posterior sampling in the following.

The methods used in this work are implemented in an R package `BaPreStoPro`, see R Core Team (2015), available on `github.com/SimoneHermann/BaPreStoPro`.

We need a criterion for the quality of the procedure. The coverage rate is one obvious criterion, since the level has to be held. But the interval size is also interesting. Gneiting and Raftery (2007) propose an interval score, which is a combination between interval size, which should be small, and coverage rate, which should be equal to or larger than the level $1-\alpha$, given by

$$S(l,u,y) = u - l + \frac{2}{\alpha}(l-y)\mathbb{1}_{(-\infty,l)}(y) + \frac{2}{\alpha}(y-u)\mathbb{1}_{(u,\infty)}(y).$$

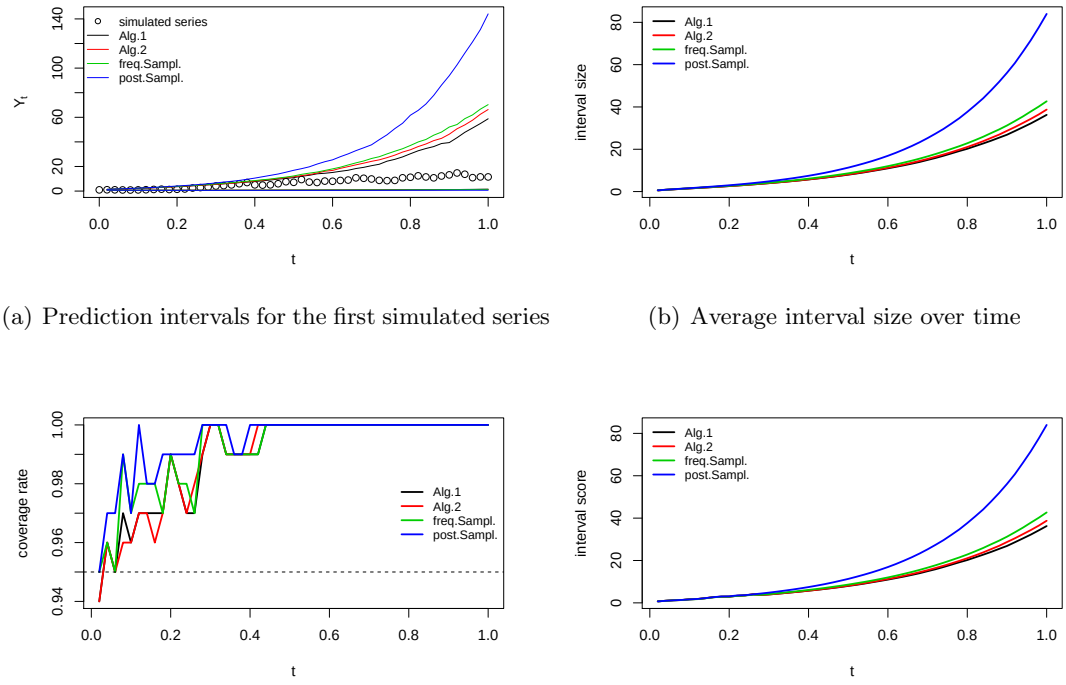
l denotes the lower and u the upper bound of the interval and y the true, i.e. the observed, value. In the case of a non-covering interval, the deviance from the respective bound is highly penalized by the deviance itself multiplied with $\frac{2}{\alpha}$.

We see in Figure 3(a) the first simulated series with the corresponding prediction intervals for the four methods. All methods have the lower bound in common. But the upper interval bounds differ. The posterior sampling, i.e. the blue line, leads to the biggest intervals which is not surprising since we simulate with all the posterior samples instead of integrating out. The difference between the red and the green lines, i.e. the Algorithm 2 and the frequentist sampling, is small. The procedures are comparable, the only difference is, that Algorithm 2 includes the estimation uncertainty, whereby the frequentist sampling only takes the point estimations into account and the only uncertainty stems from the model. The smallest intervals are given by the black line, i.e. Algorithm 1, which is also not surprising, because it calculates pointwise instead of simultaneous predictions. The results of all the 100 series can be seen in Figures 3(b), 3(c) and 3(d). Because of the large coverage rates, see Figure 3(c), there is almost no difference between the interval size and the score. In both Figures, 3(b) and 3(d), the black line, i.e. Algorithm 1, has the smallest sizes and scores, followed by the red lines, i.e. Algorithm 2. The frequentist sampling surprisingly has larger intervals and the posterior sampling, as expected, has the largest intervals. Except a small deviance in the beginning for the black and the red line, all prediction intervals hold the level.

The same simulation study was made for Example 1 as well. The differences between the Algorithms 1, 2 and frequentist sampling were not visually noticeable. The small difference between Algorithm 1 and 2 can be caused by the small model variance. This, in addition, leads to a small estimation uncertainty and, therefore, the frequentist sampling is comparable.

Algorithm 1 has also been investigated in a simulation study for six growth curves for drift function b and constant variance function $s(\gamma,t,y)=\gamma$, and applied to crack growth in aluminum alloy in Hermann et al. (2016).

Since the methods have to be applicable, running times are an interesting fact in the decision for one procedure. We test for Example 1 with $n=50$ the running times with the sampling algorithms inversion method and rejection sampling, see Devroye (1986), both on a fixed vector of candidates (discrete) and free in an interval (continuous). We take $K=2000$ samples from the posterior and calculate in each step $K=L=2000$ simulations from the



(a) Prediction intervals for the first simulated series

(b) Average interval size over time

(c) Coverage rates

(d) Average interval score over time

Figure 3: Comparison of the prediction methods in Algorithms 1 and 2, and the posterior and frequentist sampling, each depicted for time points $t_1 = 0.02, ..., t_{50} = 1$, for the diffusion model in Example 2.

predictive distribution to have a fair comparison. In addition, in each procedure, the candidate area to sample Y^*_{m+1} is chosen in the same way by

$$\Big[\min_{k=1}^{K} Y_m^{*(k)} + b(\widehat{\phi}, \tau_m, \widehat{Y}^*_m)\Delta^* - 5s(\sqrt{\widehat{\gamma}^2}, \tau_m, \widehat{Y}^*_m), \max_{k=1}^{K} Y_m^{*(k)} + b(\widehat{\phi}, \tau_m, \widehat{Y}^*_m)\Delta^* + 5s(\sqrt{\widehat{\gamma}^2}, \tau_m, \widehat{Y}^*_m)\Big]$$

with $\widehat{\phi}$ and $\widehat{\gamma}^2$ the point estimation, i.e. the posterior mean, and $\widehat{Y}^*_m = \frac{1}{K}\sum_{k=1}^{K} Y_m^{*(k)}$.

In the case of vectorial sampling, we fix for comparison $C \in \{5\,000, 10\,000\}$ points $y_l = y_1 < y_2 < ... < y_C = y_u$ between lower and upper bound of the interval. In the other case, for the inversion method, we apply binary search until a fineness degree of $\frac{y_u - y_l}{C}$ to have a fair comparison. Again, the calculations are made with an Intel(R) Core(TM) i5-3470 CPU (3.20GHz), 8 GB RAM, Windows 7 64bit.

In Table 1, the resulting running times are displayed, averaged over 10 runs. We clearly see the advantage of Algorithm 1 that turns out in the vectorial sampling.

	Algorithm 1				Algorithm 2	
	vectorial					
	$C = 5000$	$C = 10000$	$C = 5000$	$C = 10000$	$C = 5000$	$C = 10000$
Rejection sampling	68.77	141.30	354.97	359.62	349.39	356.93
Inversion method	85.47	172.14	445.68	491.02	398.23	443.01

Table 1: Overview of the running times of the two presented algorithms for the different sampling methods.

Additionally, we investigate the four prediction methods, Algorithm 1 and 2, frequentist and posterior sampling, for the full jump diffusion model.

Example 3
We will give an example by the process (1) *with* $b(\phi, t, y) = \phi$, $s(\gamma, t, y) = \gamma$, $h(\eta, t, y) = \eta y$ *and* $\Lambda_\xi(t) = \left(\frac{t}{\xi_2}\right)^{\xi_1}$. *It is approximated by*

$$Y_i = Y_{i-1} + \phi\,\Delta_i + \gamma\sqrt{\Delta_i}\,\xi_i + \eta Y_{i-1}\Delta N_i,\ i = 1, ..., n$$

with $Y_0 = y_0$, $\xi_i \sim \mathcal{N}(0, 1)$ *and* $\Delta N_i \sim Pois(\Lambda_\xi(t_i) - \Lambda_\xi(t_{i-1}))$. *We fix* $y_0 = 0.5, \phi = 0.2, \gamma = 0.5, \eta = 0.1, \xi = (2, 0.2)$ *and estimate non-informative, i.e.* $\phi \sim p(\phi) = 1, \eta \sim p(\eta) = 1, \gamma^2 \sim p(\gamma^2) = 1$ *and* $\xi \sim p(\xi) = 1$.

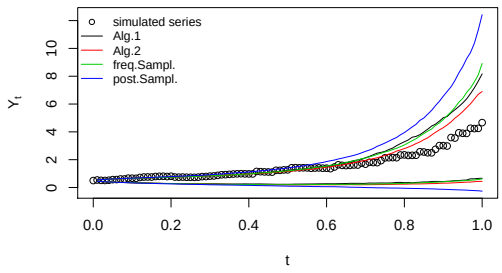
We run a simulation study with 100 series from Example 3 with sample size $n = 100$ and $t \in [0, 1]$. For the estimation, we assume the Poisson process to be observed and conduct a Metropolis within Gibbs sampler for the parameter $\theta = (\phi, \gamma^2, \eta, \xi)$.

For Algorithm 1, we sample predictions for $\Delta N_i \sim \frac{1}{K}\sum_{k=1}^{K} p(\Delta N_i | \Lambda_{\xi_k^*}(t_i) - \Lambda_{\xi_k^*}(t_{i-1}))$ in each step with $p(\cdot|\lambda)$ being the density of the Poisson distribution with parameter λ. For Algorithm 2, we simulate trajectories for the Poisson process variables. This is done by sampling trajectory samples of the event times $T_1^* < T_2^* < ...$ with Algorithm 2 until τ_M is reached. The trajectory samples of the Poisson process can be calculated as follows

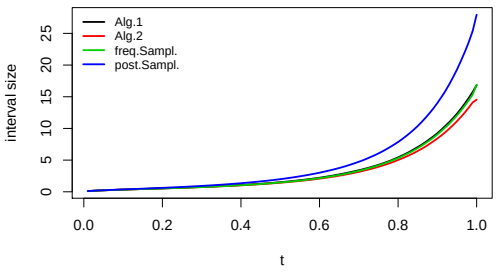
$$N_t^* = \{j : T_j^* \leq t < T_{j+1}^*\}. \tag{7}$$

In Figure 4, we see the results. From the coverage rates, it is difficult to make a qualitative comparison. The sizes and the interval score draw the same picture. The red lines show, that Algorithm 2 has the smallest interval sizes, followed by Algorithm 1 and frequentist sampling. The difference between Algorithm 1 and frequentist sampling is hardly noticeable, since the green line overdraws the black line. To investigate this effect, the same simulation study was made with Algorithm 1 based on the same trajectory-wise sampling for the Poisson process as for Algorithm 2. The result is, that the black and the red line are equal, see Figure 5, where the red line overdraws the black one.

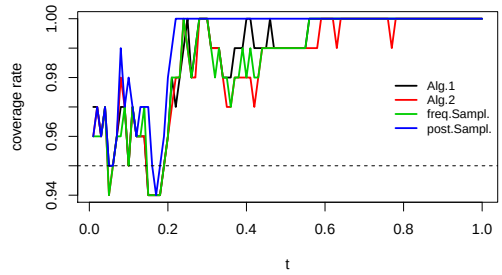
In summary, we can state, that the trajectory-wise Poisson process sampling brings the best results for the jump diffusion. But, if we compare the running times, the procedure to get the results for the black lines in Figure 4, took between 699 and 1183 seconds, whereas the procedure for the black lines in Figure 5 took between 5576 and 12116 seconds.



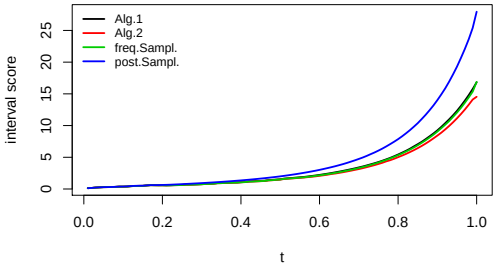
(a) Prediction intervals for the first simulated series



(b) Average interval size



(c) Coverage rates



(d) Average interval score

Figure 4: Comparison of the prediction methods in Algorithms 1 and 2 and the posterior and frequentist sampling, each in time points $t_1 = 0.01, ..., t_{100} = 1$, for Example 3.

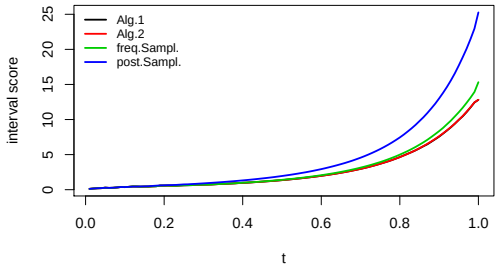


Figure 5: Interval score for Example 3 with Algorithm 1 based on trajectory-wise samples for the Poisson process vector, see (7).

5 Conclusion

A Bayesian prediction approach has been presented for stochastic processes defined by their differential equation, which is approximated with the Euler scheme. The model variance as well as estimation uncertainty is included and, in the case of the jump diffusion, the uncertainty of the Poisson process as well. Two algorithms have been presented, one for a pointwise predictive distribution, and a second for the prediction of a trajectory. We showed in a simulation study, that both procedures lead to a reliable and precise prediction.

It will be future work to investigate other approximation schemes than the one of Euler. There is also a wide field of time-discrete autoregressive process models, where the presented methods could be applicable.

Acknowledgement

This work has been supported by the Collaborative Research Center "Statistical modeling of nonlinear dynamic processes" (SFB 823) of the German Research Foundation (DFG) in project B5 "Statistical methods for damage processes under cyclic load".

References

Beskos, A., O. Papaspiliopoulos, G. O. Roberts, and P. Fearnhead (2006). "Exact and Computationally Efficient Likelihood-Based Estimation for Discretely Observed Diffusion Processes". *J. R. Statist. Soc. B* 68, pp. 333–382.

Chiquet, J., N. Limnios, and M. Eid (2009). "Piecewise Deterministic Markov Processes Applied to Fatigue Crack Growth". *Journal of Statistical Planning and Inference* 139, pp. 1657–1667.

Devroye, L. (1986). *Non-Uniform Random Variate Generation*. New York: Springer.

Dion, C., S. Hermann, and A. Samson (2016). "Mixedsde: an R Package to Fit Mixed Stochastic Differential Equations". hal-01305574.

Donnet, S., J.-L. Foulley, and A. Samson (2010). "Bayesian Analysis of Growth Curves Using Mixture Models Defined by Stochastic Differential Equations". *Biometrics* 66, pp. 733–741.

Fuchs, C. (2013). *Inference for Diffusion Processes*. 1st ed. Berlin Heidelberg: Springer.

Gneiting, T. and A. E. Raftery (2007). "Strictly Proper Scoring Rules, Prediction and Estimation". *Journal of the American Statistical Association* 102, pp. 359–378.

Hermann, S. and F. Ruggeri (2016). "Modelling Wear in Cylinder Liners". *SFB 823 discussion paper 06/16*.

Hermann, S., K. Ickstadt, and C. H. Müller (2015). "Bayesian Prediction for a Jump Diffusion Process with Application to Crack Growth in Fatigue Experiments". *SFB 823 discussion paper 30/15*.

Hermann, S., K. Ickstadt, and C. H. Müller (2016). "Bayesian Prediction of Crack Growth Based on a Hierarchical Diffusion Model". *Applied Stochastic Models in Business and Industry*. DOI: 10.1002/asmb.2175.

Kloeden, P. E. and E. Platen (1992). *Numerical Solution of Stochastic Differential Equations*. 1st ed. Berlin Heidelberg: Springer.

Platen, E. and N. Bruti-Liberati (2010). *Numerical Solution of Stochastic Differential Equations with Jumps in Finance*. First. Berlin Heidelberg: Springer.

R Core Team (2015). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria.

Rifo, L. L. R. and S. Torres (2009). "Full Bayesian Analysis for a Class of Jump-Diffusion Models". *Communications in StatisticsTheory and Methods* 38, pp. 1262–1271.

Robert, C. P. and G. Casella (2004). *Monte Carlo Statistical Methods*. 2nd ed. New York: Springer.

Roberts, G. O. and O. Papaspiliopoulos (2004). "Bayesian inference for non-Gaussian Ornstein-Uhlenbeck stochastic volatility processes". *J. R. Statist. Soc. B* 66, pp. 369–393.

Sørensen, H. (2004). "Parametric Inference for Diffusion Processes Observed at Discrete Points in Time: a Survey". *International Statistical Review* 72, pp. 337–354.

Vidoni, P. (2004). "Improved Prediction Intervals for Stochastic Process Models". *Journal of Times Series Analysis* 25, pp. 137–154.

Weinberg, J., L. D. Brown, and J. R. Stroud (2007). "Bayesian Forecasting of an Inhomogeneous Poisson Process with Application to Call Center Data". *Journal of the American Statistical Association* 102, pp. 1185–1198.