# Analytic Approximation of Transition Probabilities

Euler-Maruyama and Laplace Method Modified to Transition Probabilities

**Jonathan Foley**

Centre for Math Science
Faculty of Science
Lund University
Sweden
Date:TBD

# Contents

# 1 Abstract

A transition probability is essentially a likelihood of 'something random' transitioning from one state of being to another. Though, more formally, for all intents and purposes, the 'something random' is a sequence of random events, which is a *stochastic process*.

There are many stochastic processes that are valuable to understand. Examples can be found in a multitude of topics, from finance to biology and physics. Many naturally occurring stochastic processes cannot be explicitly defined. However, there are many stochastic models that describe their behaviour very well. Once a stochastic model is established, transition probabilities for the underlying stochastic process can also be described, which in turn allows some predictability.

Estimations of transition probabilities are often limited in terms of convergence. The Euler-Maruyama (E-M) method for instance, is a numerical method for approximating Stochastic differential equations that represent the behaviour of stochastic processes. This means that it effectively approximates transition probabilities between time-steps of a process. The E-M method is weakly convergent with order 1, which entails that the error of the approximation decreases linearly with the size of the time-step.

In this study, the aim is to beat the aforementioned linear decrease in error of approximation via an analytic approximation using a combination of the E-M method and the Laplace method for some well known stochastic models. The two methods work neatly in conjunction, however, a correctional function is necessary for the Laplace method to work due to the nature of the problem. The resulting approach shows astonishingly good results with room for further improvements.

# 2 Transition Probability

A transition probability is the probability of transition of a stochastic process from a certain state $X(\tau_i) = x_i$ to another. State changes will often be time dependent and so the transition probabilities are as well. The probability in (1) symbolizes the probability of a process $X_t$ being in state $x_N$ at time $\tau_N$ given that at time $\tau_0$ the process was in state $x_0$

$$p(X_{\tau_N} = x_N | X_{\tau_0} = x_0) \ . \tag{1}$$

When analysing transition probabilities of a time dependent process $X_t$, as in (1), it is an advantage to look at an interval of time $[0, T]$ partition into $n$ equidistant subintervals $0 = \tau_0 < \tau_1 < \tau_2 < \ldots < \tau_N = T$. For notational simplicity the transition probability in (1) is written: $p(X_{\tau_N}|X_{\tau_0})$.

The approximation of transition probabilities considered in this study is approached by attempting to adapt the probabilities to an integral form for which Laplace method is applicable. In order to express (1) in integral form, the process $X_t$ is assumed to be Markovian[1], which is the case for many relevant stochastic models. Then by using the Law of Total Probability, and the Strong Markov Property[2], the transition probability (1) can be expressed as an integral of a product of transition probabilities as follows.

$$\begin{aligned} p(X_{\tau_n}|X_{\tau_0}) &= \int_\Omega p(X_{\tau_n}, X_{\tau_{n-1}}, ..., X_{\tau_1}|X_{\tau_0}) dX_{\tau_{n-1}}, ..., dX_{\tau_1} \\ &= \int_\Omega p(X_{\tau_n}|X_{\tau_{n-1}}) \cdot p(X_{\tau_{n-1}}|X_{\tau_{n-2}}) \cdot ... \cdot p(X_{\tau_2}|X_{\tau_1}) \cdot p(X_{\tau_1}|X_{\tau_0}) dX_{\tau_{n-1}}, ..., dX_{\tau_1} \ . \end{aligned} \tag{2}$$

The transition probability factors inside the integral can be represented as Gaussian by utilizing the Euler - Maruyama (E-M) method which is focal to then applying the Laplace method to (2).

**Euler-Maruyama Method**

The E-M method is a numerical approximation of a stochastic differential equation (SDE). The main application of the E-M method in this study, is that it enables the factors $p(X_{\tau_i}|X_{\tau_{i-1}})$ in (2) to be represented as Gaussian densities. The following is to show the reasoning behind the E-M method.

The E-M method approximates a SDE of the following form. (where $a, b$ are potentially some $C^4$ smooth functions)

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

with initial condition $X_0 = x_0$ and where $W_t$ is the Weiner process with independent Gaussian increments and $W_0 = 0$.

The E-M approximation is found through means of a discretization of the interval of interest such as $0 = t_0 < t_1 < t_2 < ... < t_N = T$ in a typical case $[0, T]$, where $\delta = t_n - t_{n-1} = T/N$ for $n \in \{0, ..., N\}$ $X_n$ is now recursively defined as

$$X_{n+1} = X_n + \int_{t_n}^{t_{n+1}} a(X_t)dt + \int_{t_n}^{t_{n+1}} b(X_t)dW_t \ .$$

By Itô - Taylor expansion the following approximations are made:

$$\int_{t_n}^{t_{n+1}} a(X_t)dt \approx a(X_n)\delta \ ,$$

[1] The process is only dependent on the last known value

[2] $\tau_i$ are stopping times for the process

$$\int_{t_n}^{t_{n+1}} b(X_t)dW_t \approx b(X_n)\Delta W_n \ .$$

The following recursive relationship can be tested numerically for convergence order

$$X_{n+1} = X_n + a(X_n)\delta + b(X_n)\Delta W_n \ ,$$

with $\Delta W_n = W_{n+1} - W_n \sim N(0,\delta)$ . It is commonly known that if $X$ is Gaussian with mean $\mu$ and variance $\delta$ then

$$Z = \frac{X-\mu}{\sqrt{\delta}} \Rightarrow \Delta W_n \sim \sqrt{\delta}N(0,1)$$

where $Z$ is standard Gaussian. In other words, the probability to transition from $X_n$ to $X_{n+1}$ is Gaussian with mean $a(X_n)\delta$ and standard deviation $b(X_n)\sqrt{\delta}$, from the E-M method it follows that

$$p(X_{\tau_{n+1}}|X_{\tau_n}) \cong \varphi\left(X_{\tau_{n+1}}; X_{\tau_n} + a(X_{\tau_n})\delta, b(X_{\tau_n})\sqrt{\delta}\right) \tag{3}$$

where $\varphi\left(X;\mu,\sigma\right)$ is common convention for that $X$ is a Gaussian variable with mean $\mu$ and standard deviation $\sigma$.

**Laplace Method**

The Laplace method is now relevant as the factors $p(X_{\tau_i}|X_{\tau_{i-1}})$ are of the form

$$\frac{1}{\sigma(X_{\tau_{i-1}})\sqrt{2\pi\delta}} \exp\left(-\frac{1}{2}\left[\frac{X_{\tau_i} - a(X_{\tau_{i-1}})\delta)}{\sqrt{\delta}b(X_{\tau_{i-1}})}\right]^2\right) \ ,$$

where $b(X_{\tau_{i-1}})$ and $a(X_{\tau_{i-1}})$ represent some functions which determine the standard deviation and expectation of $X_{\tau_i}$ respectively. Utilizing the fact that products of exponentials are exponentials of sums, it is reasonable to say that the integral in (2) will have the form

$$\int_\Omega g(X_{\tau_n}, X_{\tau_{n-1}}, \ldots, X_{\tau_0}) e^{f(X_{\tau_n}, X_{\tau_{n-1}}, \ldots, X_{\tau_0})} dX_{\tau_{n-1}}, \ldots, dX_{\tau_1} \ ,$$

which is precisely the type of function that the Laplace method is made to approximate.

The Laplace approximation gives an analytic approximation of an integral of the form

$$\int_\Omega g(\mathbf{x}) e^{cf(\mathbf{x})} d\mathbf{x} \ ,$$

If the function $f$ of variables $\mathbf{x} = x_1, x_2, ..., x_n$ from some domain $\Omega$ to the real numbers $\mathbf{R}^n$, $(f : \Omega \longrightarrow \mathbf{R}^n)$, has a unique maximum $\mathbf{x}^\bigstar$ in $\Omega$, is three times differentiable close to $\mathbf{x}$ and $(g : \Omega \longrightarrow \mathbf{R}^n)$ is continuously differentiable in said neighbourhood and bounded in $\Omega$, then:

$$\int_\Omega g(\mathbf{x}) e^{cf(\mathbf{x})} d\mathbf{x} = e^{cf(\mathbf{x}^\bigstar)} \left(\frac{2\pi}{c}\right)^{\frac{n}{2}} \left(\frac{g(\mathbf{x}^\bigstar)}{\sqrt{\det H(f)(\mathbf{x}^\bigstar)}} + \frac{\mathcal{O}(c)}{\sqrt{c}}\right) \ , \tag{4}$$

where $H(f)(\mathbf{x}^\bigstar)$ represents the Hessian of $f$ evaluated at $\mathbf{x}^\bigstar$.

Here $\mathcal{O}(c)$ is an error term function which decreases faster than $\sqrt{c}$ as $c$ grows, $c$ is a scalar and the crux (arguably) of the approximation is that it grows more accurate as c increases (Łapiński, 2019). When dealing with a single variable, the approximation can be written as:

$$\int_\Omega g(x) e^{cf(x)} dx = e^{cf(x)^\bigstar} \sqrt{\frac{2\pi}{c|f''(x^\bigstar)|}} \left(g(x^\bigstar) + \frac{\mathcal{O}(c)}{\sqrt{c}}\right) \ . \tag{5}$$

## 2.1 Explicit Solution to a Simple Transition Probability

In order to test the validity of using the Laplace method, it is appropriate to look at the case when $n = 2$ since the conditional probability can be solved explicitly, and compare the result with the Laplace approximation. Following the expression in (2), the transition probability when $n = 2$ is written

$$p(X_{\tau_2}|X_{\tau_0}) = \int_\Omega p(X_{\tau_2}|X_{\tau_1}) \cdot p(X_{\tau_1}|X_{\tau_0}) dX_{\tau_1} \ . \tag{6}$$

The conditional probabilities in the integral are then evaluated using the explicit E-M method. In other words, the conditional distribution of $X_{\tau_1}$ given $X_{\tau_0}$ will inherit parameters from the distribution of the process at the previous time-step $X_{\tau_0}$ as discussed in the previous section. Then,

$$p(X_{\tau_1}|X_{\tau_0}) \cong \varphi\left(X_{\tau_1}; X_{\tau_0} + \mu(X_{\tau_0})\delta, \sigma(X_{\tau_0})\sqrt{\delta}\right) = \frac{1}{\sigma(X_{\tau_0})\sqrt{2\pi\delta}} e^{-\frac{1}{2}\left[\frac{X_{\tau_1} - (X_{\tau_0} + \mu(X_{\tau_0})\delta)}{\sqrt{\delta}\sigma(X_{\tau_0})}\right]^2} \ ,$$

with

$$\delta = \tau_i - \tau_{i-1} \ .$$

The conditional probability of $X_{\tau_2}$ on $X_{\tau_1}$ does not need to inherit parameters since information about the process at the end point of the interval is already known. So instead the implicit[3] E-M is used to infer parameters for $X_{\tau_1}$. Hence,

$$p(X_{\tau_2}|X_{\tau_1}) \cong \varphi\left(X_{\tau_2}; X_{\tau_1} + \mu(X_{\tau_2})\delta, \sigma(X_{\tau_2})\sqrt{\delta}\right) \ ,$$

*due to the form of $\varphi$ the above is identically:*

$$\varphi\left(X_{\tau_1}; X_{\tau_2} - \mu(X_{\tau_2})\delta, \sigma(X_{\tau_2})\sqrt{\delta}\right) \ .$$

The integral (6) can then be written as:

$$C_{0,2} \cdot \int_\Omega \exp\left(-\frac{1}{2}\left(\left[\frac{X_{\tau_1} - (X_{\tau_2} - \mu(X_{\tau_2})\delta)}{\sqrt{\delta}\sigma(X_{\tau_2})}\right]^2 + \left[\frac{X_{\tau_1} - (X_{\tau_0} + \mu(X_{\tau_0})\delta)}{\sqrt{\delta}\sigma(X_{\tau_0})}\right]^2\right)\right) dX_{\tau_1} \ ,$$

with

$$C_{0,2} = \frac{1}{\sigma(X_{\tau_0})\sqrt{\delta}\sigma(X_{\tau_2})\sqrt{\delta}\sqrt{2\pi}\sqrt{2\pi}} = \frac{1}{\sigma(X_{\tau_0})\sigma(X_{\tau_2})\delta 2\pi} \ .$$

This indefinite integral can be solved assuming $\mu(X_{\tau_2})$, $\mu(X_{\tau_0})$, $\sigma(X_{\tau_2})$ and $\sigma(X_{\tau_0})$ do not depend on $X_{\tau_1}$, which in this case is a reasonable assumption since the process at $\tau_0$ and $\tau_2$ are known. In the following, $\mu(X_{\tau_i})\delta = \mu_i$ and $\sigma(X_{\tau_i}) = \sigma_i$ are used for cleaner expressions.

$$\left(-\frac{1}{2\delta}\right) f(X_{\tau_1}) = \left(-\frac{1}{2\delta}\right)\left(\left[\frac{X_{\tau_1} - (X_{\tau_2} - \mu_2)}{\sigma_2}\right]^2 + \left[\frac{X_{\tau_1} - (X_{\tau_0} + \mu_0)}{\sigma_0}\right]^2\right) \ ,$$

$$\begin{aligned} f(X_{\tau_1}) &= \left(\frac{1}{{\sigma_2}^2}\right)\left[X_{\tau_1}^2 - 2X_{\tau_1}(X_{\tau_2} - \mu_2)\right] + \left(\frac{1}{{\sigma_0}^2}\right)\left[X_{\tau_1}^2 - 2X_{\tau_1}(X_{\tau_0} + \mu_0)\right] + \alpha \\ &= \left(\frac{1}{{\sigma_2}^2{\sigma_0}^2}\right)\left[({\sigma_0}^2 + {\sigma_2}^2)X_{\tau_1}^2 - 2X_{\tau_1}(\sigma_0(X_{\tau_2} - \mu_2) + \sigma_2(X_{\tau_0} + \mu_0))\right] + \alpha \ , \end{aligned}$$

where $\alpha$ does not contain any $X_{\tau_1}$ terms and

[3] Analogous to the explicit E-M but for a step 'back' in time. The implicit E-M requires more conditions than the explicit E-M on $\sigma(X_{\tau_i})$, however for the purpose of the explicit solution it can be assumed that the conditions are met without impeding on the result.

$$\alpha = \left(\frac{(X_{\tau_2} - \mu_2)^2}{{\sigma_2}^2}\right) + \left(\frac{(X_{\tau_0} + \mu_0)^2}{{\sigma_0}^2}\right) .$$

Completing the square gives

$$f(X_{\tau_1}) = \left(\frac{1}{{\sigma_2}^2{\sigma_0}^2}\right)({\sigma_0}^2 + {\sigma_2}^2)\left[X_{\tau_1} - \frac{\sigma_0(X_{\tau_2} - \mu_2) + \sigma_2(X_{\tau_0} + \mu_0)}{({\sigma_0}^2 + {\sigma_2}^2)}\right]^2 + \alpha - \beta ,$$

where $\beta$ does not contain any $X_{\tau_1}$ terms and

$$\beta = \frac{[\sigma_0(X_{\tau_2} - \mu_2) + \sigma_2(X_{\tau_0} + \mu_0)]^2}{{\sigma_2}^2{\sigma_0}^2({\sigma_0}^2 + {\sigma_2}^2)} .$$

The integral (6) can now be represented as

$$C_{0,2} \cdot \int_\Omega e^{-\frac{1}{2\delta}\left[\hat{f}(X_{\tau_1}) + \alpha - \beta\right]} dX_{\tau_1} ,$$

where $\hat{f}(X_{\tau_1}) = f(X_{\tau_1}) - \alpha + \beta$. Recall that $\alpha$ and $\beta$ do not contain the integrand, so they can be moved out of the integral. Also recognize the form of $\hat{f}(X_{\tau_1})$ and rewrite the coefficient $C_{0,2}$

$$\hat{f}(X_{\tau_1}) = \left(\frac{{\sigma_0}^2 + {\sigma_2}^2}{{\sigma_2}^2{\sigma_0}^2}\right)[X_{\tau_1} - \xi]^2 ,$$
$$C_{0,2} = \frac{1}{\sigma_0\sigma_2\delta 2\pi} = \frac{1}{\sqrt{{\sigma_0}^2 + {\sigma_2}^2}\sqrt{\delta 2\pi}}\left(\frac{\sqrt{{\sigma_0}^2 + {\sigma_2}^2}}{\sigma_0\sigma_2\sqrt{\delta 2\pi}}\right) .$$

The above entails that

$$\int_\Omega p(X_{\tau_2}|X_{\tau_1}) \cdot p(X_{\tau_1}|X_{\tau_0}) dX_{\tau_1} = \frac{e^{-\frac{1}{2\delta}[\alpha - \beta]}}{\sqrt{{\sigma_0}^2 + {\sigma_2}^2}\sqrt{\delta 2\pi}} \int_\Omega \left(\frac{\sqrt{{\sigma_0}^2 + {\sigma_2}^2}}{\sigma_0\sigma_2\sqrt{\delta 2\pi}}\right) e^{-\frac{1}{2\delta}\hat{f}(X_{\tau_1})} dX_{\tau_1} . \tag{7}$$

At this point, the form of a density function of a Gaussian random variable with some $\frac{1}{\hat{\sigma}} = \left(\frac{\sqrt{{\sigma_0}^2 + {\sigma_2}^2}}{\sigma_0\sigma_2\sqrt{\delta}}\right)$ and some $\hat{\mu} = \xi$ becomes apparent. Since the domain of integration is the entire state-space, the integral part of (5) is simply the total probability: 1, what remains is:

$$p(X_{\tau_2}|X_{\tau_0}) = \frac{e^{-\frac{1}{2\delta}[\alpha - \beta]}}{\sqrt{{\sigma_0}^2 + {\sigma_2}^2}\sqrt{\delta 2\pi}} . \tag{8}$$

Further evaluating the exponent in (6) :

$$\alpha = \left(\frac{(X_{\tau_2} - \mu_2)^2}{{\sigma_2}^2}\right) + \left(\frac{(X_{\tau_0} + \mu_0)^2}{{\sigma_0}^2}\right) ,$$
$$\beta = \frac{[\sigma_0(X_{\tau_2} - \mu_2) + \sigma_2(X_{\tau_0} + \mu_0)]^2}{{\sigma_2}^2{\sigma_0}^2({\sigma_0}^2 + {\sigma_2}^2)} ,$$

so,

$$\alpha - \beta = \frac{1}{{\sigma_2}^2{\sigma_0}^2({\sigma_0}^2 + {\sigma_2}^2)}[(X_{\tau_2} - \mu_2)^2{\sigma_0}^2({\sigma_0}^2 + {\sigma_2}^2) + (X_{\tau_0} + \mu_0)^2{\sigma_2}^2({\sigma_0}^2 + {\sigma_2}^2) - ({\sigma_0}^2)^2(X_{\tau_2} - \mu_2)^2 - {\sigma_2}^2{\sigma_0}^2(X_{\tau_2} - \mu_2)(X_{\tau_0} + \mu_0) - ({\sigma_2}^2)^2(X_{\tau_0} + \mu_0)^2]$$

$$= \frac{(X_{\tau_2} - \mu_2)^2 - (X_{\tau_2} - \mu_2)(X_{\tau_0} + \mu_0) + (X_{\tau_0} + \mu_0)^2}{{\sigma_0}^2 + {\sigma_2}^2}$$

$$= \left(\frac{((X_{\tau_2} - \mu_2) - (X_{\tau_0} + \mu_0))^2}{\sigma_0^2 + \sigma_2^2}\right) = \left(\frac{((X_{\tau_2} - X_{\tau_0}) - (\mu_2 + \mu_0))^2}{\sigma_0^2 + \sigma_2^2}\right) . \tag{9}$$

The result has the form of a Gaussian density function which depends on the initial and final points of the process

$$p(X_{\tau_2}|X_{\tau_0}) = \frac{1}{\sqrt{{\sigma_0}^2 + {\sigma_2}^2}\sqrt{\delta 2\pi}} \exp\left[\left(-\frac{1}{2\delta}\right)\frac{((X_{\tau_2} - X_{\tau_0}) - (\mu_2 + \mu_0))^2}{\sigma_0^2 + \sigma_2^2}\right]$$
$$= \varphi\left(X_{\tau_2} - X_{\tau_0}; \mu(X_{\tau_2})\delta + \mu(X_{\tau_0})\delta, \sqrt{(\sigma^2(X_{\tau_0}) + \sigma^2(X_{\tau_2}))\delta}\right) .$$

In essence, what is going on is illustrated in *figure 1.* An approximation of the intermediate distribution at time $\tau_1$ is inferred from the distributions at times $\tau_0$ and $\tau_2$. The E-M approximation becomes increasingly accurate as the distance between intermediate points decreases. In other words, the smaller $|\tau_1 - \tau_0|$ is, the smaller the error in the E-M estimate of the distribution at $X_{\tau_1}$ becomes.

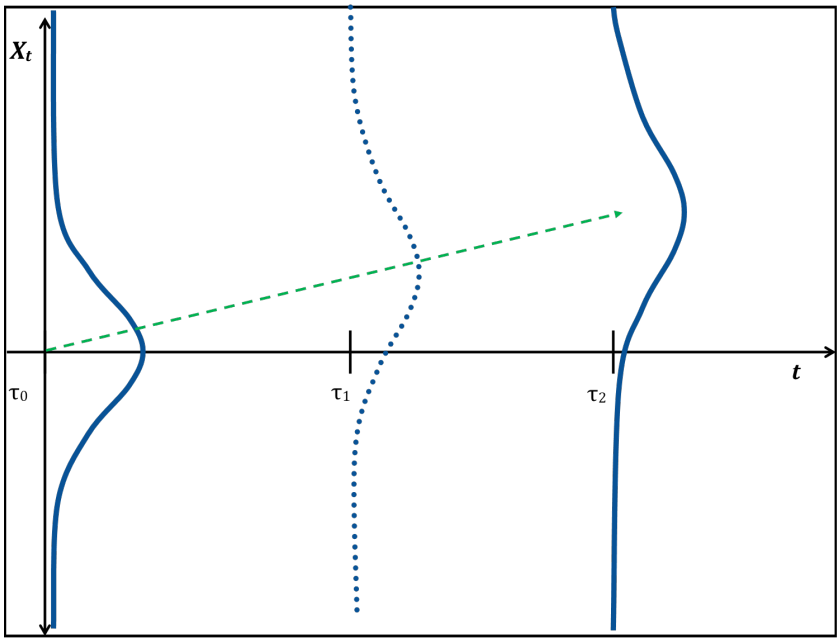


Figure 1: Illustration of possible distributions of an Arithmetic Brownian Motion process $X_t$ at times $\tau_0$ and $\tau_2$ as well as an E-M approximated distribution at time $\tau_1$. The green hatched arrow shows a hypothetical most probable path.

## 2.2 Laplace Method Solution to a Simple Transition Probability

Having solved the integral in (6) explicitly, the result can be compared to the Laplace method approximation. As before

$$
\begin{aligned}
p(X_{\tau_2}|X_{\tau_0}) &= C_{0,2} \cdot \int_\Omega \exp\left(-\frac{1}{2\delta} f(X_{\tau_1})\right) dX_{\tau_1} \ , \\
f(X_{\tau_1}) &= \left( \left[\frac{X_{\tau_1} - (X_{\tau_2} - \mu_2)}{\sigma_2}\right]^2 + \left[\frac{X_{\tau_1} - (X_{\tau_0} + \mu_0)}{\sigma_0}\right]^2 \right) \ , \\
C_{0,2} &= \frac{1}{\sigma(X_{\tau_0})\sigma(X_{\tau_2})\delta 2\pi} \ .
\end{aligned}
$$

The transition probability follows the desired form, $f(X_{\tau_1})$ is three times differentiable, and $g(x) = 1$ (while assuming $C_{0,2}$ does not depend on $X_{\tau_1}$). The derivatives are then needed to evaluate the approximation, here again it is assumed that $\mu(X_{\tau_2})$, $\mu(X_{\tau_0})$, $\sigma(X_{\tau_2})$ and $\sigma(X_{\tau_0})$ do not depend on $X_{\tau_1}$:

$$
\begin{aligned}
f'(X_{\tau_1}) &= \frac{2}{\sigma_2^2}(X_{\tau_1} - (X_{\tau_2} - \mu_2)) + \frac{2}{\sigma_0^2}(X_{\tau_1} - (X_{\tau_0} + \mu_0)) \ , \\
f''(X_{\tau_1}) &= \frac{2}{\sigma_2^2} + \frac{2}{\sigma_0^2} \ .
\end{aligned}
$$

Then setting $f'(X_{\tau_1}) = 0$ to find the maxima

$$
X_{\tau_1}\left(\frac{1}{\sigma_2^2} + \frac{1}{\sigma_0^2}\right) = \frac{1}{\sigma_2^2}(X_{\tau_2} - \mu_2) + \frac{1}{\sigma_0^2}(X_{\tau_0} + \mu_0) \ ,
$$

which implies that the maxima

$$
\begin{aligned}
X_{\tau_1}^\bigstar &= \frac{\sigma_0^2\sigma_2^2}{\sigma_0^2 + \sigma_2^2}\left(\frac{1}{\sigma_2^2}(X_{\tau_2} - \mu_2) + \frac{1}{\sigma_0^2}(X_{\tau_0} + \mu_0)\right) \\
&= \frac{\sigma_0^2}{\sigma_0^2 + \sigma_2^2}(X_{\tau_2} - \mu_2) + \frac{\sigma_2^2}{\sigma_0^2 + \sigma_2^2}(X_{\tau_0} + \mu_0) \ .
\end{aligned}
$$

It is now apparent that evaluating the function $f$ at the maxima $X_{\tau_1}^\bigstar$:

$$
\begin{aligned}
f(X_{\tau_1}^\bigstar) &= \left(\frac{1}{\sigma_0^2 + \sigma_2^2}\right)^2 (\sigma_0^2 + \sigma_2^2)((X_{\tau_2} - \mu_2) - (X_{\tau_0} + \mu_0))^2 \\
&= \left(\frac{((X_{\tau_2} - \mu_2) - (X_{\tau_0} + \mu_0))^2}{\sigma_0^2 + \sigma_2^2}\right) \ ,
\end{aligned}
$$

Since

$$
p(X_{\tau_2}|X_{\tau_0}) = C_{0,2} \cdot \int_\Omega \exp\left(-\frac{1}{2\delta} f(X_{\tau_1})\right) dX_{\tau_1} \ ,
$$

using the Laplace method approximation in (5),

$$
p(X_{\tau_2}|X_{\tau_0}) = C_{0,2} \cdot \exp\left(-\frac{1}{2\delta} f(X_{\tau_1}^\bigstar)\right) \sqrt{\frac{2\pi}{c|f''(x^\bigstar)|}} \left(1 + \frac{\mathcal{O}(c)}{\sqrt{c}}\right) \ .
$$

The above when evaluated at

$$
f(X_{\tau_1}^\bigstar) = \left(\frac{((X_{\tau_2} - \mu_2) - (X_{\tau_0} + \mu_0))^2}{\sigma_0^2 + \sigma_2^2}\right) \ ,
$$

and

$$C_{0,2}\sqrt{\frac{2\pi}{c|f''(x^{\bigstar})|}} = \frac{1}{\sigma_0\sigma_2\delta 2\pi}\sqrt{\frac{2\pi}{\frac{1}{2\delta}\left(\frac{2}{\sigma_2^2}+\frac{2}{\sigma_0^2}\right)}} = \frac{1}{\sqrt{2\pi\delta}\sqrt{{\sigma_0}^2{\sigma_2}^2\left(\frac{1}{\sigma_2^2}+\frac{1}{\sigma_0^2}\right)}} = \frac{1}{\sqrt{2\pi\delta(\sigma_2^2+\sigma_0^2)}} ,$$

gives

$$p(X_{\tau_2}|X_{\tau_0}) = \frac{1}{\sqrt{{\sigma_0}^2+{\sigma_2}^2}\sqrt{\delta 2\pi}}\exp\left[\left(\frac{1}{2\delta}\right)\left(-\frac{((X_{\tau_2}-X_{\tau_0})-(\mu_2+\mu_0))^2}{\sigma_0^2+\sigma_2^2}\right)\right]\left(1+\frac{\mathcal{O}\left(\frac{1}{2\delta}\right)}{\sqrt{\left(\frac{1}{2\delta}\right)}}\right) ,$$

The above turns out to be the exact same as the explicit solution previously solved in 2.1 if the error term is zero, which it is. This is not unexpected when considering the behaviour of the function $f(X_{\tau_1})$. The Laplace method utilizes a Taylor expansion of the function $f(X_{\tau_1})$, consequentially the error term comes from the sum of Taylor terms after the second derivative term. The function $f(X_{\tau_1})$ has no non-zero derivatives after the second derivative[4], and hence the error function term $\mathcal{O}\left(\frac{1}{2\delta}\right)$ vanishes from the above approximation.

## 2.3 Example: Arithmetic Brownian Motion

At this point it is fitting to show a simple example of what is going on using the results from 2.1 and 2.2. An appropriate simple process is the Arithmetic Brownian Motion (ABM) process $X_t$ where $\mu(\cdot)$ and $\sigma(\cdot)$ of $X_{\tau_2}$ and $X_{\tau_0}$ are both $\mu$ and $\sigma$. The SDE of an ABM is:

$$dX_t = \mu dt + \sigma dW_t .$$

Note that $\mu$ and $\sigma$ do not depend on the integrand $X_{\tau_1}$,they are in fact constants, this entails that using the explicit E-M method, the result will be exact:

$$X_{n+1} = X_n + \mu\delta + \sigma\Delta W_n .$$

In other words, the conditional probabilities can be written as

$$p(X_{\tau_1}|X_{\tau_0}) \cong \varphi(X_{\tau_1}; X_{\tau_0}+\mu(X_{\tau_0})\delta, \sigma(X_{\tau_0})\sqrt{\delta}) = \varphi(X_{\tau_1}; X_{\tau_0}+\mu\delta, \sigma\sqrt{\delta}) ,$$

$$p(X_{\tau_2}|X_{\tau_1}) \cong \varphi(X_{\tau_1}; X_{\tau_2}-\mu(X_{\tau_2})\delta, \sigma(X_{\tau_2})\sqrt{\delta}) = \varphi(X_{\tau_1}; X_{\tau_2}-\mu\delta, \sigma\sqrt{\delta}) .$$

From the solution (9), the following is found for the Arithmetic Brownian Motion transition probability:

$$p(X_{\tau_2}|X_{\tau_0}) \cong \varphi\left(X_{\tau_2}-X_{\tau_0}; 2\mu\delta, \sigma\sqrt{2\delta}\right) ,$$

where if $T=1$ and $\delta=\frac{1}{2}$

$$p(X_{\tau_2}|X_{\tau_0}) \cong \varphi\left(X_{\tau_2}-X_{\tau_0}; \mu, \sigma\right) .$$

The result is a Gaussian form distribution with parameters that agree with the model being estimated, and some representation of $X_{\tau_1}$ in terms of $X_{\tau_2}$ and $X_{\tau_0}$. In other words for this relatively straight forward process, the conditional probability corresponds to the probability of a change in the process $X_t$ of magnitude $X_{\tau_2}-X_{\tau_0}$ over the time period $T$. Note that this was for $n=2$ and $\mu(X_{\tau_2})$, $\mu(X_{\tau_0})$, $\sigma(X_{\tau_2})$ and $\sigma(X_{\tau_0})$ do not depend on $X_{\tau_1}$. However, when $n>2$ and if $\mu(X_{\tau_i})$, $\sigma(X_{\tau_i})$ are dependant on $(X_{\tau_j})$ for $i\neq j\in(1,...,n-1)$ an explicit solution will be far less trivial.

[4]Since we have assumed that $\mu(X_{\tau_2})$,$\mu(X_{\tau_0})$,$\sigma(X_{\tau_2})$ and $\sigma(X_{\tau_0})$ do not depend on $X_{\tau_1}$

# 3 The Laplace Method on General Transition probabilities

The Laplace approximation has given an exact equivalence to the explicit solution under certain conditions. As mentioned previously, it would not be simple to find an explicit solution to an integral of the form (2) when $n$ begins to grow larger, or perhaps when the integrands begin to have dependencies on one another, as is the case with several models of stochastic processes. This however, is where the Laplace approximation shines.
The Laplace approximation considers the situation when the $\sigma(X_{\tau_i})$'s may have dependence on $X_{\tau_j}$'s for $i, j \in (1, ..., n-1)$ that are being integrated in (2), since it considers a function $g(\mathbf{x})$ in in the approximation. The approximation also considers the case where the $\mu(X_{\tau_i})$'s depend on $X_{\tau_j}$'s.

It is clear from the definition of the Laplace approximation in (4) that it also allows for the multivariate case, which is the case when increasing the number of time-steps. Due to the form of the E-M method approximation, $\frac{1}{2\delta}$ can be factored out of the function in the exponent, even in the multivariate case. Then letting $\frac{1}{2\delta} = c$ such that (2) can be written

$$\int_\Omega g(X_{\tau_n}, X_{\tau_{n-1}}, \dots, X_{\tau_0}) e^{\frac{1}{2\delta}\hat{f}(X_{\tau_n}, X_{\tau_{n-1}}, \dots, X_{\tau_0})} dX_{\tau_{n-1}}, \dots, dX_{\tau_1}$$

When the number of time-steps increases, the time-step size $\delta$ decreases, which then if $\frac{1}{2\delta} = c$, $c$ will increase, which as previously discussed will decrease the error of the Laplace approximation. i.e.

$$\delta \downarrow \Rightarrow \frac{1}{2\delta} \uparrow \Rightarrow c \uparrow \Rightarrow \frac{\mathcal{O}(c)}{\sqrt{c}} \downarrow,$$

where the latter is the error term of the Laplace approximation. This is one of the core reasons for considering that the Laplace method may be a perfect candidate as estimator for transition probabilities.

However, there is (of course) a caveat. The standard Laplace method is entirely capable of estimating the transition probabilities since the functions progressively change for each increase in $c$. For each increase in number of steps, the functions $f(\mathbf{x})$ and $g(\mathbf{x})$ increase in number of variables and are, in a sense, amended to compensate for each new factor $p(X_{\tau_i}|X_{\tau_{i-1}})$. For instance, increasing $n = 2$ to $n = 3$ would change the integral from

$$\int_\Omega g(X_{\tau_2}, X_{\tau_1}, X_{\tau_0}) e^{\frac{1}{2\delta}\hat{f}(X_{\tau_2}, X_{\tau_1}, X_{\tau_0})} dX_{\tau_1}$$

to

$$\int_\Omega g(X_{\tau_3}, X_{\tau_2}, X_{\tau_1}, X_{\tau_0}) e^{\frac{1}{2\delta}\hat{f}(X_{\tau_3}, X_{\tau_2}, X_{\tau_1}, X_{\tau_0})} dX_{\tau_2} dX_{\tau_1},$$

where, though $\frac{1}{2\delta} = c$ is larger in the second case when $n = 3$, nothing is to say that the error when $n = 3$ is smaller than when $n = 2$. In other words, contrary to how the Laplace approximation is meant to be used — for explicit functions — the functions to be integrated $g(\mathbf{x}), f(\mathbf{x})$ have dependence[5] on $c$. Hence, without altering the Laplace function, $c \uparrow$ will not necessarily minimize the error term $\frac{\mathcal{O}(c)}{\sqrt{c}}$.

## 3.1 Compensating for the Function Increment Dependence

The Laplace approximation does not consider function dependence on the value $c$. Thus, the approximation does not necessarily converge to the value of the true integral solely by increasing $c$. This becomes clear when considering the evolution of $g_c(\mathbf{x})$ and $\sqrt{\det H(f_c)(\mathbf{x})}$ as $c$ grows and as the vector of variables $\mathbf{x}$ grows. In order to ensure convergence, the following factor of the Laplace Approximation must be considered:

$$\left(\frac{2\pi}{c}\right)^{\frac{n}{2}} \left(\frac{g_c(\mathbf{x}^\star)}{\sqrt{\det H(f_c)(\mathbf{x}^\star)}} + \frac{\mathcal{O}(c)}{\sqrt{c}}\right) \tag{10}$$

[5] Henceforth referred to as $f_c(x)$ & $g_c(x)$ to highlight $c$ dependence

**Considering the Hessian $H(f_c)(\mathbf{x})$:**

Owing thanks to the Markovian assumption of the stochastic processes considered, and the E-M estimation method, the Hessian of $f_c$ will be sparse since each of the $x_i$'s will only depend on the previous $x_{i-1}$ so the following holds $\forall\ i, j$ s.t $|i-j|>1$:

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = 0$$

In fact the Hessian will always be Tridiagonal when looking at functions composed by Gaussian form factors. A positive consequence is that the Hessian operations will be considerably less computationally heavy but also, the determinant of a Tridiagonal matrix of the following form (Muir & Metzler, 2003)

*Note: $\alpha$ and $\beta$ here are not related to those in 2*

$$h_n = \begin{vmatrix} \alpha_1 & \beta_1 & 0 & 0 & 0 & 0 & 0 & \dots & 0 \\ \gamma_1 & \alpha_2 & \beta_2 & 0 & 0 & 0 & 0 & \dots & \vdots \\ 0 & \gamma_2 & \alpha_3 & \beta_3 & 0 & 0 & 0 & & \\ 0 & 0 & \gamma_3 & \alpha_4 & \beta_4 & 0 & 0 & & \\ 0 & 0 & 0 & \gamma_4 & \ddots & \ddots & 0 & & \\ 0 & 0 & 0 & 0 & \ddots & & & & \\ \vdots & \vdots & & & & & & & \beta_{n-1} \\ 0 & \dots & & & & & & \gamma_{n-1} & \alpha_n \end{vmatrix} ,$$

can be found from the recurrence relation:

$$h_n = \alpha_n h_{n-1} - \gamma_{n-1}\beta_{n-1}h_{n-2}\ .$$

At this point, the form of $f_c(\mathbf{x})$ shows much symmetry and repetition, it is a sum of the exponents of Gaussian form density functions. Each $X_i \in \mathbf{x}$ is contained in only two summands of function $f_c(\mathbf{x})$:

$$\left[\frac{X_{\tau_i} - \mu(X_{\tau_{i-1}})}{\sigma(X_{\tau_{i-1}})}\right]^2 + \left[\frac{X_{\tau_{i+1}} - \mu(X_{\tau_i})}{\sigma(X_{\tau_i})}\right]^2 , \tag{11}$$

where the functions $\sigma(\cdot)$ and $\mu(\cdot)$ depend on the stochastic process $X_t$ and do not depend on $c$. Due to (11) the elements $\alpha_i, \beta_i, \gamma_i$ in the Hessian $H(f_c)(\mathbf{x})$ are functions that can be written as follows:

$$\alpha_i = \alpha(X_{\tau_{i+1}}, X_{\tau_i}, X_{\tau_{i-1}})\ , \beta_i = \beta(X_{\tau_{i+1}}, X_{\tau_i})\ , \gamma_i = \gamma(X_{\tau_i}, X_{\tau_{i-i}})\ ,$$

since

$$\alpha_i = \frac{\partial^2 f}{\partial X_i^2}\ , \beta_i = \frac{\partial^2 f}{\partial X_{i+1}\partial X_i}\ , \gamma_i = \frac{\partial^2 f}{\partial X_i \partial X_{i-1}}\ . \tag{12}$$

It is then possible to analyse how the determinant of the Hessian grows with $n = c-2$. If it is possible to find the behaviour of $\alpha_i$, $\beta_i$, $\gamma_i$, which it is if the original criteria[6] for the Laplace approximation hold, then it is possible to explicitly find how $H(f_c)(\mathbf{x}) = h_{c-2}$ grows with $c$.
When further looking at the symmetry in (11) and therefore (12), some extremely convenient results are found. To begin, (11) can be split into two parts, for simplification let

$$X_{\tau_{i-1}} = z\ ,\ X_{\tau_i} = x\ ,\ X_{\tau_{i+1}} = y$$

and

$$S^1(x,z) = \left[\frac{x-\mu(z)}{\sigma(z)}\right]^2\ ,\ S^2(y,x) = \left[\frac{y-\mu(x)}{\sigma(x)}\right]^2\ ,$$

[6] See (6)

now let $(11) = S(x,y,z) = S^1(x,z) + S^2(y,x)$ and

$$\frac{\partial^2}{\partial x^2}S = \frac{\partial}{\partial x}\left(\frac{\partial}{\partial x}S^1(x,z) + \frac{\partial}{\partial x}S^2(y,x)\right) = \frac{\partial^2}{\partial x^2}S^1(x,z) + \frac{\partial^2}{\partial x^2}S^2(y,z)\ , \tag{13}$$

$$\frac{\partial^2}{\partial x \partial z}S = \frac{\partial}{\partial x}\left(\frac{\partial}{\partial z}S^1(x,z) + \frac{\partial}{\partial z}S^2(y,x)\right) = \frac{\partial^2}{\partial x \partial z}S^1(x,z)\ , \tag{14}$$

$$\frac{\partial^2}{\partial x \partial y}S = \frac{\partial}{\partial x}\left(\frac{\partial}{\partial y}S^1(x,z) + \frac{\partial}{\partial y}S^2(y,x)\right) = \frac{\partial^2}{\partial x \partial y}S^2(y,x)\ . \tag{15}$$

Using the equality of mixed partials which only assumes that $S(x,y,z)$ is twice differentiable, which is already assumed when applying the Laplace method. It is then clearly also true that

$$\frac{\partial^2}{\partial x \partial z}S^1(x,z) = \frac{\partial^2}{\partial z \partial x}S^1(x,z) \ \ and \ \ \frac{\partial^2}{\partial x \partial y}S^2(y,x) = \frac{\partial^2}{\partial y \partial x}S^2(y,x)\ .$$

The two derivatives in (14) & (15) are identical in form. In other words, they have the property

$$\frac{\partial^2}{\partial x \partial z}S^1(x,z)\bigg|_{u,w} = \frac{\partial^2}{\partial y \partial x}S^2(y,x)\bigg|_{u,w} = \frac{\partial^2}{\partial x \partial z}S(x,y,z)\bigg|_{u,w,v} = \frac{\partial^2}{\partial y \partial x}S(x,y,z)\bigg|_{u,w,v}\ . \tag{16}$$

For some dummy variables $u,\ w,\ v$. In fact, due to symmetry in $S$, there are similar properties for the first partial derivatives

$$\frac{\partial}{\partial x}S^1(x,z)\bigg|_{u,w} = \frac{\partial}{\partial y}S^2(y,x)\bigg|_{u,w}\ ,$$

$$\frac{\partial}{\partial z}S^1(x,z)\bigg|_{u,w} = \frac{\partial}{\partial x}S^2(y,x)\bigg|_{u,w}$$

and

$$\frac{\partial}{\partial x}S(x,y,z)\bigg|_{u,w,v} = \frac{\partial}{\partial z}S^1(x,z)\bigg|_{u,v} + \frac{\partial}{\partial y}S^2(y,x)\bigg|_{w,u}\ , \tag{17}$$

then, as a consequence, due to equality of mixed partials,

$$\begin{aligned}\frac{\partial}{\partial u}\left(\frac{\partial}{\partial x}S(x,y,z)\bigg|_{u,w,v}\right) &= \frac{\partial}{\partial u}\left(\frac{\partial}{\partial z}S^1(x,z)\bigg|_{u,v}\right) + \frac{\partial}{\partial u}\left(\frac{\partial}{\partial y}S^2(y,x)\bigg|_{w,u}\right)\\ &= \frac{\partial^2}{\partial u \partial v}S^1(u,v) + \frac{\partial^2}{\partial u \partial w}S^2(w,u)\\ &= \frac{\partial^2}{\partial x \partial z}S^1(x,z) + \frac{\partial^2}{\partial x \partial y}S^2(y,x) = \frac{\partial^2}{\partial x^2}S(x,y,z)\ ,\end{aligned}$$

which combined with (17) entails that

$$\begin{aligned}&\frac{\partial^2}{\partial x^2}S(x,y,z)\bigg|_{u,w,v} = 2\cdot\frac{\partial^2}{\partial x \partial z}S^1(x,z)\bigg|_{u,v} = 2\cdot\frac{\partial^2}{\partial x \partial y}S^2(y,x)\bigg|_{w,u}\\ \Rightarrow\ &\frac{1}{2}\ \frac{\partial^2}{\partial x^2}S(x,y,z)\bigg|_{u,w,v} = \frac{\partial^2}{\partial x \partial z}S(x,y,z)\bigg|_{u,w,v} = \frac{\partial^2}{\partial x \partial y}S(x,y,z)\bigg|_{u,w,v}\end{aligned} \tag{18}$$

At this point recall that the determinant of the Hessian $H(f_c)$ can be found by the recurrence relation

$$h_n = \alpha_n h_{n-1} - \gamma_{n-1}\beta_{n-1}h_{n-2}\ ,$$

where $\gamma$ & $\beta$ are evaluated at the previous time step relative to $\alpha$, so

$$h_n = \frac{\partial^2 f}{\partial X_n^2} h_{n-1} - \frac{\partial^2 f}{\partial X_{n-1}\partial X_{n-2}} \cdot \frac{\partial^2 f}{\partial X_{n-1}\partial X_n} h_{n-2} \ ,$$

which when represented in coherent notation is

$$h_n = \frac{\partial^2 S^*}{\partial x^{*2}} h_{n-1} - \frac{\partial^2 S}{\partial x \partial z} \cdot \frac{\partial^2 S}{\partial x \partial y} h_{n-2} \ , \tag{19}$$

where $S^*$ signifies a function that maintains the shape but the variables shift to one iterative step larger, i.e.

$$z^* = x \ , \ x^* = y \ , \ y^* = X_{\tau_{i+2}} \ .$$

When using said notation, along with that $S^{-*}$ is a shift one iterative step smaller and

$$S_{z,z} = \frac{\partial^2 S}{\partial z^2} \ ,$$

then a section of the Hessian matrix would look like

$$H(f) = \begin{pmatrix} \ddots & \ddots & \ddots & & & & \\ & S^{-*}_{z,X_{\tau_i-2}} & S^{-*}_{z,z} & S^{-*}_{z,x} & & & \\ & & S_{x,z} & S_{x,x} & S_{x,y} & & \\ & & & S^*_{y,x} & S^*_{y,y} & S^*_{y,X_{\tau_{i+2}}} & \\ & & & & \ddots & \ddots & \ddots \end{pmatrix} \ ,$$

where necessarily

$$S^{-*}_{z,z}\big|_{u,w} = S_{x,x}|_{u,w} = S^*_{y,y}\big|_{u,w} \ , \tag{20}$$

which when applied to (19) gives

$$h_n = S_{x,x}|_{u,w}\, h_{n-1} - \frac{\partial^2 S}{\partial x \partial z} \cdot \frac{\partial^2 S}{\partial x \partial y} h_{n-2} \ ,$$

then from the equality in (18)

$$h_n = S_{x,x}|_{u,w}\, h_{n-1} - \frac{1}{2}\, S_{x,x}|_{u,w} \cdot \frac{1}{2}\, S_{x,x}|_{u,w}\, h_{n-2} = S_{x,x}|_{u,w}\, h_{n-1} - \left( \frac{S_{x,x}|_{u,w}}{2} \right)^2 h_{n-2} \ , \tag{21}$$

which has the solution

$$\det H(f_c)(\mathbf{x}) = h_n = (C_1 \cdot n + C_2) \left( \frac{S_{x,x}|_{u,w}}{2} \right)^n \ , \tag{22}$$

for some arbitrary $C_1$ & $C_2$ determined by the initial values of $h_n$.

**Considering $g(x)$:**

When applying the Laplace method to a product of Gaussian form density functions:

$$g_c(\mathbf{x}) = \frac{1}{\sigma(X_{\tau_0})\sqrt{\delta 2\pi}} \cdot \frac{1}{\sigma(X_{\tau_1})\sqrt{\delta 2\pi}} \cdot \ldots \cdot \frac{1}{\sigma(X_{\tau_{c-1}})\sqrt{\delta 2\pi}} \ ,$$

where, again, $n = c - 2$ since $n$ is the number of integrands in the integral (2) being estimated, and $c$ is the number of integrands but including the start and end points. The growth of $g_c(\mathbf{x})$ depends on the standard deviation function $\sigma(\cdot)$ and of course $\sqrt{\delta 2\pi}$. It is clear that $g_c(\mathbf{x})$ is affected by $c$ differently than $h_c$. This in turn entails that the quotient $\frac{g_c(\mathbf{x})}{\sqrt{h_n}}$ changes with $c$ which in turn changes the Laplace

approximation. Hence, a correctional function of $c$, let it be denoted $C(c)$ must be factored into the approximation. i.e. $C(c) = \Delta_c \frac{g_c(\mathbf{x})}{\sqrt{h_{c-2}}}$

## 3.2 Estimating Independent Function Maxima

It seems that with the appropriate correctional function, the approximation can handle most of the difficulties of solving the integral (2) which was proposed. Another problem which does arise however, is evaluating the maximum $\mathbf{x}^{\bigstar}$. In the single variable case, the maximum was found in terms of the end points of which information was already had. In the multivariate case, a maximum found by setting $f_c'(X_{\tau_i}) = 0$ will have dependence on the process at other times $\tau_j, j \in (1, ..., n-1)$ at which the process is not known, hence, another way of evaluating the maximum must be used.
One such optimization approach is to first make a crude approximation of the process at each time-step by linear-interpolation and to then improve the first approximation by taking a Newton step towards the maxima of the distribution of the process at each time-step.

$$\widetilde{x_i} = X_0 + (\tau_i - \tau_0)\frac{X_N - X_0}{\tau_N - \tau_0} = X_0 + \frac{i}{N}(X_N - X_0) \ .$$

Followed by:

$$\overline{\overline{\widetilde{x}}} = \overline{\widetilde{x}} - (H(f_c)(\overline{\widetilde{x}}))^{-1}(\bigtriangledown f_c)(\overline{\widetilde{x}}) \ , \tag{23}$$

where $\overline{\overline{\widetilde{x}}}$ and $\overline{\widetilde{x}}$ are column vectors of the approximations of the process at each intermediate time-step, $H(f_c)$ is the Hessian of $f_c$ and $(\bigtriangledown f_c)$ is the gradient vector of $f_c$ also as a column vector. The idea is to move toward the most likely outcome i.e. the maxima of the function as illustrated in *figure 2*.
The approximated maxima $\overline{\overline{\widetilde{x}}}$ will be a function of $X_N$ and $X_0$ and since the Hessian is already being calculated for the Laplace approximation, there will not be much additional computational cost for the algorithm.

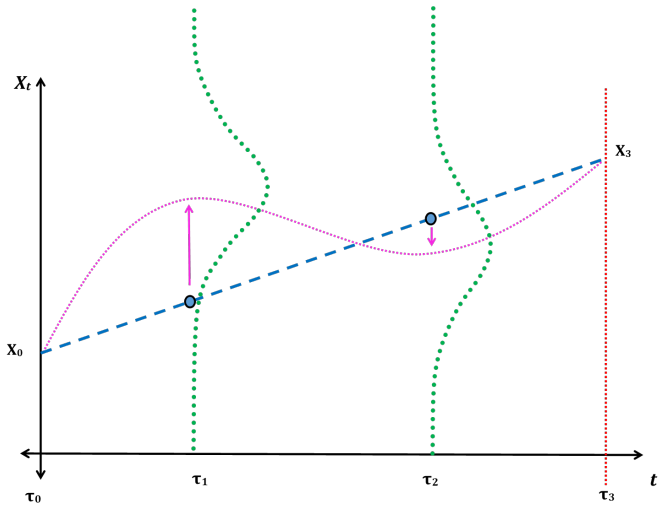


Figure 2: Illustration of Newton-Step: The blue dotted line represents the linear estimate of the maxima. The turquoise arrows show how the estimates will more towards the more probable outcome after a Newton-Step

# 4 Applying the Developed Analytic Approximation to a CIR Process

Due to the well known closed form transition probability density of the Cox-Ingersoll-Ross model and its common application in statistical probability it is a suitable candidate for evaluating the performance of the presented approximation method. Omitting particulars, the Cox-Ingersoll-Ross model follows the Stochastic differential equation:

$$dX_t = \alpha(\beta - X_t)dt + \sigma\sqrt{X_t}dW_t \ .$$

The process has a transition probability density given by:

$$p(X_{\tau_n}|X_{\tau_0}) = de^{-d(X_{\tau_n}+\gamma X_{\tau_0})}\left(\frac{X_{\tau_n}}{\gamma X_{\tau_0}}\right)^{q/2} I(q, 2d\sqrt{X_{\tau_n}\gamma X_{\tau_0}}) \ , \tag{24}$$

where

$$d = \frac{2\alpha}{\sigma^2(1-\gamma)}, \quad q = \frac{2\alpha\beta}{\sigma^2} - 1, \quad \gamma = e^{-\alpha(\tau_n - \tau_0)} \ .$$

And $I(q, Z)$ is a modified Bessel function of the first kind of order q.
From analysing the recurrence relation for the determinant and the growth of $g_c(x)$ with $c$, the correctional function $C(c)$ for the CIR process is approximately:

$$\Delta_c \frac{g_c(\mathbf{x})}{\sqrt{h_{c-2}}} = C_{CIR}(c) \cong \left(\frac{\delta^{\frac{c-3}{2}}}{\sigma\sqrt{(c-1)2^{c-2}}}\right)^{-1} \ . \tag{25}$$

The following calculations were modelled using only explicit E-M in calculations as opposed to the implicit E-M used as part of solving the simple case when n=2 in 2.1. In fact the implicit E-M require extra conditions on the function $\sigma(\cdot)$ , see (Kloeden & Platen, 2013). *Figures 3 & 4* show how increasing iterations $c = 3, 4, 5, 6, 7, 8, 9, 10, 11, 22, 35$ in the estimations approach the true transition density given the parameters.
Parameters in *Figure 3, 4 ,5 & 6* : $\alpha = 0.5$ , $\beta = 0.05$ , $\sigma = 0.15$ , $X_0 = 0.05$ and $X_n$ varies from 0.0025 to 0.2 in *Figure 3 & 4*.

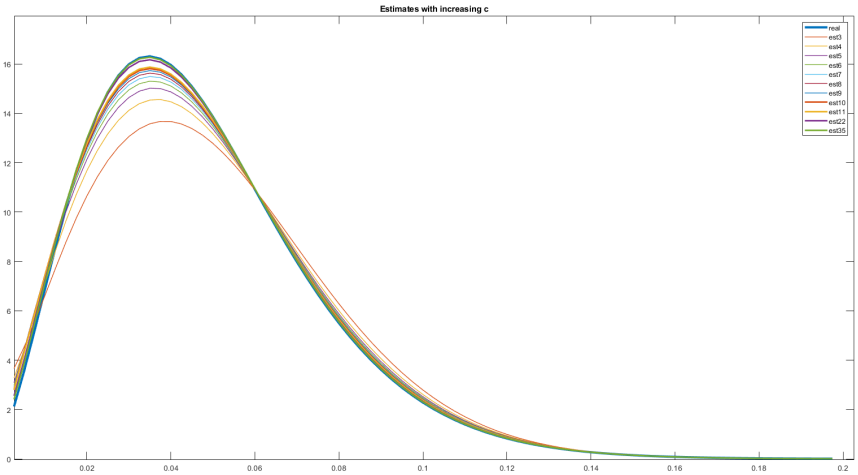


Figure 3: The estimated transition density approaching the true transition density with increasing iterations. Blue: True transition density, Red: Estimation of transition density with $c = 3$, Green: Estimation of transition density with $c = 35$

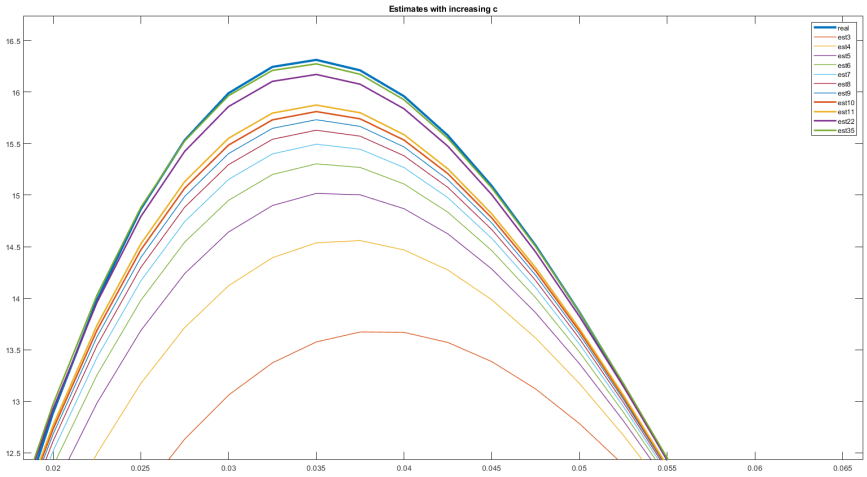


Figure 4: Zoomed in at peak of estimated transition density approaching the true transition density with increasing iterations. Blue: True transition density, Red: Estimation of transition density with $c = 3$, Green: Estimation of transition density with $c = 35$

As seen in *Figure 4* when $c = 35$ the estimate is already very close to the true density. The errors at some points of the graph are shown in *Figure 4 & 5* below.

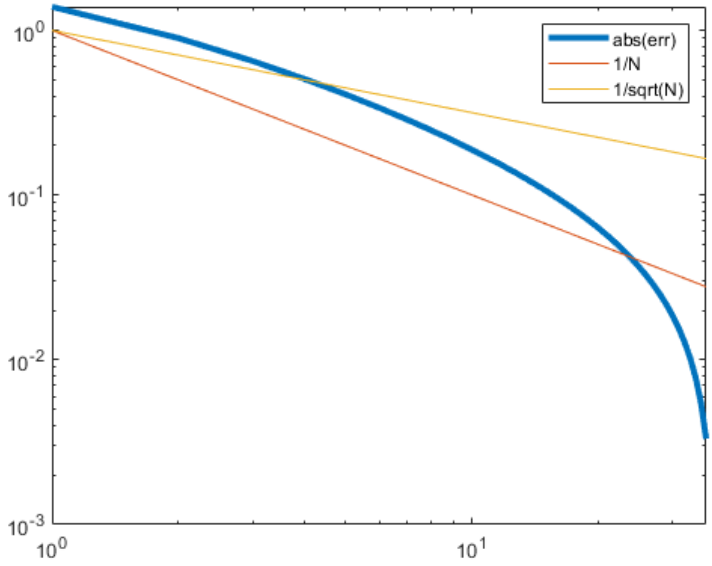


Figure 5: Loglog graph which shows how the estimates at $X_n = 0.05$ decrease with $c$ from $c = 3$ to $c = 35$ (Blue) compared to Yellow: $\frac{1}{\sqrt{c}}$ and Orange:$\frac{1}{c}$

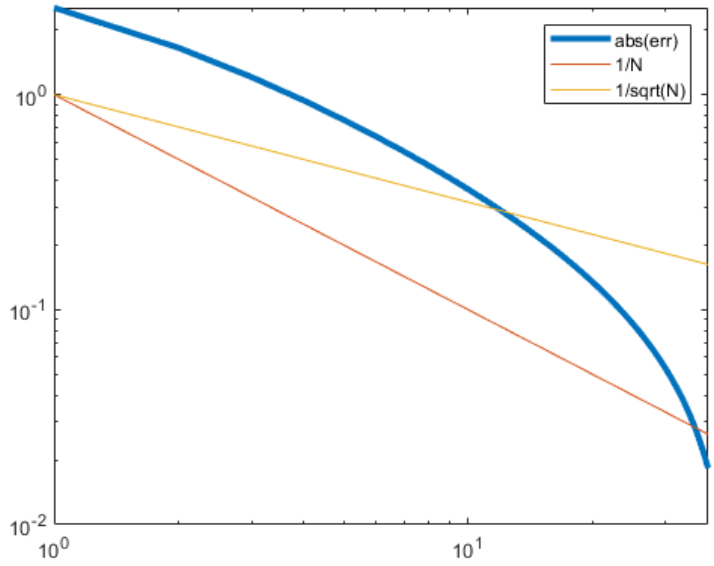


Figure 6: Loglog graph which shows how the estimates at $X_n = 0.04$ decrease with $c$ from Blue: $c = 3$ to $c = 35$, compared to Yellow: $\frac{1}{\sqrt{c}}$ and Orange:$\frac{1}{c}$

In both *Figure 5 & 6* above, clearly the errors of the estimation do not start at 0 as $\frac{1}{\sqrt{c}}$ and $\frac{1}{c}$ do, but quickly decrease faster with $c$. The errors are different at each point of the density as well, as can be seen from *Figure 5 & 6* since errors at $X_n = 0.05$ and $X_n = 0.04$ do not start at the same points. This is illustrated in *Figure 7* which compares the errors of the density estimation for $c = 11$ and $c = 22$. It is clear that errors are decreasing everywhere but start at varying points depending on $X_n$

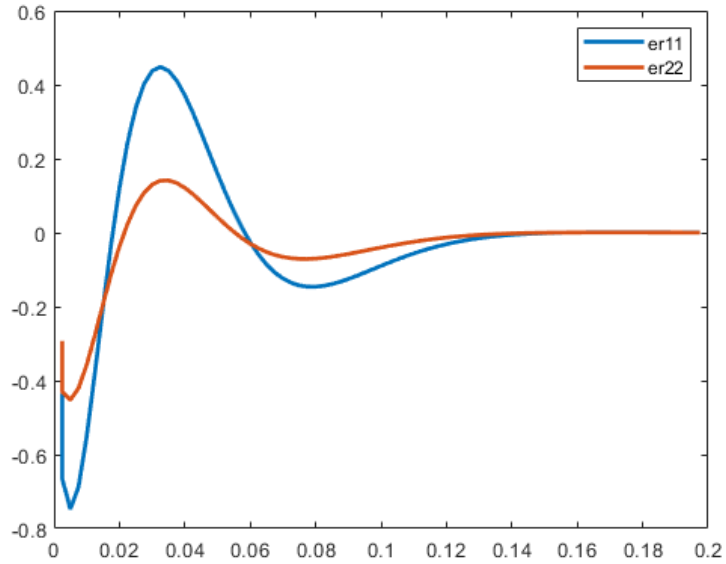


Figure 7: Errors for Blue:$c = 11$ and Orange: $c = 22$

# 5 Conclusion

The method yielded 0 error for the Arithmetic Brownian motion process, as expected since the higher derivatives that compose the error term are zero, and the linear approximation of the local maxima agree with where $f' = 0$.

Though only a few results for CIR were included, different parameters yielded very similar results. The method in general seems to perform well when $c$ is as low as 35. This is promising for quick estimations of the transition probabilities.
The code used to calculate the results for CIR were not entirely optimal, which is the reason for not exceeding $c = 35$ since the errors begin to fluctuate around $10^{-4}$. For instance, the correctional function $C(c)$ was, as mentioned, approximated, in the sense that it was not adaptive to the individual maxima points as discussed in 3.1. To improve C(c) for CIR, $h_n$ needs to be adaptive to the $\alpha(X^\star)$ since

$$h_n = (C_1 \cdot n + C_2) \left( \frac{S_{x,x}|_{u,w}}{2} \right)^n = (C_1 \cdot n + C_2) \left( \frac{\hat{\alpha}(X^\star)}{2} \right)^n ,$$

where $\hat{\alpha}$ is an average. $S_{x,x}$ should be evaluated at the estimated maxima $X^\star$.
Taking a second newton-step for the approximation of the maxima points of $f$ could also improve the approximation.

The fact that $\tau_i$'s are stopping times is not actually a necessity for the transition probability for to be represented as the integral in (2), however, the fact that they are stopping times, leads to other applications for the method used. For instance, if the process is adaptive, and $\mu(X_i)$ or $\sigma(X_i)$ change for some sub-interval of the interval in question.
The method is versatile and will work for processes with drift and volatility functions $\mu(X_i)$ and $\sigma(X_i)$ that fulfil the conditions necessary such that the Laplace method can be applied to $f$ and $g$. Judging from the error convergence, presented in for example *Figure 5*, it seems that the errors quickly converge to some small number, such that it would be redundant to use a much larger value of $c$.

# References

Kloeden, P. E., & Platen, E. (2013). *Numerical solution of stochastic differential equations* (Vol. 23). Springer Science & Business Media.

Łapiński, T. M. (2019). Multivariate laplace's approximation with estimated error and application to limit theorems. *Journal of Approximation Theory*, *248*, 105305.

Lindström, E., Madsen, H., & Nielsen, J. N. (2015). *Statistics for finance*. CRC Press.

Muir, T., & Metzler, W. H. (2003). *A treatise on the theory of determinants*. Courier Corporation.