Maximum Likelihood Estimation of an Integrate and Fire Neuronal Model

Paul Mullowney
*Tech-X Corporation, 5621 Arapahoe Avenue, Suite A, Boulder, CO, 80303 (*paulm@txcorp.com*), Department of Mathematics and Statistics, University of Canterbury, Private Bag 4800, Christchurch, NZ*

Satish Iyengar
*Department of Statistics, University of Pittsburgh, 2730 Cathedral of Learning, Pittsburgh, PA, 15260 (*si@stat.pitt.edu*)*

January 22, 2007

Abstract. The Ornstein-Uhlenbeck process has been proposed as a model for the spontaneous activity of a neuron. In this model, the firing of the neuron corresponds to the first passage of the process to a constant boundary, or threshold. While the Laplace transform of the first-passage time distribution is available, the probability distribution function has not been obtained in any tractable form. We address the problem of estimating the parameters of the process when the only available data from a neuron are the interspike intervals, or the times between firings. In particular, we give an algorithm for computing maximum likelihood estimates (MLEs) and their corresponding confidence regions for the three identifiable (of the five model) parameters by numerically inverting the Laplace transform. A comparison of the two-parameter algorithm (where the time constant τ is known a priori) to the three-parameter algorithm shows that significantly more data is required in the latter case to achieve comparable parameter resolution as measured by 95% confidence intervals widths. The computational methods described here are an efficient, biophysically-revealing alternative to the well-known moment-based estimations for OU integrate and fire (IF) models. Moreover, it could serve as a template for performing parameter inference on more complex IF neuronal models.

Keywords: Ornstein-Uhlenbeck process, parameter inference, inverse Laplace transform

1. Introduction

Mean-reverting stochastic processes are common across many areas of science. The Ornstein-Uhlenbeck diffusion (Uhlenbeck and Ornstein, 1954; Tuckwell, 1989) is perhaps the simplest mathematical model of this type and has thus received considerable attention in physics (Chandrasekhar, 1954), meteorology (Gringorten, 1968), population growth (Tuckwell, 1974), and financial asset modeling (Linetsky, 2004). It is also used to model the sub-threshold, spontaneous activity of certain types of individual neurons (Stein, 1965). Accounts of the neurophysiological background leading to the Ornstein-Uhlenbeck process are given

in (Ricciardi, 1977; Ricciardi and Sacerdote, 1977; Tuckwell, 1988). These derivations are dependent on the assumptions that the neuron is connected to a large number of cells within the membrane and that the input from no single cell dominates. Neuroanatomical studies (Tuckwell, 1974) indicate that a neuron can have as many as 10^4 to 10^5 synapses. Bursting neurons are an example of a type of neuron for which second assumption fails and the Ornstein-Uhlenbeck model does not apply.

In many cases, due to technical limitations only the extracellular region is accessible, so the neurons successive firings, or spike train, are the available data. These data are used in several ways. For instance, they provide a quantitative characterization of the neuron in terms of indices such as the mean and variance of the interspike intervals, and their autocorrelation. The effects of different experimental conditions can then be inferred from changes in these indices. Such uses, however, are somewhat crude in the sense that they do not relate the data or indices to any known (or hypothesized) details of the electrophysiology or anatomy of the neuron. When such information has been gathered from experiments, more realistic mathematical models for neural activity have been proposed (Hodgkin and Huxley, 1952).

The main difficulty with the study of first-passage times using the Ornstein-Uhlenbeck model is that the probability density function (pdf) of these times is not analytically tractable. In fact, the inverse Gaussian distribution (Chhikara and Folks, 1988; Jorgensen, 1981) and generalized inverse Gaussian distribution families (Iyengar and Liao, 1997) are the only ones that are amenable to standard statistical inference such as maximum likelihood or Bayes estimation and hypothesis testing. These families arise from simple Brownian motion models and more general diffusion models of neural activity that includes reversal potentials (Iyengar and Liao, 1997). Thus, in the case Ornstein-Uhlenbeck, it is hard to estimate parameters from available data, to judge the goodness of fit, or to compare various models.

Recently, Ricciardi and Sato, (Ricciardi and Sato, 1988), did a detailed study of this first-passage time density for the Ornstein-Uhlenbeck model and its moments, providing useful approximations for certain ranges of parameter values. Their work has made it easier to use numerical methods to evaluate the density, so that maximum likelihood estimation is now feasible. In the present paper, we outline an algorithm for computing maximum likelihood estimates for the three identifiable parameters, which are themselves functions of the five Ornstein-Uhlenbeck model parameters. In particular, we use the properties of the Laplace transform and known inversion algorithms to compute the first-passage time density and its parameter partial derivatives. We

assume that the firing threshold is constant, and that each time the neuron fires, it is instantly reset to its resting potential; that is, we ignore the refractory period so the spike train forms a renewal process.

In §2, we summarize the known properties of the Ornstein-Uhlenbeck process and we extract the estimable or identifiable parameters of the process. Then, we give an expression for the first-passage time density in terms of its Laplace transform and the identifiable parameters. In §3, the algorithm for generating maximum likelihood estimates is described in detail. This includes an overview of the maximum likelihood method for known densities §3.1 and techniques for calculating the densities by inverting the Laplace transform §3.2. This section also discusses methods for maximizing the efficiency §3.3 and robustness of the algorithm §3.4 and finishes with a comparison of the algorithm in the case where the exact pdf is known §3.5. §4 opens with a description of techniques for calculating confidence intervals and regions for the estimations. This is followed by numerical results for two and three-parameter algorithms. We conclude with a discussion of the main results in §5.

2. SDE, First-Passage Time

Let $\{X(t), t \geq 0, X(0) = X_0\}$ be a stationary, Gaussian, Markovian process that is continuous in probability. Any such process is referred to as an Ornstein-Uhlenbeck (OU) diffusion and it satisfies a linear stochastic differential equation (SDE) of the form (Arnold, 1974):

$$\begin{aligned} \mathbf{d}X(t) &= \left(\mu - \frac{X(t)}{\tau}\right)\mathbf{d}t + \sigma\mathbf{d}W(t) \\ X(0) &= X_0 \, . \end{aligned} \tag{1}$$

Here, $\{W(t), t \geq 0\}$ denotes Brownian motion with zero mean and unit variance and $\mathbf{d}W(t)$ represents white noise. OU diffusions have continuous sample paths. That is, for any $\epsilon > 0$,

$$\lim_{h\to 0} \mathrm{P}\left\{|X(t+h) - X(t)| > \epsilon\right\} = 0 \, ,$$

where P is the probability measure. The infinitesimal drift and variance, μ and σ^2, are defined through expectation operators:

$$\begin{aligned} \mu &= \lim_{h\to 0} \frac{1}{h} E\left\{X(t+h) - X(t)\right\} \\ \sigma^2 &= \lim_{h\to 0} \frac{1}{h} E\left\{[X(t+h) - X(t)]^2\right\} \, . \end{aligned}$$

More general diffusions can have both spatial and temporal dependence in the drift and variance, $\mu(X(t), t)$ and $\sigma^2(X(t), t)$ (Karlin and Taylor,

1981), although this study focuses on the case where μ and σ^2 are constant. In the context of a neuron, $X(t)$ represents the depolarization or membrane potential at time t. μ captures the mean excitatory ($\mu > 0$) or inhibitory ($\mu < 0$) input from the surrounding neural network while the net variance in that signal is stored within σ^2. The remaining parameter, τ, is a physiological parameter of the cell. It defines the rate at which $X(t)$ decays to its resting potential, X_0, in the absence of external stimuli ($\mu = 0, \sigma^2 = 0$). We can further simplify by assuming $X_0 = 0$ by translation. OU processes are commonly referred to as mean-reverting due to the presence of the attracting equilibrium, $X = \mu\tau$, in the deterministic limit of the SDE ($\sigma \to 0$). The solution to (1) is given in terms of $W(t)$:

$$X(t) = \mu\tau\left(1 - e^{-t/\tau}\right) + X_0 e^{-t/\tau} + \sigma\sqrt{\frac{\tau}{2}} e^{-t/\tau} W\left(e^{2t/\tau} - 1\right) . \quad (2)$$

Note that in the limit $\sigma \to 0$, (2) converges to the solution of the deterministic case.

The first-passage time of the OU process to a horizontal barrier, $X(t) = X_f$, is a random variable T:

$$\begin{aligned} T &= \inf\{t > 0 : X(t) = X_f\} \\ &= \inf\left\{t > 0 : \theta_2 = \theta_1 e^{-t/\theta_3} + \frac{e^{-t/\theta_3}}{\sqrt{2}} W(e^{2t/\theta_3} - 1)\right\} , \end{aligned} \quad (3)$$

where θ_i ($i = 1, 2, 3$) are the *identifiable* parameters:

$$\begin{aligned} \Theta &= (\theta_1, \theta_2, \theta_3)^T & (4) \\ &= \left(\frac{X_0 - \mu\tau}{\sigma\sqrt{\tau}}, \frac{X_f - \mu\tau}{\sigma\sqrt{\tau}}, \tau\right)^T . & (5) \end{aligned}$$

Given only a set of first-passage time data $\vec{T} = \{T_1, \cdots, T_n\}$ and a method for calculating the pdf of (3), estimates of Θ can be generated using the maximum likelihood method. It is not possible though, to estimate the full set of parameters $\{X_f, X_0, \mu, \sigma, \tau\}$ given only first-passage time data. This would require auxiliary experiments to fix two of the parameters to determine the full parameter set.

Tractable expressions for the pdf of T exist only in the special case $X_f = \mu\tau$ (Darling and Siegert, 1953):

$$\begin{aligned} p(t, \zeta, \theta_3) &= \frac{2\zeta e^{2t/\theta_3}}{\sqrt{\pi}\theta_3\left(e^{2t/\theta_3} - 1\right)^{3/2}} \exp\left[-\frac{\zeta^2}{e^{2t/\theta_3} - 1}\right] & (6) \\ \zeta &= \frac{X_f - X_0}{\sigma\sqrt{\tau}} \end{aligned}$$

For all other parameter values, exact representations for $p(t,\Theta)$ are more diffcult to generate. Typically one starts with the Laplace transform, $\hat{p}(\nu,\Theta) = \mathcal{L}(p(t,\Theta))$ which has a simple form (Darling and Siegert, 1953; Siegert, 1951):

$$\hat{p}(\nu,\Theta) = \frac{H_{\nu\theta_3}(\theta_1)}{H_{\nu\theta_3}(\theta_2)} \; . \tag{7}$$

Here, $H_\nu(\theta)$ are parabolic cylinder or Hermite functions that satisfy the ODE,

$$H''_\nu(\theta) - 2\theta H'_\nu(\theta) - 2\nu H_\nu(\theta) = 0 \; , \tag{8}$$

and the derivative are with respect to θ. Solutions of (8) are uniformly convergent power series for all θ (Lebedev, 1972):

$$H_\nu(\theta) = \sum_{n=0}^{\infty} \frac{\Gamma\left(\frac{n+\nu}{2}\right)(2\theta)^n}{\Gamma\left(\frac{\nu}{2}\right) n!} \; . \tag{9}$$

ν is the variable of the Laplace transform domain and is called the order parameter. When $\nu \in \mathbb{Z}^+$, the Hermite functions reduce to the Hermite polynomials. The real density, $p(t,\Theta)$, can then be obtained through an inverse Laplace transform, also known as the Bromwich integral (Churchill, 1981):

$$\begin{aligned} p(t,\Theta) &= \mathcal{L}^{-1}\left(\hat{p}(\nu,\Theta)\right) \\ &= \frac{1}{2\pi i}\int_{\sigma_0-i\infty}^{\sigma_0+i\infty} e^{t\nu}\hat{p}(\nu,\Theta)\mathbf{d}\nu \; . \end{aligned} \tag{10}$$

The contour integration in (10) extends ν to the complex plane ($\nu = \nu_{real} + i\nu_{imag}$). The integration path ($\nu_{real} = \sigma_0, -\infty < \nu_{imag} < \infty$) is chosen such that the abscissa of convergence, σ_0, is greater than the real part of all singularities of $\hat{p}(\nu,\Theta)$. Properties of the Hermite functions simplify this contour integration. For instance, $H_\nu(\theta)$ is an entire function of $\nu \in \mathbb{C}$ and does not vanish for $\nu_{real} \geq 0$. This implies $\sigma_0 \leq 0$. Moreover, $H_\nu(\theta) = 0$ at a countable set of points residing along the line $\nu_{real} < 0$, $\nu_{imag} = 0$. Therefore, $\hat{p}(\nu,\Theta)$ has a countable set of simple poles at the zeros of the Hermite function $H_{\nu\theta_3}(\theta_2)$. If one extends the contour to contain these poles, then residue theory can be applied to generate a formal expression for $p(t,\Theta)$ (see Eq.(28) in (Ricciardi and Sato, 1988)); however these representations are cumbersome.

Despite the existence of the formal representation for the real density, it is not used in the algorithm discussed in §3 and §4. The decision was motivated by computational efficiency and implementation considerations. The formal representation for $p(t,\Theta)$ is an infinite series whose

coefficients are also power series in the parameters $\theta_{1,2,3}$. The asymptotic behavior of these series is not completely understood, therefore robust and efficient calculation for large parameter ranges could prove difficult. More importantly, the MLE will be generated via Newtons method and requires all parameter derivatives out to second order. Generating and implementing these derivatives would be difficult for the formal representation. However, these calculations are comparatively simple in the Laplace transform domain.

3. Algorithm

We begin by describing the maximum likelihood method for parameter estimation. This method assumes that the pdf and its parameter partial derivatives can be readily computed. This is followed by a summary of the algorithms for generating a function when only its Laplace transform is available. Then, the focus turns to the computational efficiency of the ML method. Next, we discuss techniques for maximizing the algorithms effectiveness over the broadest possible parameter space. Finally, we give results of the algorithm for the specific case where the pdf is known.

3.1. Maximum Likelihood Estimation

The method of maximum likelihood (ML) provides estimates of parameters embedded within a distribution. The goal of ML is to maximize the likelihood function,

$$L(\Theta|\vec{T}) = \prod_{i=1}^{n} p(T_i, \Theta) \ , \tag{11}$$

with respect to the identifiable parameters Θ. Here, the switch in the order of the arguments, $\vec{T}$ and Θ, reflects the idea that the observed data $\vec{T}$ now acts as the parameters of the function. Maximizing (11) is equivalent to maximizing the log-likelihood function,

$$\ln L(\Theta|\vec{T}) = \sum_{i=1}^{n} \ln p(T_i, \Theta) \ , \tag{12}$$

with respect to Θ. Moreover, ML assumes that the pdf is available and can be evaluated at the data points $p(T_i, \Theta)$. With sufficient smoothness, the maximum likelihood estimate (MLE), $\hat{\Theta} = \left(\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3\right)^T$, is

then the solution to the nonlinear system of equations:

$$\begin{aligned} \frac{\partial}{\partial\theta_1}\ln L(\Theta|\vec{T}) &= 0 \\ \frac{\partial}{\partial\theta_2}\ln L(\Theta|\vec{T}) &= 0 \\ \frac{\partial}{\partial\theta_3}\ln L(\Theta|\vec{T}) &= 0\ . \end{aligned} \tag{13}$$

This system can be solved numerically with a multidimensional root-finding algorithm. If the derivatives of $p(T_i,\Theta)$ with respect to $\theta_{1,2,3}$ are available up to second order, then application of Newtons method gives an algorithm that is asymptotically efficient (Lehmann, 1983):

$$\begin{aligned} \Theta^{(n+1)} &= \Theta^{(n)} - H(\Theta^{(n)})F(\Theta^{(n)}) \\ \Theta^{(n)} &= \left(\theta_1^{(n)},\theta_2^{(n)},\theta_3^{(n)}\right)^T\ . \end{aligned}$$

Here, $H(\Theta^{(n)})$ is the Hessian matrix containing the second derivatives of $\ln L(\Theta|\vec{T})$, and $F(\Theta^{(n)})$ is the system (13). Then, the MLE is the limiting value of the sequence:

$$\hat{\Theta} = \lim_{n\to\infty}\Theta^{(n)} = \lim_{n\to\infty}\left(\theta_1^{(n)},\theta_2^{(n)},\theta_3^{(n)}\right)^T = \left(\hat{\theta}_1,\hat{\theta}_2,\hat{\theta}_3\right)^T\ .$$

Typically it is found that only a few iterations are needed for convergence. This is due to the fact that $\ln L(\Theta|\vec{T})$ is a log-concave function of Θ (Iyengar and Mullowney, 2006). Moreover, convergence of the algorithm requires an initial guess within the basin of attraction. Estimates for θ_1 and θ_2 can be generated through the method of moments (Inoue et al., 1995), although no such technique exists for θ_3 . For the typical neuron, θ_3 is on the order of $5-20$ milliseconds. This range can be used to formulate an appropriate initial guess, although a bit of guesswork is certainly required.

3.2. Inversion of the Laplace Transform

In this section, we consider techniques for computing the inverse Laplace transform of a function (10). There are a number of methods available with each having their own benefits and drawbacks. In Weeks' method (Weeks, 1966), the desired function is written as an expansion in Laguerre polynomials. This routine can be computationally efficient if one needs to perform multiple evaluations in the time domain. However, the implementation is not as straightforward as the other methods. Alternatively, the Post-Widder algorithm (Kano et al., 2005) has a simple

form for its expansion and is easy to use. However the convergence is slow and the method is computationally prohibitive when multiple time domain evaluations are needed. The Fourier Series technique of De Hoog(De Hoog et al., 1982; D'Amore et al., 1999) is a good choice because it is highly efficient for multiple time evaluations and straightforward to program. In this algorithm, the path of the contour integration of (10) is discretized as:

$$\nu_k = \sigma_0 + ik\Delta\nu \quad 0 \leq k \leq N \; . \tag{14}$$

The grid spacing parameter is given by $\Delta\nu = \pi/T_{max}$ where T_{max} is a parameter to be determined. Any numerical integration technique could be employed at this juncture; the trapezoidal rule gives a simple Fourier Series expansion:

$$\tilde{p}_N(t,\Theta) = \frac{e^{\sigma_0 t}}{T_{max}} \mathbf{Re}\left[\frac{\hat{p}(\sigma_0,\Theta)}{2} + \sum_{k=1}^{N} \hat{p}\left(\sigma_0 + \frac{ik\pi}{T_{max}},\Theta\right) e^{\frac{ik\pi}{T_{max}}}\right] \; . \tag{15}$$

The error in the approximation, $|\tilde{p}_N(t,\Theta) - p(t,\Theta)|$ converges to the trapezoidal discretization error as $N \to \infty$. This holds for $t \in [0, T_{max}]$ provided $p(t,\Theta)$ has period T_{max} . Otherwise, the method suffers from the Runge phenomenon at the boundaries of the interval. In the context of the problem at hand, $\sigma_0 = 0$ due to the properties of the Hermite functions. T_{max} is chosen to be greater than the maximum stopping time $\max\{T_1, \cdots, T_n\}$ while N is chosen sufficiently large to achieve a desired accuracy in the Fourier Series approximation. (15) also requires the evaluation of the Laplace transform, $\hat{p}(\nu,\Theta)$, at an evenly spaced grid on the complex ν-axis.

3.3. Efficiency

Given this technique for inverting Laplace transform, we now consider the requirements of the ML algorithm:

$$\begin{aligned}
\textbf{density} \; : \quad & p(t,\Theta) \\
\textbf{firstderivatives} \; : \quad & \partial_{\theta_1} p(t,\Theta), \partial_{\theta_2} p(t,\Theta), \partial_{\theta_3} p(t,\Theta) \\
\textbf{secondderivatives} \; : \quad & \partial^2_{\theta_1} p(t,\Theta), \partial^2_{\theta_2} p(t,\Theta), \partial^2_{\theta_3} p(t,\Theta), \\
& \partial^2_{\theta_1\theta_2} p(t,\Theta), \partial^2_{\theta_1\theta_3} p(t,\Theta), \partial^2_{\theta_2\theta_3} p(t,\Theta) \; .
\end{aligned}$$

Although the formal representation of $p(t,\Theta)$ exists and the required derivatives could be analytically computed, such expressions are impractical. Alternatively, one can work in the Laplace transform domain where the density and its derivatives have simple representations that can be computed efficiently and accurately for all parameter values.

Then, the Fourier series inversion method can be used to generate the real density and its parameter derivatives. For $p(t,\Theta)$, a careful application of (15) gives the real density. For the parameter derivatives more work is required. In (Iyengar and Mullowney, 2006), it has been shown that the order of the differentiation and contour integration can be switched for any parameter partial derivative of $p(t,\Theta)$, i.e.

$$\begin{aligned}\frac{\partial}{\partial\theta_i}p(t,\Theta) &= \frac{\partial}{\partial\theta_i}\left(\frac{1}{2\pi i}\int_{\sigma_0-i\infty}^{\sigma_0+i\infty}e^{t\nu}\hat{p}(\nu,\Theta)\mathbf{d}\nu\right)\\ &= \frac{1}{2\pi i}\int_{\sigma_0-i\infty}^{\sigma_0+i\infty}e^{t\nu}\frac{\partial}{\partial\theta_i}\hat{p}(\nu,\Theta)\mathbf{d}\nu\end{aligned}$$

The efficiency of the algorithm would be significantly hindered in the event that this switch were not possible. First, the parameter derivatives would need to be approximated using finite differences. Second order accuracy in all the parameter partial derivatives using the traditional $[1,-2,1]$ stencils or some variation therein would require 27 separate inversions of $\hat{p}(\nu,\Theta)$ for parameter values near $\Theta^{(n)}$. This would be necessary at each ML iteration. Moreover, it would introduce another source of error in the computations. Although one could use Broydens method (Broyden, 1965), this would require a second initial approximation to the parameters which is not readily available. For these reasons, the ML method is optimal provided the parameter partial derivatives of $\hat{p}(\nu,\Theta)$ can be generated in a time-efficient manner.

3.4. Algorithmic Parameter Range

In order for the ML algorithm to be robust, accurate representations of $\hat{p}(\nu,\Theta)$ and all its parameter partial derivatives are necessary for the widest possible parameter range: $-\infty<\theta_2<\theta_1<\infty, 0<\theta_3<\infty$. This is a nontrivial computational task whose details merits some explanation. Consider first, the computation of $\hat{p}(\nu,\Theta)$. Recall from (7) that $\hat{p}(\nu,\Theta)$ is defined as a ratio of Hermite functions that are uniformly convergent power series in their arguments, $\theta_{1,2}$, and that the order parameter always has the form $\nu\theta_3$. In addition, recall that the Fourier inversion algorithm (15) requires evaluations of $\hat{p}(\nu,\Theta)$ along the complex axis of the order parameter. When the argument of the Hermite function is small ($|\theta_{1,2}|<2$ roughly), the power series can be used to calculate H and thus $\hat{p}(\nu,\Theta)$ for all $\nu\theta_3$ values on the complex axis with approximately 100 terms. However, for larger $|\theta_{1,2}|$, the power series suffers from catastrophic cancellation as $\nu\theta_3\to\infty$ (on the complex axis). Catastrophic cancellation occurs when the sum of an alternating series results in cancellation across orders of magnitude greater than

the accuracy of the chosen floating point representation. The onset of the phenomenon is observed for smaller $\nu\theta_3$ with increasing $|\theta_{1,2}|$ until eventually, the power series gives unreliable results.

The observations above indicate that limiting representations of the Hermite functions are needed for $|\theta_{1,2}| \to \infty$ and $|\nu\theta_3| \to \infty$. Using the known relationships between the Hermite and parabolic cylinder functions (Lebedev, 1972; Abramowitz and Stegun, 1972), the following asymptotic representation can be derived for the case $\xi = \sqrt{x^2 + 4\nu\theta_3 - 2} \to \infty$:

$$\begin{aligned} H_{\nu\theta_3}(x) &= 2^{3/4}\sqrt{\frac{\Gamma\left(\frac{\nu\theta_3+1}{2}\right)}{\Gamma\left(\frac{\nu\theta_3}{2}\right)}}\exp\left[\frac{x^2}{2} \pm \vartheta + G\left(\nu\theta_3 - \frac{1}{2}, x\sqrt{2}\right)\right] \\ \vartheta &= \frac{x\xi}{4} + \left(\nu\theta_3 - \frac{1}{2}\right)\ln\left(\frac{x+\xi}{2\sqrt{\nu\theta_3 - \frac{1}{2}}}\right) \qquad (16) \\ G\left(\nu\theta_3 - \frac{1}{2}, x\sqrt{2}\right) &= -\frac{\ln\xi}{2} + \frac{g_3}{\xi^3} + \frac{g_6}{\xi^6} + \frac{g_9}{\xi^9} + \frac{g_{12}}{\xi^{12}} + O\left(\frac{1}{|\xi|^{15}}\right) \end{aligned}$$

The exact form of the coefficients $g_3, g_6, \cdots$ is given by (19.10.13) in (Abramowitz and Stegun, 1972); note that each is a function of x and $\nu\theta_3$.

Now, consider a situation where the power series representation of H gives reliable results for small $\nu\theta_3$ before succumbing to catastrophic cancellation. (16) is particularly useful in this scenario because it can accurately compute H well before the onset of the numerical instability. Thereafter, the approximation improves for larger $\nu\theta_3$ thus giving a precise representation of H over the entire integration range. Next, consider the situation where the power series never gives reliable results ($|x| \to \infty$). (16) gives reasonable approximations of H when $|\nu\theta_3| << |x|$. In this regime though, another expansion is far more accurate (Lebedev, 1972):

$$\begin{aligned} H_{\nu\theta_3}(x) &= \frac{2}{\Gamma\left(\frac{\nu\theta_3}{2}\right)}\left\{(-2x)^{-\nu\theta_3}\left[\sum_{k=0}^{n}\frac{(-1)^k\Gamma(\nu\theta_3+2k)}{k!(2x)^{2k}} + O\left(|x|^{-2n-2}\right)\right]\right. \\ &+ \left. h(x)\sqrt{\pi}e^{x^2}x^{\nu\theta_3-1}\left[\sum_{k=0}^{n}\frac{\Gamma(1-\nu\theta_3+2k)}{\Gamma(1-\nu\theta_3)k!(2x)^{2k}} + O\left(|x|^{-2n-2}\right)\right]\right\}, \end{aligned} \qquad (17)$$

where $h(x)$ is the Heaviside function. For larger $\nu\theta_3$, (17) loses validity while the (16) approximation improves. Thus, one can generate an

accurate computation of H over the entire integration range for any parameter values by piecing together the different representations.

Fig. 1 shows the relative error between the three representations of $H_{\nu\theta_3}(x)$ for $\nu\theta_3$ along the complex axis. In this case, x is moderately large ($x = -4.47$) and the power series suffers from catastrophic cancellation giving at least $O(1)$ errors for $|\nu\theta_3| > 3$. The dotted trajectory shows that (17) gives an excellent approximation of (9) as $|\nu\theta_3| \to 0$. As expected though, the approximation diminishes as $|\nu\theta_3|$ increases. On the other hand, (16) gives a poor approximation for small $|\nu\theta_3|$ but improves significantly for larger values. Eventually though the approximation degrades due to the onset of the cancellation in (9). For $|x| < 2$, this error would go to 0 because the power series does not suffer from cancellation in this regime. Finally, the dashed line gives the error between the two asymptotic approximations. While (16) should improve with increasing $|\nu\theta_3|$ for any x, (17) loses validity when $|x| \sim |\nu\theta_3|$. This picture is typical for any x value, the only difference between the location where the different representations gain/lose validity.

We now address the computation of the parameter derivatives of H and thus $\hat{p}(\nu, \Theta)$. In the power series representation (9), uniform convergence implies that the infinite series can be differentiated term by term to give a series that is also uniformly convergent:

$$\frac{\partial}{\partial x} H_{\nu\theta_3}(x) = 2\nu\theta_3 H_{\nu\theta_3+1}(x) \; .$$

The same holds for the second derivatives with respect to x although for computation purposes, it would be easier to use (8). For the θ_3 derivatives, direct differentiation of (9) also yields convergent power series. However, this brings about an interesting side issue that must be addressed; the θ_3 dependence in θ_1 and θ_2. When computing θ_3 partial derivatives, this dependence is ignored effectively treating $\theta_{1,2,3}$ as independent parameters. Were this not done, then the $\partial_{\theta_3}\theta_1(\theta_3)$ and $\partial_{\theta_3}\theta_2(\theta_3)$ derivatives would yield terms that could not be written only as a function of $\theta_{1,2,3}$. The ML algorithm would then require information on two out of four parameters: X_0, X_f, μ, or σ. This cannot be generated through first-passage time data since (3) is a function of $\theta_{1,2,3}$ only. Thus, the parameters are considered to be independent.

Parameter derivatives of all types can be computed from the asymptotic expansion (16). Here, convergence is guaranteed because it is generated as a solution to a differential equation (Miller, 1955; Bender and Orzag, 1978). This is not the case when differentiating (17) with respect to the parameters because it was derived as an approximation to an integral (Lebedev, 1972). In practice though, it has never been shown to fail.

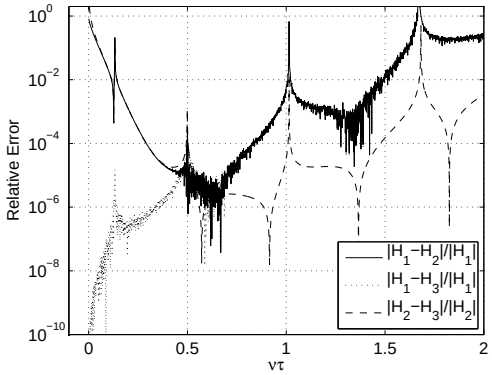

Figure 1. Relative Error versus $\nu\theta_3$ for the three Hermite function representations for $\theta_3 = 5$, $\theta = -4.47$. The solid line gives the error between the power series (9) and the Darwin expansion (16), the dotted line compares (9) and (17), and the dashed line measures the error between (16) and (17).

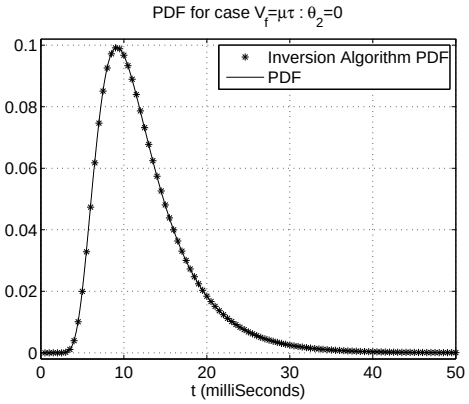

Figure 2. Probability distribution function (pdf) versus t for $\{X_0, X_f, \mu, \sigma, \tau\} = \{0, 15, 3, 1.5, 5\}$. (-) gives the exact representation, (6) while the markers (*) give the pdf through (15).

3.5. The case $\theta_2 = 0$

Recall the special case, $\theta_2 = 0$ or $X_f = \mu\theta_3$, where the exact pdf is given by (6). Fig. 2 displays (6) and the numerically inverted pdf from (15). To the eye, (15) accurately computes the exact distribution. To quantify the accuracy of the inversion, the relative error is computed between (6) and (15) on a uniform grid in t. The results are then averaged across the grid giving a mean relative error. In this particular example, the mean relative error between the two representations is $1.5 * 10^{-3}$. Since, the grid spacing parameter for the inversion is $\Delta\nu = \pi/50$, we find agreement with the trapezoidal rule composite error $O(\Delta\nu^2)$. As

expected, refinement of the grid spacing parameter $\Delta\nu$ yields a smaller mean relative error always in agreement with the theoretical composite error.

4. Numerical Results

In this section, numerical results for the algorithm are given. First, we consider the case of estimating the two parameters, (θ_1, θ_2), when θ_3 is fixed. In particular, we focus on the construction and testing of approximate confidence intervals and regions for the estimates. Then, results for the full three-parameter estimation $\theta_{1,2,3}$ are given. Comparisons with the two-parameter case will be discussed. Here, we note that simulated data from (2) is used to generate the first-passage time samples for all parameter estimation studies. The data sets are generated via discrete mapping in time:

$$\begin{aligned} t_i &= i\Delta t \\ W_i &= \sqrt{\frac{2\Delta t}{\theta_3}} e^{t_i/\theta_3} Z_i + W_{i-1} \\ X_i &= (X_0 - \mu\theta_3)e^{-t_i/\theta_3} + \mu\theta_3 + \sigma\sqrt{\frac{\theta_3}{2}} e^{-t_i/\theta_3} W_i \ . \end{aligned} \tag{18}$$

Here, Z_i are standard, independent Gaussian variates and W_i is the increment of the Wiener process with $W_0 = 0$. Since the probability that $X_n = X_f$ at some $t_n = n\Delta t$ is zero, linear interpolation is used to find the time of the boundary crossing once X_n first exceeds X_f. It is important to emphasize that the reliability of the sample in representing a true OU diffusion is inversely proportional to Δt. This fact will have an important effect on both the quality of the estimates as well as the confidence intervals and regions as we shall see in the two-parameter estimation case study.

In order to construct approximate confidence intervals and regions for the estimates, we make use of the Hessian evaluated at the MLE, $H(\hat{\Theta})$, or the observed Fisher information matrix:

$$J(\hat{\Theta}) = -H(\hat{\Theta}) \ .$$

It has been shown in (Iyengar and Mullowney, 2006) that the MLEs of the embedded parameters Θ are asymptotically normal as the number of first-passage time samples, n, tends to ∞, i.e.:

$$\sqrt{J(\hat{\Theta}_n)}(\hat{\Theta}_n - \Theta) \to N(0, I) \tag{19}$$

$$J(\hat{\Theta}_n)^{-1} = \begin{bmatrix} \mathbf{Var}(\hat{\theta}_{1,n}) & \mathbf{Cov}(\hat{\theta}_{1,n},\hat{\theta}_{2,n}) & \mathbf{Cov}(\hat{\theta}_{1,n},\hat{\theta}_{3,n}) \\ \mathbf{Cov}(\hat{\theta}_{2,n},\hat{\theta}_{1,n}) & \mathbf{Var}(\hat{\theta}_{2,n}) & \mathbf{Cov}(\hat{\theta}_{2,n},\hat{\theta}_{3,n}) \\ \mathbf{Cov}(\hat{\theta}_{3,n},\hat{\theta}_{1,n}) & \mathbf{Cov}(\hat{\theta}_{3,n},\hat{\theta}_{2,n}) & \mathbf{Var}(\hat{\theta}_{3,n}) \end{bmatrix} . \quad (20)$$

Here, we adopt the notation $\hat{\Theta}_n = \left(\hat{\theta}_{1,n}, \hat{\theta}_{2,n}, \hat{\theta}_{3,n}\right)^T$ to specify the sample size dependence of the estimate thereby distinguishing it from the iteration of the ML algorithm given by the superscript $\hat{\Theta}^{(n)}$. Typically J is referred to as the covariance or error matrix. I is a vector of ones whose dimension equals the number of parameters being estimated. Using (19) and (20), one can construct individual confidence intervals for each parameter in the usual manner via:

$$\hat{\theta}_{1,n} \pm z_{\frac{1-\alpha}{2}} \sqrt{J(\hat{\Theta}_n)^{-1}_{11}} \quad (21)$$

$$\hat{\theta}_{2,n} \pm z_{\frac{1-\alpha}{2}} \sqrt{J(\hat{\Theta}_n)^{-1}_{22}} \quad (22)$$

$$\hat{\theta}_{3,n} \pm z_{\frac{1-\alpha}{2}} \sqrt{J(\hat{\Theta}_n)^{-1}_{33}} \quad (23)$$

where $z_{\frac{1-\alpha}{2}} = 1.96$ when $\alpha = .95$[1]. The intersection of these regions in the $\theta_1\theta_2\theta_3$ space defines an α percent confidence box. For (21)-(23), note that the square root operation is taken after computing $J(\hat{\Theta}_n)^1_{ii}$

Alternatively, one can construct more restrictive confidence ellipsoids. To do this, we assume that (12) can be modeled with an ellipsoid around the MLE:

$$\begin{aligned} F(\theta_1,\theta_2,\theta_3) &= a_1\theta_1^2 + a_2\theta_2^2 + a_3\theta_3^2 + 2a_4\theta_1\theta_2 + 2a_5\theta_1\theta_3 \\ &+ 2a_6\theta_2\theta_3 + 2a_7\theta_1 + 2a_8\theta_2 + 2a_9\theta_3 + a_{10}. \end{aligned} \quad (24)$$

Then, use the following facts to determine the parameters, a_i :

- (12) is maximized at $\hat{\Theta}_n$ and thus F is as well. This gives 3 constraints.
- The second derivatives of (12), i.e. the Hessian, must coincide with the second derivatives of F. This gives 6 constraints since the Hessian is symmetric.
- Equating the two functions at the MLE gives 1 constraint and determines the center of the ellipsoid:

$$\ln L(\hat{\Theta}_n|\vec{T}) = F(\hat{\theta}_{1,n}, \hat{\theta}_{2,n}, \hat{\theta}_{3,n}) .$$

[1] The decimal and percentage interpretations for α are used interchangeably hereafter

Given this, we must now determine which contour of F defines the $\alpha\%$ confidence region. To do this, we first compute the square root of the covariance matrix. This is done via eigenvalue decomposition since the covariance matrix is always positive definite. Let $\{\lambda_{max}, v_{max}\}$ be the eigenvalue-eigenvector pair corresponding to the maximum eigenvalue. v_{max} defines the direction of the ellipsoids semi-major axis. The other eigenvectors define the minor axes. Then, the contour of F, $z_{\frac{1-\alpha}{4}}\sqrt{\lambda_{max}}$ units from the center of the region in the direction of v_{vmax} yields the approximate $\alpha\%$ confidence ellipsoid. This contour could equally be determined using the other eigenvalue-eigenvector pairs. Notice here that $z_{\frac{1-\alpha}{4}}$ is used rather than $z_{\frac{1-\alpha}{2}}$[2]. This arises from the fact that confidence ellipsoids measure the joint probability of the true parameters residing within the region. Thus, a larger z value must be used to accomodate the combined variation. For uncorrelated parameters, $z_{\frac{1-\alpha}{4}}$ is sufficient for a reasonable approximation. For highly correlated parameters, $z > z_{\frac{1-\alpha}{4}}$ is necessary.

4.1. 2-D RESULTS: FIXED θ_3

The results in §3.5 instill confidence in the accuracy and reliability of the inversion algorithm. We now proceed with an analysis of the MLEs in the case where θ_3 is fixed. Fig. 3a shows a histogram of a simulated first-passage time sample for $\theta_1 \approx -3.354$ and $\theta_2 \approx 1.118$. Fig. 3b displays the true pdf evaluated on a grid of points in t with the exact parameters used to generate the sample in Fig. 3a. It also shows the maximum likelihood pdf evaluated at the first-passage sample shown in Fig. 3a using the MLE parameters: $\hat{\theta}_{1,n} \approx -3.231$, $\hat{\theta}_{2,n} \approx 1.134$. Each density is calculated from the inversion technique (15). To the eye, the maximum likelihood pdf does a good job of approximating the true distribution although deviations are certainly evident especially near the maximum.

Contours of the log-likelihood function (12) are shown in Fig. 4 in the $\theta_1\theta_2$ plane for two cases of θ_2. The true values of the parameters, (θ_1, θ_2), and the estimate $(\hat{\theta}_{1,n}, \hat{\theta}_{2,n})$ are also given for reference. The 95% and 99% confidence regions/ellipses are plotted as thick, dark contours. The 95% and 99% confidence intervals/boxes are also shown to contrast the ellipses. At first we glance, we note that the contours are approximately elliptical and that the MLE resides at the center of the concentric ellipses. Note the different scales of the axes in the two pictures. The ratio of the ellipses major to minor axes is roughly $10:1$ in Fig. 4a and $4:1$ in Fig. 4b. These pictures are representative of

[2] For $\alpha = .95$, $z_{\frac{1-\alpha}{4}} = 2.243$

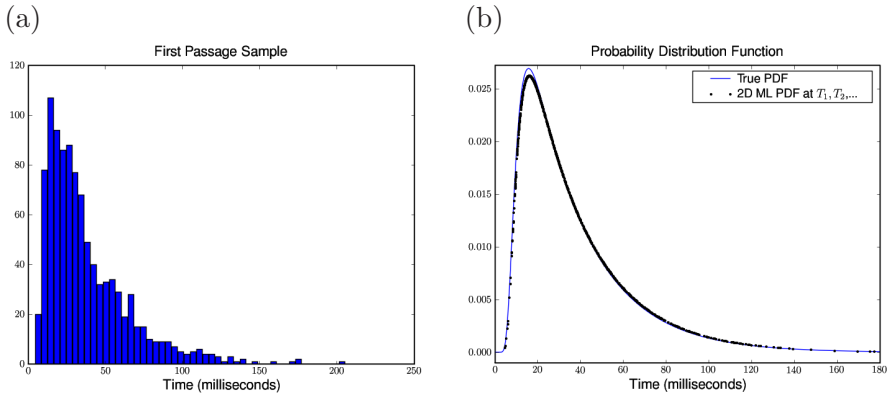

Figure 3. $\{X_0, X_f, \mu, \sigma, \tau\} = \{0, 20, 3, 2, 5\} \Leftrightarrow \Theta = (-3.354, 1.118, 5)^T$. Simulation parameters: $n = 1000$ samples with time step $\Delta t = 10^{-4}$. (a) distribution of first–passage times and (b) Maximum likelihood pdf using $\hat{\theta}_{1,n} \approx -3.231, \hat{\theta}_{2,n} \approx 1.134$ (.) versus true pdf (-). Each distribution was generated from (15).

the log-likelihood (12) when $\theta_2 > 0$ (Fig. 4a) and $\theta_2 < 0$ (Fig. 4b). In Fig. 4a, $\ln L(\Theta|\vec{T})$ is broad in the θ_1 direction and sharp in the θ_2 direction. These features are a byproduct of the mean reverting nature of the OU process. As $\theta_1 \to -\infty$, the process has a stronger inclination to revert back to $\mu\theta_3$ as indicated by (1). Therefore, large changes in θ_1 in either direction should have little effect on the first-passage time distribution. On the other hand, large positive increases in θ_2 yields significantly longer first-passage times since the natural tendency of the process is to revert back to $\mu\theta_3$ (recall that $\theta_2 > 0$ gives a threshold greater than $\mu\theta_3$). Then given enough samples ($n = 1000$ for both cases), it is not surprising that one has enough information to make a precise estimate of the θ_2 as indicated by the length of the semi-minor axis of the 95% confidence ellipsoid or the size of the 95% confidence interval in θ_2. If one chooses $X_f < \mu\theta_3$ as in Fig. 4b, the aspect ratio decreases dramatically as noted above, however, the correlation between θ_1 and θ_2 increases as measured by the rotation angle of the ellipse. The correlation increases as $\theta_2 \to \theta_1$ and the ellipses tend to circles. Due to this higher correlation, it is likely that the calculated 95% confidence ellipse in Fig. 4b underestimates the true region [3].

Now we consider the quality of the estimations with regard to the confidence intervals and regions. In particular, we expect asymptotic normality in the parameters for large sample sizes and we expect an α% confidence interval/region to contain the true parameter in α% of a set

[3] The first-passage sample in Fig. 3 gives the MLEs and contours in Fig. 4a. A different first-passage sample was used in Fig. 4b.

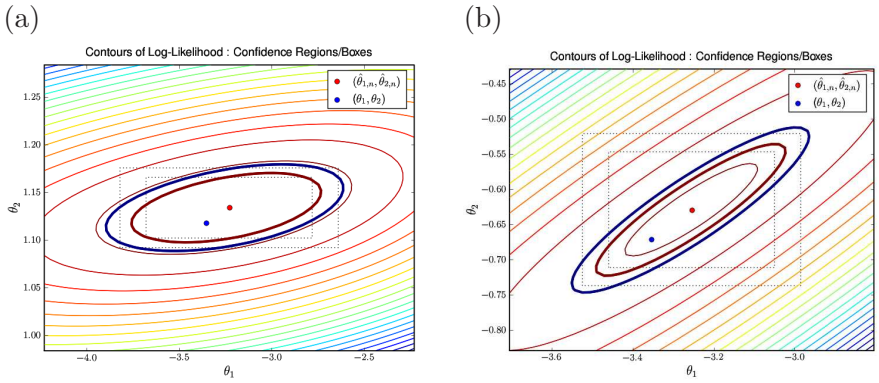

Figure 4. $\{X_0, \mu, \sigma, \tau\} = \{0, 3, 2, 5\} \Rightarrow \theta_1 = -3.354$. Simulation parameters: $n = 1000$ samples with time step $\Delta t = 10^{-4}$. Contours of log-likelihood function (12) in the $\theta_1\theta_2$ plane. The thick contours give the 95% and 99% confidence ellipses. (a) $X_f = 20 \Leftrightarrow \theta_2 = 1.118$ and (b) $X_f = 12 \Leftrightarrow \theta_2 = -.67082$.

of trials. To test this, 100 simulations are run with each consisting of 1000 first-passage times. Asymptotic normality is measured by looking at the distribution of (19) using the Anderson-Darling test at the 95% level (Stephens, 1974). The confidence regions/intervals can be tested by counting the number of times the true parameter(s) fall within the appropriate regions for all simulations. These tests are performed as a function of the simulation time increment Δt. The results are shown in Table I Clearly, refining the simulations with smaller times steps yields statistics consistent with the theoretical predictions when considering the asymptotic normality of the estimates, and the predictive power of the confidence intervals and regions/ellipses. More importantly, the data suggests that simulated data is only reliable when small time steps are taken. It should also be mentioned that the results in Table I are consistent for different values of θ_1 and θ_2 in all aspects

4.2. 3-D RESULTS

In the previous section, estimates for θ_1 and θ_2 were computed from first-passage time samples for fixed θ_3. This may incorrectly constrain the optimization since (3) is clearly a function of three identifiable parameters: $\theta_{1,2,3}$. Thus, we now allow for variation in θ_3 and see what effect this has on the estimates. First, the maximum likelihood pdf is computed from the sample shown in Fig. 3a. The algorithm yields the estimates: $\hat{\Theta}_n = (-2.653, 1.035, 5.984)^T$. The associated pdf is shown in Fig. 5. To the eye there is little discernible difference with the two-parameter estimation given in Fig. 3b. Although, once again, there are deviations from the true pdf around the maximum. The degree

Table I. $\{X_0, X_f, \mu, \sigma, \tau\} = \{0, 20, 3, 2, 5\} \Leftrightarrow \Theta = (-3.354, 1.118, 5)^T$. Number of times θ_j falls within $\alpha\%$ confidence interval, the number of times (θ_1, θ_2) falls within $\alpha\%$ confidence region/ellipse, and the asymptotic normality of (19) for 100 Samples, 1000 first-passage times per sample generated with time step Δt.

	θ_1				θ_2				(θ_1, θ_2)		
	Interval			Normal	Interval			Normal	Ellipse		
Δt	90	95	99	95	90	95	99	95	90	95	99
10^{-1}	66	75	92	No	6	8	25	No	6	12	31
10^{-2}	86	95	99	No	74	83	92	No	73	82	91
10^{-3}	90	98	100	Yes	84	94	98	No	83	90	99
10^{-4}	94	97	100	Yes	87	95	100	Yes	89	94	100

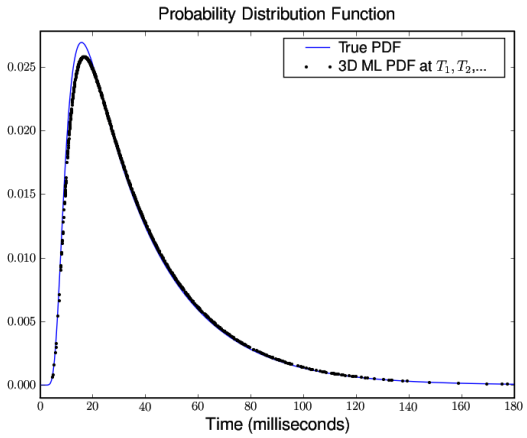

Figure 5. $\{X_0, X_f, \mu, \sigma, \tau\} = \{0, 20, 3, 2, 5\} \Leftrightarrow \Theta = (-3.354, 1.118, 5)^T$. Simulation parameters: $n = 1000$ samples with time step $\Delta t = 10^{-4}$. Maximum likelihood pdf (.) using $\hat{\Theta}_n = (-2.653, 1.035, 5.984)^T$ versus true pdf (-). Each distribution is generated from (15).

of similarity can be quantified by computing the r.m.s error between the true pdf evaluated at the first-passage times and the maximum likelihood pdf for both the two and three parameter estimations, i.e.:

$$\epsilon_j = \sqrt{\sum_{i=1}^{n} [p(T_i, \Theta) - p_j(T_i, \hat{\Theta}_n)]^2}$$

The subscript j denotes the two and three parameter estimates respectively and $p(T_i, \Theta)$ is the true pdf with parameters $\Theta = (-3.354, 1.118, 5)^T$.

For the data given in Fig. 3 and Fig. 5, we find that $\epsilon_2 \approx 3.7 * 10^{-4}$ and $\epsilon_3 \approx 6.7 * 10^{-4}$. Both errors are of the same order as the trapezoidal rule integration error $O(10^{-4})(\Delta\nu = .015)$ in the inversions.

The previous computations measured the relative accuracy of the maximum likelihood pdf for the two and three parameter algorithms. Now, we consider their respective degrees of variability, i.e. the confidence intervals for the estimates. To do this, we compute the 95% confidence intervals from (21)-(23) as a function of the sample size n. 20 unique simulations are performed for each n. For each sample size, we tabulate the number of times, N, that the true parameter value resides within the 95% interval as well as the average width (resolution) of the 95% confidence interval:

$$\langle \Delta CI \rangle_{i,n} = \frac{2z_{\frac{1-\alpha}{2}}}{20} \sum_{j=1}^{20} \sqrt{J(\hat{\Theta}_{n_j})^{-1}_{ii}} \; . \tag{25}$$

The results are given in Table II for both the two and three-parameter algorithms. From this data, it is obvious that the two-parameter algorithm provides a much higher resolution of the true parameter values for each sample size. This is because $\langle \Delta CI \rangle_{1,n}$ and $\langle \Delta CI \rangle_{2,n}$ are $4-5$ and $7-9$ times larger than their respective two-parameter algorithm counterparts. Even at 10000 samples, $\langle \Delta CI \rangle_{1,n} \approx 30\%$ larger than the same interval using the two-parameter algorithm and 1000 samples. This suggests that a much larger range of parameters is capable of representing the true distribution with reasonable accuracy. Thus, if θ_3 is not known a priori, then a significantly larger sample size is required to resolve the parameters with the same accuracy as the two-parameter algorithm.

In Fig. 6, $\langle \Delta CI \rangle_{i,n}$ is plotted against n for each parameter and each algorithm on a loglog scale. Each data set is fit with a function of form $\langle \Delta CI \rangle_{i,n} = \beta n^{\gamma}$. The results show that $\langle \Delta CI \rangle_{i,n} \sim n^{-1/2}$ as expected. Extrapolating these trends to larger n, we find that using the three-parameter algorithm with roughly 16000 samples gives the same resolution as the two-parameter algorithm with 1000 samples on $\hat{\theta}_{1,n}$. For $\hat{\theta}_{2,n}$, nearly 50000 samples are required.

A close inspection of the data yielding the statistics in Table II also reveals a strong correlation between $\hat{\theta}_{2,n}$ and $\hat{\theta}_{3,n}$. Fig. 7 gives a scatter plot of $\hat{\theta}_{3,n}$ versus $\hat{\theta}_{2,n}$ for every simulation in every sample size n. A marker denoting the true parameter value used for the simulation is given for reference. These results are not at all surprising considering the physics of the OU process, although the strength of the correlation is surprising. To see this, imagine two distinct OU processes labeled A and B that are subject to the same input drift μ and variance σ.

Table II. $\{X_0, X_f, \mu, \sigma, \tau\} = \{0, 20, 3, 2, 5\} \Leftrightarrow \Theta = (-3.354, 1.118, 5)^T$. Simulation Parameters; $\Delta t = 10^{-4}$, 20 simulations with n samples in each simulation. Number of times, N, that the true parameter falls within the 95% confidence interval for $\hat{\theta}_{1,n}$, $\hat{\theta}_{2,n}$, and $\hat{\theta}_{3,n}$ for the two and three parameter algorithms. Also, the average width of the 95% confidence interval, $\langle \Delta CI \rangle_{i,n}$ versus n.

	$\hat{\theta}_{1,n}$				$\hat{\theta}_{2,n}$				$\hat{\theta}_{3,n}$	
	3 Parameter		2 Parameter		3 Parameter		2 Parameter		3 Parameter	
n	N	$\langle \Delta CI \rangle_{1,n}$	N	$\langle \Delta CI \rangle_{1,n}$	N	$\langle \Delta CI \rangle_{2,n}$	N	$\langle \Delta CI \rangle_{2,n}$	N	$\langle \Delta CI \rangle_{3,n}$
1000	18	3.972	19	0.936	20	0.524	19	0.064	19	5.292
2000	19	2.588	19	0.646	20	0.370	17	0.045	20	3.612
3000	18	2.395	19	0.534	20	0.273	20	0.037	20	2.508
4000	20	2.029	19	0.466	20	0.237	19	0.032	20	2.180
5000	19	1.679	18	0.406	19	0.218	20	0.029	19	2.045
6000	20	1.536	18	0.374	20	0.198	18	0.026	20	1.846
7000	19	1.458	20	0.346	19	0.187	18	0.024	19	1.747
8000	20	1.350	17	0.324	20	0.169	20	0.023	19	1.561
9000	20	1.323	20	0.307	19	0.159	18	0.021	19	1.449
10000	19	1.230	18	0.290	19	0.153	20	0.020	19	1.407

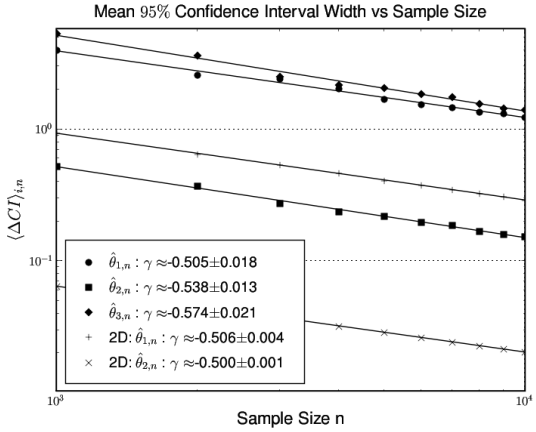

Figure 6. $\{X_0, X_f, \mu, \sigma, \tau\} = \{0, 20, 3, 2, 5\} \Leftrightarrow \Theta = (-3.354, 1.118, 5)^T$. Simulation parameters: with time step $\Delta t = 10^{-4}$. Average width of the 95% confidence interval, $\langle \Delta CI \rangle_{i,n}$, versus the number of samples for each parameter and each algorithm.

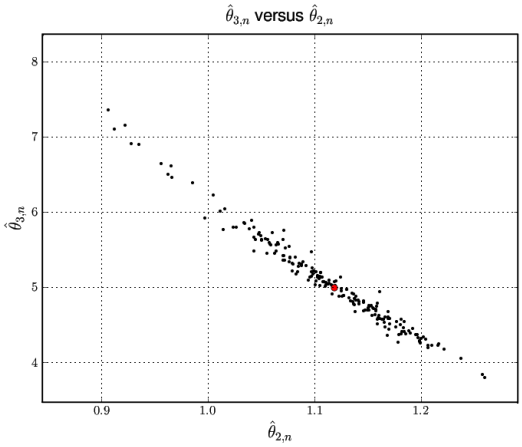

Figure 7. $\{X_0, X_f, \mu, \sigma, \tau\} = \{0, 20, 3, 2, 5\} \Leftrightarrow \Theta = (-3.354, 1.118, 5)^T$. Simulation parameters: $\Delta t = 10^{-4}$. $\hat{\theta}_{3,n}$ versus $\hat{\theta}_{2,n}$ for all sample sizes and simulations in Table II. The true value of Θ (i.e. the simulation parameters) is given by the red dot.

Moreover, let them have the same resting and firing thresholds X_0 and X_f. Lastly, let A have a larger time constant, θ_3. Since A has a larger time constant, the state variable of A, $X_A(t)$, will decay more slowly than $X_B(t)$ and thus can be expected to have shorter first-passage times. Shorter first passage-times implies a smaller value of θ_2 due to mean-reversion. The same argument can be applied in reverse thus explaining the strong negative correlation between these parameters.

Although a variety of additional parameter studies could be performed at this juncture, we believe that the main differences between the two and three-parameter algorithms have been demonstrated. Moreover, since the results are typical for any set of parameters, Θ, the conclusions are completely general.

5. Discussion

We have given an algorithm for computing maximum likelihood estimates (MLEs) for the three identifiable parameters of the Ornstein-Uhlenbeck process. Previous efforts (Ricciardi and Sato, 1988; Inoue et al., 1995; Ditlevsen and Ditlevsen, 2006) have focused on the two-parameter case where the time constant τ is known a priori. When this assumption is introduced into our algorithm, the resulting parameter estimates and their corresponding confidence intervals and regions are consistent with theoretical predictions.

However, when only first-passage time data is available, a priori assumptions on τ cannot be made and thus the full three-parameter estimation is necessary. Our results show unequivocably that the addition of the τ parameter in the estimation procedure dramatically decreases the resolution of $\theta_{1,2}$ for a fixed sample size. Here, a decrease in resolution corresponds to an increase in the width of the parameters' 95% confidence intervals. In fact, case studies have shown that roughly 15 times more samples are required to resolve θ_1 to the same accuracy as the two-parameter algorithm. For θ_2 , the situation is considerably worse since roughly 50 times more samples are required for comparable resolution between the algorithms. However, since the typical neuron yields between $500 - 50000$ samples in the spike train (Ricciardi and Sato, 1988), this algorithm could resolve the identifiable parameters with good accuracy. Moreover, the algorithm enables estimates of τ which could be corroborated through independent experiments.

The computational technique described here is based on the fact that the OU diffusion is temporally homogenous. This allowed us to write the Laplace transform of the first-passage time pdf as the Laplace transform of the transition density (Darling and Siegert, 1953). With this in mind, we expect the algorithm to be applicable (after appropriate changes) to other integrate-and-fire neuronal models with this feature; the Feller process is a good example. The availability of robust and efficient algorithms that relate the measurable data to the biophysical parameters would be an invaluable tool for any neurobiologist. In particular, it would allow for rapid model evaluation and comparison based on rigorous estimates of the parameters and their standard errors. Moreover, changes in experimental conditions that impact the neuron's behavior could be evaluated with algorithms of this type enabling a more precise understanding of the biophysics.

Our results also indicate that simulated first-passage time data can only be trusted when the time step of the simulation is small. This suggests the need for greater computational power when simulating and testing OU processes. This point is further reinforced by the fact that the data in Table II took roughly 30 hours to generate on a dual-core multiprocessing cpu. Moreover, it is reasonable to conclude that this result would hold for any type of stochastic integrate and fire model with comparable complexity like the CIR or Feller process thus suggesting the need for greater computational power when simulating stochastic neural behavior. A natural extension of this work to coupled, stochastic integrate and fire models would impose an even greater computational burden thus demanding parallel implementations for the purpose of simulation. This demand will be unavoidable as the size of the coupled network grows, and complexity of the cellular models increases.

References

M. Abramowitz and I.R. Stegun. Handbook of Mathematical Functions. Dover Publications Inc, 9 edition, 1972.

L. Arnold. Stochastic Differential Equations: Theory and Applications. John Wiley and Sons, 1974.

C.M. Bender and S.A. Orzag. Advanced mathematical methods for scientists and Engineers. McGraw Hill, 1978.

C. G. Broyden. A class of methods for solving nonlinear simultaneous equations. Mathematical Computing, 19:577–593, 1965.

S. Chandrasekhar. Stochastic problems in physics and astronomy (1943). In Selected papers in Noise and Stochastic Processes. Dover Publications, n. wax edition, 1954.

R. Chhikara and J.L. Folks. The Inverse Gaussian Distribution: Theory, Methodology, and Applications. Marcel-Dekker, 1988.

R.V. Churchill. Operational Mathematics. McGraw-Hill, 1981.

L. D'Amore, G. Laccetti, and A. Murli. An implementation of a fourier series method for the numerical inversion of the laplace transform. ACM Transactions on Mathematical Software, 25:279–305, 1999.

D. Darling and A. Siegert. The first passage problem for a continuous markov process. Annals of Mathematical Statistics, 24:624–639, 1953.

F.R. De Hoog, J.H. Knight, and A.N. Stokes. An improved method for numerical inversion of laplace transforms. SIAM Journal of Scientific and Statistical Computing, 3:357–366, 1982.

S. Ditlevsen and O. Ditlevsen. Parameter estimation from observations of first-passage times of the ornstein-uhlenbeck process and the feller process. Presented at the Fifth Computational Stochastic Mechanics Conference, 2006.

I.I. Gringorten. Estimating finite time maxima and minima of a stationary gaussian ornstein-uhlenbeck process by monte carlo simulation. Journal of the American Statistical Association, 63:1517–1521, 1968.

A.L. Hodgkin and A.F. Huxley. A quantitative description of membrane current and its application to conduction and excitation in nerves. Journal of Physiology, 117:500–544, 1952.

J. Inoue, S. Sato, and L. Ricciardi. On the parameter estimation for diffusion models of single neuron's activities. Biological Cybernetics, 73:209–221, 1995.

S. Iyengar and Q. Liao. Modeling neural activity using the generalized inverse gaussian distribution. Biological Cybernetics, 77:289–295, 1997.

S. Iyengar and P. Mullowney. Inference for the Ornstein-Uhlenbeck Model for Neural Activity. Submitted to: submitted to The Annals of Statistics, 2007.

B. Jorgensen. The Generalized Inverse Gaussian Distribution. Springer, 1981.

P.O. Kano, B. Moysey, and J.V. Moloney. Application of weeks method for the numerical inversion of the laplace transform of the matrix exponential. Computational Mathematical Sciences, 3:335–372, 2005.

S. Karlin and H. Taylor. A Second Course in Stochastic Processes. Academic Press, 1981.

N.N. Lebedev. Special Functions and their Applications. Dover Publications, 1972.

E.L. Lehmann. Theory of Point Estimation. John Wiley and Sons, 1983.

V. Linetsky. Computing hitting time densities for cir and ou diffusions: applications to mean-reverting models. Journal of Computational Finance, 7(4):1–22, 1004.

J.C.P Miller. National Physical Laboratory, Tables of Weber parabolic cylinder functions. Her Majesty's Stationary Office, 1955.

L. Ricciardi. Diffusion processes and related topics in biology. In Lecture Notes in Biomathematics, volume 16. Springer-Verlag, 1977.

L. Ricciardi and L. Sacerdote. The ornstein-uhlenbeck process as a model for neuronal activity. Biological Cybernetics, 35:1–9, 1977.

L. Ricciardi and S. Sato. First passage time density and moments of the ornstein-uhlenbeck process. J. of Applied Probability, 25:43–57, 1988.

A.J.F. Siegert. On the first passage time probablity functioin. Physical Review, 81: 617–623, 1951.

R.B. Stein. A theoretical analysis of neuronal variability. Biophysical Journal, 5: 173–194, 1965.

M.A. Stephens. Edf statistics for goodness of fit and some comparisons. Journal of the American Statistical Association, 69:730–737, 1974.

H. Tuckwell. Introduction to Theoretical Neurobiology. Volume 2: Nonlinear and Stochastic Theories. Cambridge University Press, 1988.

H. Tuckwell. A study of some diffusion models of population growth. Theoretical Population Biology, 5:345–357, 1974.

H. Tuckwell. Stochastic Processes in the Neurosciences. SIAM, 1989.

G. Uhlenbeck and L. Ornstein. On the theory of brownian motion (1930). In Selected papers in Noise and Stochastic Processes. Dover Publications, n. wax edition, 1954.

W.T. Weeks. Numerical inversion of the laplace transform using laguerre functions. Journal of Association of Computational Mathematics, 13:419–429, 1966.