Efficient estimation for diffusions

Nina Munkholt Jakobsen[*1]

[1] *Department of Mathematical Sciences, University of Copenhagen*

We consider estimation of the diffusion parameter of a diffusion process observed over a fixed time interval. We present conditions on approximate martingale estimating functions under which estimators are rate optimal and efficient in the case of in-fill asymptotics. In this setup, limit distributions of the estimators are non-standard, in the sense that they are usually normal variance-mixtures. In particular, the mixing distribution depends on the full sample path of the diffusion process over the observation time interval. We also present the more applicable result that, after a suitable data-dependent normalisation, estimators converge in distribution to a standard Gaussian limit. The results presented here are joint work with Michael Sørensen, and published in [10].

Keywords: Stochastic differential equations, approximate martingale estimating functions, in-fill asymptotics, rate optimality, stable convergence

1 Introduction

Diffusion processes are used in a variety of fields to model continuous-time dynamics, for instance, in biology, finance, and neuroscience. However, the corresponding data are usually only observable at discrete time-points. Except in a few simple cases, the likelihood function based on the discrete-time observations is not known explicitly. Thus, for parameter estimation, alternatives to maximum likelihood estimation must be considered.

Here, we focus on a one-dimensional diffusion process $(X_t^\theta)_{t\geq 0}$, which solves a stochastic differential equation of the form

$$dX_t^\theta = a(X_t^\theta)\,dt + b(X_t^\theta,\theta)\,dW_t\,,$$

$\theta \in \Theta$, where $(W_t)_{t\geq 0}$ is a standard Wiener process. Let $\theta_0 \in \Theta$ denote the true, unknown parameter. We assume observations of $(X_t^{\theta_0})_{t\geq 0}$ over the fixed time-interval $[0,1]$ at times $t_i^n = i\Delta_n$, $i = 0, 1, \ldots, n$, with $\Delta_n = 1/n$. In the following, we put $X_t = X_t^{\theta_0}$ and $X_i^n = X_{t_i^n}^{\theta_0}$.

[*] Corresponding author: munkholt@math.ku.dk

For simplicity, we assume that $\Theta \subseteq \mathbb{R}$, but an extension of the following results to a multidimensional parameter would be straightforward. Similarly, the observation time interval $[0,1]$ may be generalised to other compact time intervals by rescaling of the drift and diffusion coefficients a and b.

We consider estimators of the diffusion parameter θ, which are based on approximate martingale estimating functions. Many well-known estimators proposed in the literature may be formulated in terms of these estimating functions, see [12]. Our aim is to give a simple characterisation of the estimating functions that produce efficient estimators of the diffusion parameter when the sample size n increases to infinity.

Here, an approximate martingale estimating function $G_n(\theta)$ may be written on the form

$$G_n(\theta) = \sum_{i=1}^{n} g(\Delta_n, X_i^n, X_{i-1}^n, \theta)\,.$$

It is given by a real-valued function $g(t,y,x,\theta)$, which satisfies that for all $\theta \in \Theta$, the conditional expectation

$$\mathbb{E}\left(g\left(\Delta_n, X_{t_i^n}^\theta, X_{t_{i-1}^n}^\theta, \theta\right) \mid X_{t_{i-1}^n}^\theta\right)$$

is of order Δ_n^γ, for some constant $\gamma \geq 2$. A G_n-estimator solves the estimating equation $G_n(\theta) = 0$.

Under other asymptotic scenarios often considered for diffusion processes, limit distributions of estimators are typically Gaussian, with variances depending on θ_0, see e.g. [2, 4, 6, 11, 12]. Under the sampling scheme considered here, the limit distributions are usually normal variance-mixture distributions. In addition to depending on θ_0, these distributions may also depend on the full sample path of the diffusion process over the observation time interval. Estimation and asymptotics under the current observation scheme have previously been treated by, e.g., [1, 3, 5].

It was shown in [1, 5] that under suitable regularity conditions, the model and observation scheme considered here satisfy the local asymptotic mixed normality property with rate $\sqrt{n}$ and random asymptotic Fisher information

$$\mathcal{I}(\theta_0) = 2\int_0^1 \frac{\partial_\theta b(X_s,\theta_0)^2}{b^2(X_s,\theta_0)}\, ds\,.$$

Here, $\partial_\theta b(x,\theta)$ denotes the first partial derivative of b with respect to θ. This result is used to characterise a consistent estimator $\hat{\theta}_n$ as rate optimal and efficient if

$$\sqrt{n}\left(\hat{\theta}_n - \theta_0\right) \xrightarrow{\mathcal{D}} L$$

as $n \to \infty$, where $L = \mathcal{I}(\theta_0)^{-1/2} Z$, with Z standard normal distributed and independent of $\mathcal{I}(\theta_0)$. We may interpret $\sqrt{n}$ as the fastest possible rate of convergence in distribution, and L as the limit distribution with the smallest possible variance, conditionally on $\mathcal{I}(\theta_0)$.

2 Main results

The work presented in [10] establishes existence, uniqueness, and asymptotic distribution results concerning consistent G_n-estimators, addressing the question of their rate optimality and efficiency. The essence of the main results of [10], Theorem 3.2 and Corollary 3.4, is summarized in the following Theorem 1, and Corollaries 1 and 2. Technicalities, as well as the existence and uniqueness results, are omitted here.

Theorem 1. *Assume suitable regularity assumptions. Suppose that*

$$\partial_y g(0, y, x, \theta)\big|_{y=x} = 0 \tag{1}$$

for all x and θ. Then, for any consistent G_n-estimator $\hat{\theta}_n$,

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{D}} W(\theta_0) Z$$

as $n \to \infty$, where Z is standard normal distributed and independent of

$$W(\theta_0) = \frac{\left(2\displaystyle\int_0^1 b^4(X_s, \theta_0)\partial_y^2 g(0, X_s, X_s, \theta_0)^2\, ds\right)^{1/2}}{\displaystyle\int_0^1 \partial_\theta b^2(X_s, \theta_0)\partial_y^2 g(0, X_s, X_s, \theta_0)\, ds}. \tag{2}$$

Here, $\partial_y^2 g$ denotes the second partial derivative of g with respect to y. Condition (1) ensures estimators that converge at the optimal rate $\sqrt{n}$. The proof of Theorem 1 relies on, among others, results from [8, 9], including a stable central limit theorem, Theorem IX.7.28, from [8]. The expression (2) reveals that $W(\theta_0)$ is usually random, and depends on the full sample path of the diffusion process over the observation time interval. For finite sample sizes, this sample path is only observed at discrete time-points. We use properties of stable convergence in distribution to deal with these complications. The result in Corollary 1 below shows that when suitably normalised, the estimators from Theorem 1 converge in distribution to a standard Gaussian limit.

Corollary 1. Assume suitable regularity assumptions, and suppose that (1) holds. Let $\hat{\theta}_n$ be any consistent G_n-estimator. Then

$$\widehat{W}_n = -\frac{\left(\frac{1}{\Delta_n}\sum_{i=1}^{n} g^2(\Delta_n, X_i^n, X_{i-1}^n, \hat{\theta}_n)\right)^{1/2}}{\sum_{i=1}^{n} \partial_\theta g(\Delta_n, X_i^n, X_{i-1}^n, \hat{\theta}_n)}$$

satisfies that $\widehat{W}_n \xrightarrow{\mathcal{P}} W(\theta_0)$, and it holds that

$$\sqrt{n}\,\widehat{W}_n^{-1}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)\,.$$

Finally, the additional condition (3) ensures efficiency of the estimators.

Corollary 2. Assume suitable regularity assumptions. Suppose that (1) and

$$\partial_y^2 g(0,y,x,\theta)\big|_{y=x} = C_\theta \frac{\partial_\theta b^2(x,\theta)}{b^4(x,\theta)} \tag{3}$$

hold for all x and θ, where C_θ is a non-zero constant. Then any consistent G_n-estimator is efficient.

For example, it may be verified that the estimating function given by

$$\tilde{g}(t,y,x,\theta) = \frac{\partial_\theta b^2(x,\theta)}{b^4(x,\theta)}\left((y-x)^2 - tb^2(x,\theta)\right)$$

satisfies (1) and (3), and corresponds to the efficient contrast function in [3], Theorem 5. It should be noted that conditions (1) and (3) also appear in [7, 12] under other sampling scenarios. Consequently, a number of approximate martingale estimating functions discussed in those papers satisfy our rate optimality and efficiency conditions.

3 Simulation study

The paper [10] also includes a simulation study. Visual comparisons are made of distributions pertaining to estimators based on two approximate martingale estimating functions, which are not covered by the theory of [3]. An excerpt from this simulation study is summarized here. Ten thousand sample paths of a diffusion process given by

$$dX_t^\theta = -2X_t^\theta\,dt + (\theta + (X_t^\theta)^2)^{-1/2}\,dW_t \tag{4}$$

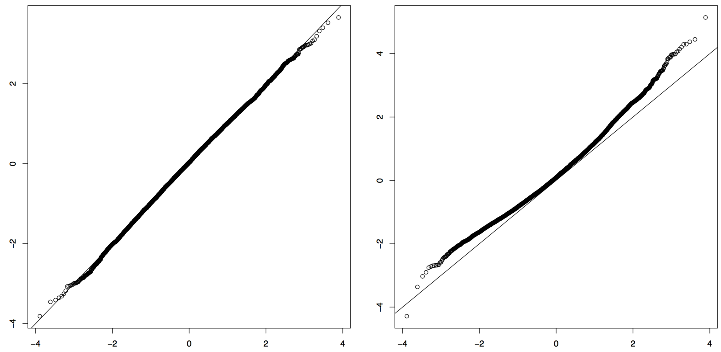

Figure 1: Q-Q plots comparing the distribution of $\sqrt{n}\,\widehat{W}_n^{-1}(\hat{\theta}_n - \theta_0)$ for the efficient (left) and inefficient (right) estimator, respectively, to the standard normal distribution, when $n = 1000$.

were simulated with $\theta_0 = 1$ and $X_0 = 0$, and parameter estimates were computed using the two estimating functions. These estimating functions were given by h and $\tilde{h}$, respectively:

$$h(t, y, x, \theta) = (y - (1 - 2t)x)^2 - (\theta + x^2)^{-1}t$$
$$\tilde{h}(t, y, x, \theta) = (\theta + x^2)^{10} h(t, y, x, \theta)$$

The functions h and $\tilde{h}$ both satisfy the rate-optimality condition (1). However, only h satisfies the efficiency condition (3) for the model (4). Figure 1 shows Q-Q plots comparing the distribution of $\sqrt{n}\,\widehat{W}_n^{-1}(\hat{\theta}_n - \theta_0)$ for the efficient (left) and inefficient (right) estimating function, respectively, to the standard normal distribution, when the sample size is $n = 1000$. In this example from [10], it seems that as the sample size increases, the standard normal distribution becomes a good approximation faster in the efficient case than in the inefficient case. This is an interesting observation, as the current theory does not speak about the speed of this convergence.

Acknowledgements: The work was partially supported by the Danish Council for Independent Research – Natural Science, through a grant to Susanne Ditlevsen. The research is also part of the Dynamical Systems Interdisciplinary Network funded by the University of Copenhagen Programme of Excellence.

References

[1] G. Dohnal. On estimating the diffusion coefficient. *Journal of Applied Probability*, 24(1):105–114, 1987.

[2] V. Genon-Catalot. Maximum contrast estimation for diffusion processes from discrete observations. *Statistics*, 21(1):99–116, 1990.

[3] V. Genon-Catalot and J. Jacod. On the estimation of the diffusion coefficient for multi-dimensional diffusion processes. *Annales de l'institut Henri Poincaré (B) Probabilités et Statistiques*, 29(1):119–151, 1993.

[4] A. Gloter and M. Sørensen. Estimation for stochastic differential equations with a small diffusion coefficient. *Stochastic Processes and their Applications*, 119(3):679–699, 2009.

[5] E. Gobet. Local asymptotic mixed normality property for elliptic diffusion: a Malliavin calculus approach. *Bernoulli*, 7(6):899–912, 2001.

[6] M. Jacobsen. Discretely observed diffusions: Classes of estimating functions and small Delta-optimality. *Scandinavian Journal of Statistics*, 28(1):123–149, 2001.

[7] M. Jacobsen. Optimality and small Delta-optimality of martingale estimating functions. *Bernoulli*, 8(5):643–668, 2002.

[8] J. Jacod and A. Shiryaev. *Limit Theorems for Stochastic Processes*. Number 288 in Grundlehren der mathematischen Wissenschaften. Springer-Verlag, Berlin Heidelberg New York, 2nd edition, 2003.

[9] J. Jacod and M. Sørensen. Aspects of asymptotic statistical theory for stochastic processes. Preprint. University of Copenhagen, 2012.

[10] N. M. Jakobsen and M. Sørensen. Efficient estimation for diffusions sampled at high frequency over a fixed time interval. *Bernoulli*, 23(3):1874–1910, 2017.

[11] M. Kessler. Estimation of an ergodic diffusion from discrete observations. *Scandinavian Journal of Statistics*, 24:211–229, 1997.

[12] M. Sørensen. Efficient estimation for ergodic diffusions sampled at high frequency. Preprint. University of Copenhagen, 2015.