

1 Introduction

We consider two-dimensional correlated Brownian motion with absorbing boundaries:

$$X(t) = x_0 + \mu_x t + \sigma_x W_x(t) \quad a_x \leq X(t) \leq b_x \quad (1)$$

$$Y(t) = y_0 + \mu_y t + \sigma_y W_y(t) \quad a_y \leq Y(t) \leq b_y \quad (2)$$

where W_i are standard Brownian motions with $\text{Cov}(W_1(t), W_2(t)) = \rho t$. In particular, we find the joint transition density function for $(X(t), Y(t))$ under the boundary conditions:

$$\frac{1}{dxdy} \Pr \left(\begin{array}{l} X(t) \in [x, x+dx), Y(t) \in [y, y+dy), \\ \min_{t' \in [0, t]} X(t') \geq a_x, \max_{t' \in [0, t]} X(t') \leq b_x, \\ \min_{t' \in [0, t]} Y(t') \geq a_y, \max_{t' \in [0, t]} Y(t') \leq b_y \end{array} \middle| X(0) = x_0, Y(0) = y_0, \theta \right), \quad (3)$$

with $\theta := (\mu_x, \mu_y, \sigma_x, \sigma_y, \rho)$. **Why not introduce a notation such as $M_X(t) = \max_{0 \leq s \leq t} X(s)$, $m_X(t) = \min_{0 \leq s \leq t} X(s)$, $M_Y(t) = \max_{0 \leq s \leq t} Y(s)$ and $m_Y(t) = \min_{0 \leq s \leq t} Y(s)$ to make the expression above more compact? I would also use a slightly different notation that matches that of the double lookup papers. Hongyun and I might be giving you conflicting feedback here, so it might be best for the three of us to talk.** This function, which we shorten to $q(x, y, t)$ from now on, is the solution to the Fokker-Planck equation [?]:

$$\frac{\partial}{\partial t} q(x, y, t) = -\mu_x \frac{\partial}{\partial x} q(x, y, t) - \mu_y \frac{\partial}{\partial y} q(x, y, t) + \frac{1}{2} \sigma_x^2 \frac{\partial^2}{\partial x^2} q(x, y, t) + \rho \sigma_x \sigma_y \frac{\partial^2}{\partial x \partial y} q(x, y, t) + \frac{1}{2} \sigma_y^2 \frac{\partial^2}{\partial y^2} q(x, y, t), \quad (4)$$

$$q(a_x, y, t) = q(b_x, y, t) = q(x, a_y, t) = q(x, b_y, t) = 0, \quad (5)$$

$$0 < t' \leq t.$$

Differentiating $q(x, y, t)$ with respect to the boundaries produces the transition density of a particle beginning and ending at the points $(X_1(0), X_2(0))$ and $(X_1(t), X_2(t))$, respectively, while attaining the minima a_x/a_y and maxima b_x/b_y in each coordinate direction:

$$\frac{\partial^4}{\partial a_x \partial b_x \partial a_y \partial b_y} q(x, y, t) = \text{density} \left(\begin{array}{l} X(t) = x, Y(t) = y, \\ \min_{t' \in [0, t]} X(t') = a_x, \max_{t' \in [0, t]} X(t') = b_x, \\ \min_{t' \in [0, t]} Y(t') = a_y, \max_{t' \in [0, t]} Y(t') = b_y \end{array} \middle| X(0) = x_0, Y(0) = y_0, \theta \right) \quad (6)$$

I have issues here with notation that are similar to those before. Let's talk the three of us, and please bring the double lookup paper to the discussion. The transition density (3) with less than all four boundaries has been used in computing first passage times [??], with application to structural models in credit risk and default correlations [??]. ? use variants of (6) with respect to some of the boundaries to price financial derivative instruments whose payoff depends on **some** of the observed maxima/minima. ? use a full likelihood-based (Bayesian) approach to estimate volatility in *univariate* financial timeseries where open, closing, highest, and lowest prices are included. Their work fits into a body of literature and collection of techniques by practitioners where the observed range of prices is used to make similar estimates. In this paper we will provide the efficient numerical method necessary for carrying out inferential procedures with correlated financial timeseries.

Closed-form solutions to (4) - (5) are available for some parameter regimes. When $\rho = 0$, the transition density of the process is the solution to a well-understood Sturm-Liouville problem where the eigenfunctions of the differential operator are sine functions. When $a_x = -\infty$ and $b_x = \infty$, the method of images can be used to enforce the remaining boundaries. For either $a_x, a_y = -\infty$ or $b_x, b_y = \infty$, eigenfunction of the Fokker-Planck equation can be found in radial coordinates. Both of these techniques are used and detailed by ?. However, to the best of our knowledge, there is no closed-form solution to the general problem in (4) - (5). This also limits the available ways to compute (6), with the most straightforward approach being finite

difference with respect to the boundary conditions. This, however, requires one to solve Fokker-Planck equation (4) - (5) accurately for at least 16 slightly different sets of boundaries and combine the results with numerical differentiation to evaluate the density function for a just single observation (see Section 2.1.1), motivating the need for an efficient numerical method to solve (4) - (5).

For non-zero correlation, it is still possible to approach the general problem by proposing a biorthogonal expansion in time and space (2, sections 6.2), where the eigenfunctions for the differential operator are approximated by the eigenvectors of a linear system obtained using a truncated expansion based on a set of separable basis function, each of which is a product of two sine functions (one in each dimension) satisfying the boundary conditions. However, a drawback of this out-of-the-box solution is that the system matrix for the corresponding eigenvalue problem is dense. Additionally, for nonzero ρ , it is necessary to have a large number of basis elements for an accurate approximation. The denseness and size of the resultant system matrices makes the expansion a slow, if not unfeasible, solution. An alternative is to use a finite difference scheme to directly solve the evolution problem after suitable transformations. However, both of these methods need a high degree of numerical resolution to produce practically useful approximations of the transition density. We conjecture that these inefficiencies come from either using a separable representation for the differential operator that is intrinsically correlated in the two dimensions (trigonometric series) or introducing numerical diffusion (finite difference). **I am having a bit of trouble with this last sentence, let's discuss.**

In this paper, we propose a robust and efficient solution to the general problem (4) - (5). The solution is obtained by combining a small-time analytic solution with a Galerkin discretization based on basis functions that are correlated. Our method directly takes into account the correlation parameter present in the differential operator in order to efficiently represent the analytic small-time solution and propagate it forward in time. We apply our computational method to estimate equation parameters with a maximum likelihood approach in settings where the model assumptions of constant $(\mu_x, \mu_y, \sigma_x, \sigma_y, \rho)$ and Brownian motion driving stochastic evolution are appropriate. Section 2 outlines some methods we considered, including the one we used. Section [] includes our numerical experiments.

2 Approximate Numerical Solutions

Before considering solutions to the full Fokker-Planck equation (4) - (5), we simplify the PDE by proposing a scaling transformation and an exponential decomposition of the solution, so that we can construct

$$p(x, y, t) = \exp(\alpha x + \beta y + \gamma t) q(x(b_x - a_x) + a_x, y(b_y - a_y) + a_y, t), \quad (7)$$

where

$$\begin{aligned} \alpha &= -\frac{\mu_x}{\sigma_x^2} - \frac{\rho}{\sigma_x \sigma_y (1 - \rho^2)} \left(-\frac{\mu_y}{\sigma_y^2} + \frac{\mu_x \rho}{\sigma_x \sigma_y} \right), \\ \beta &= \left(-\frac{\mu_y}{\sigma_y^2} + \frac{\mu_x \rho}{\sigma_x \sigma_y} \right), \\ \gamma &= \frac{1}{2} \left(\frac{\sigma_x}{(b_x - a_x)} \right)^2 \alpha^2 + \frac{1}{2} \left(\frac{\sigma_y}{(b_y - a_y)} \right)^2 \beta^2 + \alpha \beta. \end{aligned}$$

This new formula satisfies the **normalized** diffusion equation:

$$\frac{\partial}{\partial t} p(x, y, t) = \mathcal{L} p(x, y, t), \quad (x, y) \in (0, 1) \times (0, 1) := \Omega \quad (8)$$

subject to the constraints

$$\begin{aligned} p(x, y, t) &= 0, & \text{for } (x, y) \in \{x|x=0\} \cup \{y|y=0\}, \\ p(x, y, 0) &= \delta \left(x - \frac{x_0 - a_x}{b_x - a_x} \right) \delta \left(y - \frac{y_0 - a_y}{b_y - a_y} \right), \end{aligned}$$

where the differential operator \mathcal{L} takes the form

$$\mathcal{L} = \frac{1}{2}\tau_x^2 \frac{\partial^2}{\partial x^2} + \rho\tau_x\tau_y \frac{\partial^2}{\partial x\partial y} + \frac{1}{2}\tau_y^2 \frac{\partial^2}{\partial y^2},$$

and $\tau_x = \frac{\sigma_x}{(b_x - a_x)}$, $\tau_y = \frac{\sigma_y}{(b_y - a_y)}$. Note here that under this transformation ρ remains the same as in the original problem. **Without loss of generality, we will concentrate in the solution of the normalized problem (8) in terms of the diffusion parameters (τ_x, τ_y, ρ) .**

2.1 Eigenfunction Expansion

Following Section 6.2 of ?, we may use the biorthogonal decomposition of the solution as a sum of eigenfunctions and time-dependent coefficients determined by eigenvalues:

$$p(x, y, t) = \sum_{\mathbf{v}} h_{\mathbf{v}} \phi_{\mathbf{v}}(x, y) e^{-\lambda_{\mathbf{v}} t}, \quad (9)$$

where $h_{\mathbf{v}}$ is the coefficient of $\phi_{\mathbf{v}}(x, y)$ is the eigenfunction expansion of $p(x, y, 0)$. Because the differential operator \mathcal{L} is self-adjoint, the family of eigenfunctions is complete in the Hilbert space $L_2(\Omega)$. Moreover, the eigenvalues are bounded below by a positive constant c , so that the solution behaves as expected (see section 6.3 of ?).

Since we require $\phi_{\mathbf{v}}(x, y)$ to be zero on the boundaries, we approximate the eigenfunction using a finite set of orthogonal basis functions satisfying boundary conditions, i.e., a finite sequence of sines,

$$\phi_{\mathbf{v}}(x, y) = \sum_{l=1}^L \sum_{m=1}^M c_{l,m,\mathbf{v}} \sin(2\pi l x) \sin(2\pi m y) := \Psi(x, y)^T c_{\mathbf{v}},$$

where we have truncated the infinite series for some suitably large L and M and defined

$$\begin{aligned} \psi_{l,m}(x, y) &= \sin(2\pi l x) \sin(2\pi m y), \\ \Psi(x, y) &= (\psi_{1,1}(x, y), \dots, \psi_{L,M}(x, y))^T, \\ c_{\mathbf{v}} &= (c_{1,1,\mathbf{v}}, \dots, c_{L,M,\mathbf{v}})^T. \end{aligned}$$

The biorthogonal representation (9) leads to the eigenvalue problem

$$\mathcal{L}\phi_{\mathbf{v}} = -\lambda_{\mathbf{v}}\phi_{\mathbf{v}}, \quad (10)$$

where \mathcal{L} is the differential operator in the normalized Fokker-Planck equation. Applying \mathcal{L} to the basis function expansion of $\phi_{\mathbf{v}}$ and again approximating the result using the finite set of basis functions yields

$$\mathcal{L}\phi_{\mathbf{v}} = \mathcal{L}(\Psi(x, y)^T c_{\mathbf{v}}) = \mathcal{L}(\Psi(x, y)^T) c_{\mathbf{v}} = (A\Psi(x, y))^T c_{\mathbf{v}},$$

where A is a constant matrix dependent on $\theta = (\mu_x, \mu_y, \sigma_x, \sigma_y, \rho)$. **Well, before you said you would work with the normalized problem as a function of τ_x , τ_y and ρ , so θ should correspond to those three parameters. This is a detail, but an important one.** In the last part of the equation above, we have truncated the infinite sine series expansion of $\mathcal{L}\psi_{l,m}(x, y)$. In the case where $\rho = 0$, A is diagonal because $\{\psi_{l,m}(x, y)\}$ are the eigenfunctions to \mathcal{L} . When $\rho \neq 0$, A is no longer diagonal and is in fact dense. **The dense structure of matrix A is caused by the mixing terms**

$$\frac{\partial^2}{\partial x \partial y} \sin(2\pi l x) \sin(2\pi m y) = (2\pi l)(2\pi m) \cos(2\pi l x) \cos(2\pi m y)$$

being the products of cosine functions, which have an inefficient sine series representation [CITE]. THIS COULD BE REMOVED? Substituting the linear representation of $\mathcal{L}\phi_{\mathbf{v}}$ into the eigenvalue problem (10), we arrive at the matrix eigenvalue problem

$$\Psi(x, y)^T A^T c_{\mathbf{v}} = -\lambda_{\mathbf{v}} \Psi(x, y)^T c_{\mathbf{v}} \quad \Leftrightarrow \quad A^T c_{\mathbf{v}} = -\lambda_{\mathbf{v}} c_{\mathbf{v}}$$

whose solution gives the family of orthonormal eigenfunctions. As mentioned already, the efficiency of this approach is dependent on the cost of solving the eigenvalue problem $A^T c_V = -\lambda_V c_V$. With all of the eigenpairs (c_V, λ_V) , the approximate solution is then

$$p(x, y, t) \approx \Psi(x, y)^T \sum_V h_V c_V e^{-\lambda_V t}.$$

2.1.1 Likelihood Calculation

I would call it “Computing the density function” and not likelihood. Likelihood implies that you are going to only do inference, but we are trying to present this in a more general context

Recalling equation (6), to compute the **density** we are interested in we must take derivatives of the solution $p(x, y, t)$ with respect to the boundary parameters (a_x, b_x, a_y, b_y) . Because the eigenvalues and eigenvectors for this problem are functions of the parameters without explicit analytic form, $\frac{\partial^4}{\partial a_x \partial b_x \partial a_y \partial b_y} p(x, y, t)$ must be computed numerically using a finite difference approximation. We must therefore find the eigenvalues and eigenvectors for each of the sixteen perturbed problems

$$p(x, y, t | a_x \pm \epsilon, b_x \pm \epsilon, a_y \pm \epsilon, b_y \pm \epsilon).$$

The 4th derivative of p with respect to the 4 boundaries is approximated as

$$\begin{aligned} & \frac{\partial^4}{\partial a_x \partial b_x \partial a_y \partial b_y} p(x, y, t) \\ & \approx \frac{\sum_{k_1, k_2, k_3, k_4 = \pm 1} c_{\{k_1, k_2, k_3, k_4\}} p(x, y, t | a_x + k_1 \epsilon, b_x + k_2 \epsilon, a_y + k_3 \epsilon, b_y + k_4 \epsilon)}{(2\epsilon)^4} \end{aligned}$$

Because a good, working approximation to $p(x, y, t)$ requires many terms in the basis function expansions of the eigenfunctions, and because each resultant system matrix is dense, repeated computation of the likelihood in this way is unfeasible.

Much of what is mentioned here is applicable to all methods in the sense that all methods discussed here require the solution of 16 perturbed problems. Maybe we should reposition somewhere else? Maybe right before 2.1?

2.2 Finite Difference

The A matrix here is not the same as in the previous section, right? If so, we need to use a different letter to avoid confusion.

A finite difference method which approximates the spatial derivatives in problem (8) requires the solution a system of differential equations

$$\dot{c}(t) = A c(t) \Rightarrow c(t) = \exp(At) c(0), \quad (11)$$

which reduces to the eigenvalue decomposition of a matrix A . Here, $c(t)$ is a vector consisting of values of the solution in (8) on a set of grid points over Ω at time t , and the product $A c(t)$ approximates $\mathcal{L} p(x, y, t)$. The system matrix A is dependent on the discretization scheme used to approximate \mathcal{L} . Using a **central-in-space** scheme over a **regular** grid on Ω with $\Delta x = \Delta y = h$,

$$A = \frac{1}{2} \tau_x^2 \frac{1}{h^2} A_{x,x} + \rho \tau_x \tau_y \frac{1}{4h^2} A_{x,y} + \frac{1}{2} \tau_y^2 \frac{1}{h^2} A_{y,y},$$

where each of the matrices $A_{x,x}, A_{x,y}, A_{y,y}$ approximate $\frac{\partial^2}{\partial x^2}, \frac{\partial^2}{\partial x \partial y}, \frac{\partial^2}{\partial y^2}$, respectively. For example, using $c_{l,m}(t)$ to denote the approximation of the solution at grid point (x_l, y_m) , and assembling vector $c(t)$ using the index

scheme $c_{l,m}(t) \rightarrow c_k(t)$ with $k = (l-1)M + m$, we can approximate the operator $\frac{\partial^2}{\partial x^2}$ as

$$\frac{1}{2}\tau_x^2 \frac{\partial^2}{\partial x^2} p(x_l, y_m, t) \approx \frac{1}{2}\tau_x^2 \left(\frac{c_{l+1,m}(t) - 2c_{l,m}(t) + c_{l-1,m}(t)}{h^2} \right) = \frac{1}{2}\tau_x^2 \frac{1}{h^2} A_{x,x}[k,] c(t)$$

where $A_{x,x}[k,]$ is the k^{th} row of $A_{x,x}$. Similarly,

$$\rho\tau_x\tau_y \frac{\partial^2}{\partial x\partial y} p(x_l, y_m, t) \approx \rho\tau_x\tau_y \left(\frac{c_{l+1,m+1}(t) - c_{l+1,m-1}(t) - c_{l-1,m+1}(t) + c_{l-1,m-1}(t)}{4h^2} \right) = \rho\tau_x\tau_y \frac{1}{4h^2} A_{x,y}[k,] c(t)$$

MAYBE SIMPLIFY TO SOMETHING LIKE:

where the matrices $A_{x,x}, A_{x,y}, A_{y,y}$ provide finite difference approximations to the three differential operators, e.g.,

$$A_{x,x} = \begin{pmatrix} 1 & -2 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & -2 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & 1 \end{pmatrix}$$

AND ADD THE FORM FOR $A_{y,y}$ and $A_{x,y}$? It should be noted here that a regular grid approach with a constant h independent of parameters is appealing, because it allows us to construct once and store the matrices $A_{x,x}, A_{x,y}, A_{y,y}$, which saves valuable computational resources if we are to solve the finite difference eigenproblem (11) repeatedly for different parameter values (τ_x, τ_y, ρ) .

Unlike the system matrix for the trigonometric expansion, the system matrix A here is sparse: each row of $(A_{x,x}, A_{x,y}, A_{y,y})$ is composed of all zeros except for three or four entries. This structure does not change as $h \rightarrow 0$. The eigenvalue problem is therefore much cheaper to solve.

How much of the discussion that follows overlaps with that in 2.1? And if we really need to have separate discussions for differentiability for each of the two methods, why isn't thee a subtitle for this one?

However, the fundamental limitation of using a finite difference method is that differentiation with respect to boundaries is also done using finite differences and it is of fourth order. Given this high order differentiation, h cannot be made too small because round-off error and the irregular part of the truncation error become an issue relatively quickly. More explicitly, assuming consistency and stability of the finite difference approximation to \mathcal{L} , we can write down the finite difference solution as

$$p_{FD}(x, y, t|b) = p(x, y, t|b) + h^2 F_{reg}(b) + h^2 F_{irreg}(b) + \epsilon_{mach} F_{round}(b),$$

where we have included parameter b explicitly as a simplified notation for $[a_x, b_x]$ and $[a_y, b_y]$ after shifting a_x and a_y to 0.

Note that when expressed using the chain rule, both $\frac{\partial}{\partial a_x}$ and $\frac{\partial}{\partial b_x}$ contain $\frac{\partial}{\partial b}$. As a result, $\frac{\partial^2}{\partial a_x \partial b_x}$ leads to $\frac{\partial^2}{\partial b^2}$. Although in the discussion below, for simplicity, we only illustrate the numerical differentiation on the first derivative, keep in mind that it is the second derivative that is more relevant in the calculation of $\frac{\partial^4}{\partial a_x \partial b_x \partial a_y \partial b_y} p(x, y, t)$.

In the expression of finite difference solution above, $h^2 F_{reg}(b)$ is the regular part of the truncation error from discretizing the differential operator on the grid; $h^2 F_{irreg}(b)$ is the irregular part of the truncation error from the interpolations invoked by $\{x_0, y_0, x, y\}$ not being on grid points; and $\epsilon_{mach} F_{round}(b)$ is the effect of round-off errors with $\epsilon_{mach} \sim 10^{-16}$ denoting the machine epsilon for IEEE double precision system. The coefficient, $F_{reg}(b)$, of the regular part of truncation error is a smooth function of b with derivative $= O(1)$. The coefficient, $F_{round}(b)$, in the effect of round-off errors, behaves virtually like a random variable, discontinuous in b . For the irregular part of truncation error, the coefficient $F_{irreg}(b)$ in general is continuous

in b but not smooth in b where the derivative has discontinuities of magnitude $O(\frac{1}{h})$. For example, the error in a piece-wise linear interpolation of a smooth function at position b using step h has the general form of

$$\text{Interpolation error} = O(h^2)(1 - \text{rem}(b/h, 1))\text{rem}(b/h, 1)$$

The coefficient part $F_{\text{irreg}}(b) = (1 - \text{rem}(b/h, 1))\text{rem}(b/h, 1)$ is continuous in b but not differentiable. Its first derivative has the behavior of

$$\frac{\partial}{\partial b} F_{\text{irreg}}(b) = \frac{1}{h} (1 - 2\text{rem}(b/h, 1))$$

Based on the expression we wrote out above for the finite difference solution, applying the numerical differentiation on t with step ϵ yields:

$$\begin{aligned} & \frac{p_{FD}(x, y, t|b + \epsilon) - p_{FD}(x, y, t|b - \epsilon)}{2\epsilon} \\ &= \frac{\partial}{\partial b} p(x, y, t|b) + O(\epsilon^2) + h^2 \frac{F_{\text{reg}}(b + \epsilon) - F_{\text{reg}}(b - \epsilon)}{2\epsilon} \\ &+ h^2 \frac{F_{\text{irreg}}(b + \epsilon) - F_{\text{irreg}}(b - \epsilon)}{2\epsilon} + \epsilon_{\text{mach}} \frac{F_{\text{round}}(b + \epsilon) - F_{\text{round}}(b - \epsilon)}{2\epsilon} \end{aligned}$$

In the equation above, as the step in the numerical differentiation is refined, the first line of the RHS is well behaved, converging to the true value $\frac{\partial}{\partial b} p(x, y, t|b)$ as $\epsilon \rightarrow 0$. The second line of RHS, however, is problematic. As $\epsilon \rightarrow 0$, the contribution from round-off error blows up to infinity

$$\epsilon_{\text{mach}} \frac{F_{\text{round}}(b + \epsilon) - F_{\text{round}}(b - \epsilon)}{2\epsilon} = O\left(\frac{\epsilon_{\text{mach}}}{\epsilon}\right) \rightarrow \infty \quad \text{as } \epsilon \rightarrow 0$$

The contribution from the irregular part of truncation error is

$$h^2 \frac{F_{\text{irreg}}(b + \epsilon) - F_{\text{irreg}}(b - \epsilon)}{2\epsilon} = O\left(\frac{h^2}{\max(\epsilon, h)}\right)$$

In the second order numerical differentiation, however, the contribution from the irregular part of truncation error behaves like

$$h^2 \frac{F_{\text{irreg}}(b + \epsilon) - 2F_{\text{irreg}}(b) + F_{\text{irreg}}(b - \epsilon)}{\epsilon^2} = O\left(\frac{h^2}{\max(\epsilon^2, h^2)}\right)$$

3 A Novel Semidiscrete Galerkin Method

We propose a semidiscrete Galerkin (continuous in time, discrete in space) solution to the general diffusion problem (8). The method relies on an analytic approximation for the solution $p(x, y, t)$ for small time t and a basic convergence estimate from approximation theory for semidiscrete Galerkin-type solutions to parabolic problems.

Our numerical method produces a functional representation for the approximate solution which 1) imposes a computational burden comparable to or better than that of the finite difference method and 2) is infinitely differentiable with respect to the boundary parameters, allowing us to perform the crucial numerical differentiation with respect to these parameters.

As described in Section 2.1, the solution to the model problem has the eigenfunction expansion

$$p(x, y, t) = \sum_{v=0}^{\infty} \phi_v(x, y) e^{-\lambda_v t}.$$

Proceeding with the standard Galerkin approach, we propose a solution $p^{(k)}(x, y, t)$ of similar form

$$p^{(k)}(x, y, t) = \sum_{i=0}^k c_i(t) \psi_i(x, y),$$

where the basis functions $\psi_i(x, y)$ satisfy the boundary conditions. We also require that all first- and second-order derivatives of $\psi_i(x, y)$ are in $L_2(\Omega)$, i.e. $\psi_k(x, y) \in W_2^2(\Omega)$. Since $p^{(k)}$ is an approximation to the solution p , it does not follow the differential equation exactly nor can it represent the initial condition fully. We capture this by defining residuals

$$\begin{aligned} \frac{\partial}{\partial t} p^{(k)}(x, y, t) - \mathcal{L} p^{(k)}(x, y, t) &:= R_e(k), \\ p(x, y, 0) - p^{(k)}(x, y, 0) &:= R_0(k). \end{aligned}$$

There are various conditions that could be imposed on the residual functions (see Section 2.10.3 of ? for a summary). The *orthogonality* condition coincides with the Galerkin procedure:

$$\int_{\Omega} R_e(k) \psi_i(x, y) dx dy = 0, \quad \int_{\Omega} R_0(k) \psi_i(x, y) dx dy = 0, \quad i = 0, \dots, k, \quad (12)$$

which is equivalent to the weak formulation of the heat problem

$$\begin{aligned} \langle \partial_t p^{(k)}(x, y, t), \psi_i \rangle &= \langle \mathcal{L} p^{(k)}(x, y, t), \psi_i \rangle, \\ \langle p^{(k)}(x, y, 0), \psi_i \rangle &= \langle p(x, y, 0), \psi_i \rangle, \end{aligned}$$

where $\langle \cdot, \cdot \rangle$ is the usual inner product in $L_2(\Omega)$. The orthogonality conditions (12) lead to the system of equations

$$M \dot{\mathbf{c}}(t) = S \mathbf{c}(t), \quad (13)$$

$$M \mathbf{c}(0) = \mathbf{p}(0), \quad (14)$$

where M is the mass matrix, S is the stiffness matrix, and $\mathbf{p}(0)$ is the vector projection of $p(x, y, 0)$ onto the span of S_k with elements

$$\begin{aligned} [M]_{ij} &= \int_{\Omega} \psi_i \psi_j dx dy, \\ [S]_{ij} &= -\frac{1}{2} \tau_x^2 \int_{\Omega} \left(\frac{\partial}{\partial x} \psi_i(x, y) \right) \left(\frac{\partial}{\partial x} \psi_j(x, y) \right) dx dy - \rho \tau_x \tau_y \int_{\Omega} \left(\frac{\partial}{\partial x} \psi_i(x, y) \right) \left(\frac{\partial}{\partial y} \psi_j(x, y) \right) dx dy \\ &\quad - \frac{1}{2} \tau_y^2 \int_{\Omega} \left(\frac{\partial}{\partial y} \psi_i(x, y) \right) \left(\frac{\partial}{\partial y} \psi_j(x, y) \right) dx dy \\ [\mathbf{p}(0)]_i &= \int_{\Omega} p(x, y, 0) \psi_i(x, y) dx dy. \end{aligned}$$

The entries of S_{ij} are computed with integration by parts and using the boundary conditions. The semidiscrete Galerkin approximation becomes

$$p^{(k)}(x, y, t) = \psi(x, y)^T \exp(M^{-1} S t) \mathbf{c}(0),$$

with $\psi(x, y) = (\psi_0(x, y), \dots, \psi_k(x, y))^T$.

3.1 Small-time Solution

Before solving the Galerkin equations (13) - (14), we develop a small-time analytic solution to the problem. The small-time solution is derived by considering the fundamental solution $G(x, y, t)$ for the unbounded problem in (8), which is the bivariate Gaussian density with mean and covariance determined by the initial condition and the diffusion parameters [?]. We can then find a small enough t_ϵ such that $G\left(x, y, t_\epsilon \left| \frac{x_0 - a_x}{b_x - a_x}, \frac{y_0 - a_y}{b_y - a_y} \right. \right)$ is numerically zero on the boundaries of Ω . This is done in the following way.

1. Scale and rotate the coordinate axes by

$$\begin{pmatrix} \xi \\ \eta \end{pmatrix} = \text{Transformation goes here} \begin{pmatrix} x \\ y \end{pmatrix}$$

so that the problem obeys the diffusion equation. The computational domain is transformed to the parallelogram $\tilde{\Omega}$. The transformed initial condition will be denoted as (ξ_0, η_0) .

2. The fundamental solution $G(\xi, \eta, t | \xi_0, \eta_0)$ in this coordinate frame which does not take into account boundaries follows the bivariate Gaussian probability density function

$$G(\xi, \eta, t | \xi_0, \eta_0) = \frac{1}{2\pi t} \exp\left(-\frac{(\xi - \xi_0)^2 + (\eta - \eta_0)^2}{2t}\right).$$

We define **distance** between $G(\xi, \eta, t | \xi_0, \eta_0)$ and any of the linear boundaries of $\tilde{\Omega}$ as the Euclidean norm of the segment of the vector extending from (ξ_0, η_0) and normal to the boundary. There are four such distances d_1, d_2, d_3, d_4 . Assume that they are listed in increasing magnitude.

Setting $t_\epsilon = d_2/8$ ensures that the fundamental solution $G(\xi, \eta, t_\epsilon | \xi_0, \eta_0)$ is *at most* $\approx 10^{-15}$ on the second-farthest boundary, as well as the other two farthest boundaries. In this way, $G(\xi, \eta, t_\epsilon | \xi_0, \eta_0)$ satisfies the boundary condition on the three farthest boundaries numerically.

3. Reflect the point $(\xi_0, \eta_0) \rightarrow (\xi'_0, \eta'_0)$ about the closest boundary. The image function $G(\xi, \eta, t_\epsilon | \xi'_0, \eta'_0)$ satisfies the diffusion equation and is equal to $G(\xi, \eta, t_\epsilon | \xi_0, \eta_0)$ on the closest boundary. Further, $G(\xi, \eta, t_\epsilon | \xi'_0, \eta'_0)$ takes on values less than 10^{-15} on all other boundaries, because it is outside of $\tilde{\Omega}$.

Considering the difference of the two images, the small-time solution

$$p(\xi, \eta, t_\epsilon) := G(\xi, \eta, t_\epsilon | \xi_0, \eta_0) - G(\xi, \eta, t_\epsilon | \xi'_0, \eta'_0)$$

satisfies all of the boundaries numerically and also satisfies the governing diffusion equation. Performing a change of variables produces the small-time solution $p(x, y, t_\epsilon)$.

Using Theorem 5.E of ?, we can solve for $p(x, y, t)$ by considering the smooth $p(x, y, t_\epsilon)$ as an initial condition and evolving it forward in time by $t - t_\epsilon$. This replaces initial condition orthogonality condition in (14) with

$$\begin{aligned} M\mathbf{c}(t_\epsilon) &= \mathbf{p}(t_\epsilon), \\ [\mathbf{p}(t_\epsilon)]_i &= \int_{\Omega} p(x, y, t_\epsilon) \Psi_i(x, y) dx dy. \end{aligned} \tag{15}$$

Intuitively, the bigger t_ϵ , the smaller both $R_e(k)$ and $R_0(k)$ will be, and the better our method will do.

3.2 Orthonormal Basis Family

We motivate the construction of the orthonormal basis functions by once again considering the fundamental solution for the unbounded problem (8). In the absence of boundaries, (8) is solved by the function

$$G(x, y, t | x'_0, y'_0) = \frac{1}{2\pi t \tau_x \tau_y \sqrt{1 - \rho^2}} \exp\left\{-\frac{1}{2t(1 - \rho^2)} \left(\frac{(x - x'_0)^2}{\tau_x^2} - 2\rho \frac{(x - x'_0)(y - y'_0)}{\tau_x \tau_y} + \frac{(y - y'_0)^2}{\tau_y^2} \right)\right\},$$

with $x'_0 = (x_0 - a_x)/(b_x - a_x)$, $y'_0 = (y_0 - a_y)/(b_y - a_y)$. We choose the family of basis functions $S_k = (\psi_0(x, y), \dots, \psi_k(x, y))$

$$\psi_i(x, y) = \frac{1}{2\pi\sigma^2\sqrt{1-\tilde{\rho}^2}} \exp\left\{-\frac{1}{2(1-\tilde{\rho}^2)\sigma^2}((x-x_i)^2 - 2\tilde{\rho}(x-x_i)(y-y_i) + (y-y_i)^2)\right\} x(1-x)y(1-y) \quad (16)$$

for some parameters $(\tilde{\rho}, \sigma)$ and a collection of points $\{(x_i, y_i)\}_{i=0}^k$ which form a grid over Ω .

The critical advantage of these elements, however, is that they better resolve the small-time solution $p(x, y, t_\epsilon)$ by taking into account some degree of correlation $\tilde{\rho}$ in the covariance of each kernel. Essentially, the collection $\{\psi_i(x, y|x_i, y_i, \tilde{\rho}, \sigma)\}_{i=0}^k$ is composed of fundamental solutions to a heat diffusion problem tuned by σ and $\tilde{\rho}$, tampered such that their support is on Ω , zero on the boundaries, still smooth, and inheriting the correlation structure of the fundamental solution to the problem. Moreover, all of the basis functions are infinitely differentiable with respect to the boundary and diffusion parameters, so that $\frac{\partial^4 p^{(k)}(x, y, t)}{\partial a_x \partial b_x \partial a_y \partial b_y}$ exists and can be computed numerically using finite differences.

3.3 Error Bound

A bound on the closeness of the approximate solution $p^{(k)}(x, y, t)$ to the strong solution $p(x, y, t)$ is developed in ?. Their result shows that the Galerkin approximation we use converges to the strong solution in $L_2(\Omega)$, and it motivates the thrust of our numerical solution. First, we define the *error* term

$$e^{(k)}(t) = p(x, y, t) - p^{(k)}(x, y, t),$$

as well as the norm

$$\|w\|_2 = \sum_{j=0}^{\infty} \lambda_j^2 \langle w, \phi_j \rangle$$

for the eigenpairs (λ_j, ϕ_j) of the operator \mathcal{L} . As referred to in ?, functions $w \in L_2(\Omega)$ with $\|w\|_2 < \infty$ are also in $W_2^2(\Omega)$. Finally, if we have the condition (corresponding to equation 2.1 in ?)

$$\|p(x, y, t_\epsilon) - p^{(k)}(x, y, t_\epsilon)\|_{L_2(\Omega)} \leq Ch(k)^2 \|p(x, y, t_\epsilon)\|_2 \quad (17)$$

as a decreasing function $h(k) \geq 0$, Theorem 2.1 in ? applies and we have the error estimate

$$\|e^{(k)}(t)\|_{L_2(\Omega)} \leq Ch(k)^2 \|p(x, y, t_\epsilon)\|_2. \quad (18)$$

We can ensure condition (17) is met if S_k is complete in $L_2(\Omega)$ as k grows. The other two conditions necessary for the error bound to apply are demonstrated by ? for the Galerkin method. Equation (18) can be summarized in a simple way: the *error* of the method is controlled by how much variation the small-time solution has; the *rate* of decrease of the error is controlled by how well the span of S_k represents the small-time solution compared to the variation of the initial condition as k increases..

In the context of (18), our method, with its small-time analytic solution and choice of basis functions, is specifically tailored to minimize the error between the strong solution and its Galerkin approximation under $L_2(\Omega)$.

4 Estimation

Consider the problem of estimating the parameters $(\mu_x, \mu_y, \sigma_x, \sigma_y, \rho)$ from an i.i.d. sample (Z_1, \dots, Z_n) from the process in (1) - (2), where each Z_i is the vector of random variables

$$Z_i = (X_i(t), Y_i(t), m_{x,i}, M_{x,i}, m_{y,i}, M_{y,i}).$$

We say that Z_i is sampled from the distribution corresponding to the probability density function (6)

$$Z_i \sim F(\theta),$$

where the distribution F has the usual interpretation

$$F(z = (x, y, a_x, b_x, a_y, b_y) | \theta) = \Pr(X(t) \leq x, Y(t) \leq y, m_x \leq a_x, M_x \leq b_x, m_y \leq a_y, M_y \leq b_y).$$

This estimation problem is of particular importance in quantitative finance where the model equations (1) - (2) (with various bells and whistles attached) are widely used. However, to the best of our knowledge, all current *likelihood* methods in the literature either ignore the observed maximum/minimum information or use some of it. Likelihood-free approaches, like that of ?, on the other hand suffer from not being able to be easily integrated into a bigger inferential framework.

Since we do not have a closed-form solution for the likelihood, we will use an iterative derivative-free maximization algorithm (the Nelder-Mead method; see ? for review and convergence properties) which requires repeated evaluation of the likelihood. For moderate to large samples sizes this is feasible, because our numerical method is specifically designed for computational efficiency for repeatedly evaluating the density function (6). The maximum likelihood estimator (MLE) for the true parameters based on n samples, which we call $\hat{\theta}_n := (\hat{\mu}_x, \hat{\mu}_y, \hat{\sigma}_x, \hat{\sigma}_y, \hat{\rho})$, is especially useful in practical settings when it exhibits *consistency*, i.e.: the MLE gets closer to the true parameter vector θ as more data is collected and included in the likelihood (assuming the model and its parameters remain constant during the data collection). More precise, the estimator $\hat{\theta}_n$ is consistent if

$$\Pr(|\hat{\theta}_n - \theta|) \rightarrow 0 \quad \text{as} \quad n \rightarrow \infty.$$

4.1 Consistency

In this section we prove that the MLE based on the Galerkin approximation $p^{(k)}(x, y, t)$ to the governing Fokker-Planck equation (4) is consistent. To do so, we first show that the distribution on Z based on the approximate $p^{(k)}(x, y, t)$ converges to the true distribution $F(\cdot | \theta)$. Define the probability densities

$$f(z) = f(x, y, a_x, b_x, a_y, b_y) := \Pr(X(t) \in dx, Y(t) \in dy, m_x \in da_x, M_x \in db_x, m_y \in da_y, M_y \in db_y | \theta) \quad (19)$$

$$f^{(k)}(z) = f^{(k)}(x, y, a_x, b_x, a_y, b_y) := \frac{\partial^4}{\partial a_x \partial b_x \partial a_y \partial b_y} q^{(k)}(x, y, t | a_x, b_x, a_y, b_y). \quad (20)$$

The function $f(z)$ is the probability density function corresponding to $q(x, y, t)$, the strong solution to the governing Fokker-Planck equation; $f^{(k)}(z)$ corresponds to the Galerkin approximation $q^{(k)}$. Before proceeding, we should note that both $f(z)$ and $f^{(k)}(z)$ exist. So see the case for $f(z)$, consider the relation

$$\begin{aligned} f(x, y, a_x, b_x, a_y, b_y) = \\ \Pr(X(t) \in dx, Y(t) \in dy, m_x \in da_x, M_x \in db_x, m_y \in da_y, M_y \in db_y | \theta) = \\ \mathbb{P}_W \left(\underbrace{\left\{ \omega \in W \mid X_\omega(t) = x, Y_\omega(t) = y, \inf_{t' \in [0,1]} X_\omega(t') = a_x, \sup_{t' \in [0,1]} X_\omega(t') = b_x, \inf_{t' \in [0,1]} Y_\omega(t') = a_y, \sup_{t' \in [0,1]} Y_\omega(t') = b_y \right\}}_{A(x,y,a_x,b_x,a_y,b_y)=A(z)} \right) \end{aligned}$$

where \mathbb{P}_W is the Wiener measure on the sample space W of all realizations (paths) $(X_\omega(t), Y_\omega(t))$ from the stochastic process (1) - (2) defined in the usual way using Kolmogorov's extension of measure over cylinder sets on $t \rightarrow \mathbb{R}^2$ (see ?, Section 1.2). Sets of the form $A(x, y, a_x, b_x, a_y, b_y)$ can be defined as a countable intersection/union of cylinder sets on $t \rightarrow \mathbb{R}^2$, hence they are measurable under \mathbb{P}_W ; therefore $f(z)$ exists and is bounded above by 1. Moreover, since $q(x, y, t | a_x, b_x, a_y, b_y)$ represents the integral of $f(z)$ with respect to the boundary variables, we have

$$f(x, y, a_x, b_x, a_y, b_y) = \frac{\partial^4}{\partial a_x \partial b_x \partial a_y \partial b_y} q(x, y, t | a_x, b_x, a_y, b_y).$$

The existence of $f^{(k)}(z)$ as defined in (20) is guaranteed by Theorem 4.1 in ?. The result states that weak solutions to parabolic problems are differentiable with respect to parameters as long as the weak (Galerkin-form) operator $\langle \mathcal{L}\Psi_i(x, y), \Psi_j(x, y) \rangle$ is differentiable with respect to the parameters. The condition certainly holds for the normalized problem (8) with our choice of basis functions S_k , as they are infinitely differentiable with respect to x, y , the boundaries, and the diffusion parameters.

At this point we will assume that for a sufficiently large k , $f^{(k)}(z)$ is positive for all $z \in Z$ and is integrable over Z . As such, we may regard it as a proper probability density function with the cumulative probability density and probability measure over Z being

$$F^{(k)}(z|\theta) = \int_{-\infty}^{a_x} \int_{-\infty}^{a_y} \int_{-\infty}^{b_x} \int_{-\infty}^{b_y} \int_{-\infty}^x \int_{-\infty}^y \frac{\partial^4}{\partial a'_x \partial b'_x \partial a'_y \partial b'_y} q^{(k)}(x', y', t | a'_x, b'_x, a'_y, b'_y) dx' dy' da'_x db'_x da'_y db'_y, \quad (21)$$

$$\Pr_k(A) := \int_A f^{(k)}(z) dz, \quad \text{for any measurable } A \subset Z. \quad (22)$$

We will prove that for every $z \in Z$,

$$\lim_{k \rightarrow \infty} F^{(k)}(z|\theta) = F(z|\theta).$$

First, we prove the Lemma

Lemma 1. For $z = (x, y, a_x, b_x, a_y, b_y)$,

$$\lim_{k \rightarrow \infty} \int_{a_x}^{b_x} \int_{a_y}^{b_y} f^{(k)}(x, y, a_x, b_x, a_y, b_y) dx dy = \int_{a_x}^{b_x} \int_{a_y}^{b_y} f(x, y, a_x, b_x, a_y, b_y) dx dy.$$

Proof. Define the sets of form, for $z = (x, y, a_x, b_x, a_y, b_y)$,

$$B(a_x, b_x, a_y, b_y) = \{z' \in Z \mid z' \in [a_x, b_x] \times [a_y, b_y] \times [a_x, \infty) \times (-\infty, b_x] \times [a_y, \infty) \times (-\infty, b_y]\}.$$

Elements within $B(a_x, b_x, a_y, b_y)$ are equivalent to sample paths that stay within the region $[a_x, b_x] \times [a_y, b_y]$:

$$B(a_x, b_x, a_y, b_y) = \{\omega \in W \mid X_\omega(t) \in [a_x, b_x], Y_\omega(t) \in [a_y, b_y], m_x \in [a_x, b_x], M_x \in (a_x, b_x]\}$$

Then

$$\begin{aligned} \Pr(B(a_x, b_x, a_y, b_y)) &= \int_{B(a_x, b_x, a_y, b_y)} f(z) dz \\ &= \int_{a_x}^{\infty} \int_{-\infty}^{b_x} \int_{a_y}^{\infty} \int_{-\infty}^{b_y} \int_{-\infty}^{a_x} \int_{-\infty}^{a_y} f(x', y', a'_x, b'_x, a'_y, b'_y) dx' dy' da'_x db'_x da'_y db'_y \\ &= \int_{a_x}^{b_x} \int_{a_y}^{b_y} q(x, y, a_x, b_x, a_y, b_y) dx dy, \end{aligned} \quad (23)$$

where the last equality employs the interpretation of q in (3) and we can freely change the order of integration as f is bounded for all $z \in Z$. Similarly,

$$\Pr_k(B(a_x, b_x, a_y, b_y)) = \int_{a_x}^{b_x} \int_{a_y}^{b_y} q^{(k)}(x, y, a_x, b_x, a_y, b_y) dx dy.$$

From (23), the partial derivative of the probability $\Pr(B(a_x, b_x, a_y, b_y))$ is well-defined and can be written as

$$\frac{\partial^4}{\partial a_x \partial b_x \partial a_y \partial b_y} \Pr(B(a_x, b_x, a_y, b_y)) = \int_{a_x}^{b_x} \int_{a_y}^{b_y} f(x, y, a_x, b_x, a_y, b_y) dx dy.$$

The first-order **finite difference** approximation the above expression can be expressed as a linear combination of probabilities of perturbed sets $B(\cdot)$:

$$\lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon^4} \sum_{i=1}^{16} c(i) \Pr(B(a_x + k_1(i)\varepsilon, b_x + k_2(i)\varepsilon, a_y + k_3(i)\varepsilon, b_y + k_4(i)\varepsilon)) = \frac{\partial^4}{\partial a_x \partial b_x \partial a_y \partial b_y} \Pr(B(a_x, b_x, a_y, b_y)),$$

where c and k are functions from i to the appropriate coefficients in the first-order finite difference approximation $c(i) \rightarrow \{-1, 1\}$, $k_j(i) \rightarrow \{-1, 1\}$. Using Big-O notation, for a sufficiently small ϵ

$$\sum_{i=1}^{16} c(i) \Pr_k(B(a_x + k_1(i)\epsilon, b_x + k_2(i)\epsilon, a_y + k_3(i)\epsilon, b_y + k_4(i)\epsilon)) = \epsilon^4 \frac{\partial^4}{\partial a_x \partial b_x \partial a_y \partial b_y} \Pr(B(a_x, b_x, a_y, b_y)) + O(\epsilon^5; a_x, b_x, a_y, b_y). \quad (24)$$

The convergence result in Section 3.3 implies that

$$\begin{aligned} \Pr_k(B(a_x + k_1(i)\epsilon, b_x + k_2(i)\epsilon, a_y + k_3(i)\epsilon, b_y + k_4(i)\epsilon)) &= \int_{a_x + k_1(i)}^{b_x + k_2(i)} \int_{a_y + k_3(i)}^{b_y + k_4(i)} q^{(k)}(x, y, a_x, b_x, a_y, b_y) dx dy \rightarrow \\ &\int_{a_x + k_1(i)}^{b_x + k_2(i)} \int_{a_y + k_3(i)}^{b_y + k_4(i)} q(x, y, a_x, b_x, a_y, b_y) dx dy = \Pr(B(a_x + k_1(i)\epsilon, b_x + k_2(i)\epsilon, a_y + k_3(i)\epsilon, b_y + k_4(i)\epsilon)) \\ &\text{as } k \rightarrow \infty \text{ in } L_2(\Omega). \end{aligned}$$

Hence, for a sufficiently large k dependent on the supremum over i , and given the error estimate in (18), we have the relation

$$\begin{aligned} \Pr_k(B(a_x + k_1(i)\epsilon, b_x + k_2(i)\epsilon, a_y + k_3(i)\epsilon, b_y + k_4(i)\epsilon)) &= \Pr(B(a_x + k_1(i)\epsilon, b_x + k_2(i)\epsilon, a_y + k_3(i)\epsilon, b_y + k_4(i)\epsilon)) \\ &+ O(h(k)^2; a_x + k_1(i)\epsilon, b_x + k_2(i)\epsilon, a_y + k_3(i)\epsilon, b_y + k_4(i)\epsilon) \end{aligned}$$

Note here that the dominating terms $O(h(k)^2; \cdot)$ are differentiable with respect to the boundary parameters (a_x, b_x, a_y, b_y) since q and $q^{(k)}$ have this property. Therefore, if we replace $\Pr(\cdot)$ in (24) with $\Pr_k(\cdot)$

$$\begin{aligned} \sum_{i=1}^{16} c(i) \Pr_k(B(a_x + k_1(i)\epsilon, b_x + k_2(i)\epsilon, a_y + k_3(i)\epsilon, b_y + k_4(i)\epsilon)) &= \sum_{i=1}^{16} c(i) \Pr(B(a_x + k_1(i)\epsilon, b_x + k_2(i)\epsilon, a_y + k_3(i)\epsilon, b_y + k_4(i)\epsilon)) \\ &+ \sum_{i=1}^{16} c(i) O(h(k)^2; a_x + k_1(i)\epsilon, b_x + k_2(i)\epsilon, a_y + k_3(i)\epsilon, b_y + k_4(i)\epsilon) \\ &= \epsilon^4 \frac{\partial^4}{\partial a_x \partial b_x \partial a_y \partial b_y} \Pr(B(a_x, b_x, a_y, b_y)) + O(\epsilon^5; a_x, b_x, a_y, b_y) \\ &+ \sum_{i=1}^{16} c(i) O(h(k)^2; a_x + k_1(i)\epsilon, b_x + k_2(i)\epsilon, a_y + k_3(i)\epsilon, b_y + k_4(i)\epsilon). \end{aligned}$$

Dividing both sides by ϵ^4 produces

$$\begin{aligned} \frac{1}{\epsilon^4} \sum_{i=1}^{16} c(i) \Pr_k(B(a_x + k_1(i)\epsilon, b_x + k_2(i)\epsilon, a_y + k_3(i)\epsilon, b_y + k_4(i)\epsilon)) &= \frac{\partial^4}{\partial a_x \partial b_x \partial a_y \partial b_y} \Pr(B(a_x, b_x, a_y, b_y)) + O(\epsilon; a_x, b_x, a_y, b_y) \\ &+ \frac{1}{\epsilon^4} \sum_{i=1}^{16} c(i) O(h(k)^2; a_x + k_1(i)\epsilon, b_x + k_2(i)\epsilon, a_y + k_3(i)\epsilon, b_y + k_4(i)\epsilon) \end{aligned}$$

As mentioned above $O(h(k)^2; \cdot)$ is differentiable with respect to the boundary parameters, so that the rightmost term is still $O(h(k)^2)$ as $\epsilon \rightarrow 0$. Taking the limit in ϵ , we have

$$\begin{aligned} \int_{a_x}^{b_x} \int_{a_y}^{b_y} f^{(k)}(x, y, a_x, b_x, a_y, b_y) dx dy &= \frac{\partial^4}{\partial a_x \partial b_x \partial a_y \partial b_y} \Pr_k(B(a_x, b_x, a_y, b_y)) \\ &= \frac{\partial^4}{\partial a_x \partial b_x \partial a_y \partial b_y} \Pr(B(a_x, b_x, a_y, b_y)) + O(h(k)^2; a_x, b_x, a_y, b_y) \\ &= \int_{a_x}^{b_x} \int_{a_y}^{b_y} f(x, y, a_x, b_x, a_y, b_y) dx dy + O(h(k)^2; a_x, b_x, a_y, b_y). \end{aligned}$$

Therefore, we have the desired result:

$$\lim_{k \rightarrow \infty} \int_{a_x}^{b_x} \int_{a_y}^{b_y} f^{(k)}(x, y, a_x, b_x, a_y, b_y) dx dy = \int_{a_x}^{b_x} \int_{a_y}^{b_y} f(x, y, a_x, b_x, a_y, b_y) dx dy.$$

□

Lemma 2 (Convergence in distribution). *For any $z \in Z$, $\lim_{k \rightarrow \infty} F^{(k)}(z|\theta) = F(z)$.*

Proof. Let $I^{(k)}(z) = \int_{a_x}^x \int_{a_y}^y f^{(k)}(u, v, a_x, b_x, a_y, b_y) du dv$ and let $I(z) = \int_{a_x}^x \int_{a_y}^y f(u, v, a_x, b_x, a_y, b_y) du dv$. It is possible to show that $\lim_{k \rightarrow \infty} I^{(k)}(z) = I(z)$ as a consequence of Lemma 1 by considering some $\chi_k(z) \in S_k$ approximating the indicator $1(u \leq x, v \leq y)$ as $k \rightarrow \infty$ and setting up a triangle inequality. However, we will omit this technical detail here.

Next, we know that $I(z)$ is integrable over (a_x, b_x, a_y, b_y) as $\Pr(Z) = \mathbb{P}_W(W) = 1$. The Dominated Convergence Theorem applies, and we therefore have

$$\lim_{k \rightarrow \infty} \int_{-\infty}^{a_x} \int_{-\infty}^{b_x} \int_{-\infty}^{a_y} \int_{-\infty}^{b_y} I^{(k)}(z) da'_x db'_x da'_y db'_y = \int_{-\infty}^{a_x} \int_{-\infty}^{b_x} \int_{-\infty}^{a_y} \int_{-\infty}^{b_y} I(z) da'_x db'_x da'_y db'_y,$$

which implies the result of the Lemma. □

Lemma 3. *The maximum likelihood estimator is consistent as $n \rightarrow \infty$ and $m \rightarrow \infty$:*

$$\hat{\theta}_{n,k} \rightarrow \theta$$

Proof. By Lemma 2

$$Z_k \xrightarrow{d} Z \text{ as } k \rightarrow \infty.$$

Next, given Theorem 4.1 in ?, we know that, for each k , q_k is analytic in both the diffusion parameters and boundary parameters. Hence, the probability density function satisfies the criteria A1 - A6 in ? to guarantee that, for data $Z_k \sim F_k(\theta)$,

$$\hat{\theta}_{n,k}(Z_k) \xrightarrow{p} \theta \text{ as } n \rightarrow \infty.$$

Now we need to show that the same holds for data sampled from F as $k \rightarrow \infty$. To do this, we will use Chebyshev's inequality:

$$\Pr_Z(|\hat{\theta}_{n,k}(Z) - \theta| \geq \epsilon) \leq \frac{\mathbb{E}_Z[(\hat{\theta}_{n,k}(Z) - \theta)^2]}{\epsilon^2}.$$

By the Maximum theorem [REFERENCE], $\hat{\theta}_{n,k}(x)$ is a continuous function with respect to x , and further because we have bounded $\hat{\theta}$ from below and above,

$$\mathbb{E}_{Z_k}[(\hat{\theta}_{n,k}(Z_k) - \theta)^2] \rightarrow \mathbb{E}_Z[(\hat{\theta}_{n,k}(Z) - \theta)^2] \text{ as } k \rightarrow \infty$$

by the portmanteau lemma. Finally, we can show that

$$\mathbb{E}_{Z_k}[(\hat{\theta}_{n,k}(Z_k) - \theta)^2] \rightarrow 0 \text{ as } n \rightarrow \infty, \tag{25}$$

since the expected value of the estimator tends to θ and its variance goes to 0 when $n \rightarrow \infty$. Therefore, given any $\epsilon > 0$ and $\delta > 0$, we can find a sufficiently large n and k such that

$$\Pr_Z(|\hat{\theta}_{n,k}(Z) - \theta| \geq \epsilon) \leq \frac{\mathbb{E}_Z[(\hat{\theta}_{n,k}(Z) - \theta)^2]}{\epsilon^2} < \delta$$

□