1 Introduction

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

$$X(t) = x_0 + \mu_x t + \sigma_x W_x(t) \qquad a_x < X(t) < b_x \tag{1}$$
$$Y(t) = y_0 + \mu_y t + \sigma_y W_y(t) \qquad a_y < Y(t) < b_y \tag{2}$$

where W_i are standard Brownian motions with $\text{Cov}(W_1(t), W_2(t)) = \rho t$ for $0 < t' \leq t$. In particular, we find the joint transition density function for $(X(t), Y(t))$ under the boundary conditions:

$$\Pr\left(X(t) \in dx, Y(t) \in dy | \forall t' \in [0,t]\ X(t') \in [a_x, b_x], \forall t' \in [0,t]\ Y(t') \in [a_y, b_y], X(0) = x_0, Y(0) = y_0, \theta\right), \tag{3}$$

with $\theta := (\mu_x, \mu_y, \sigma_x, \sigma_y, \rho)$. This function, which we shorten to $q(x,y,t)$ from now on, is the solution to the Fokker-Planck equation [Oksendal, 2013]:

$$\frac{\partial}{\partial t} q(x,y,t') = -\mu_x \frac{\partial}{\partial x} q(x,y,t') - \mu_y \frac{\partial}{\partial y} q(x,y,t') + \frac{1}{2}\sigma_x^2 \frac{\partial^2}{\partial x^2} q(x,y,t') + \rho\sigma_x\sigma_y \frac{\partial^2}{\partial x \partial y} q(x,y,t') + \frac{1}{2}\sigma_y^2 \frac{\partial^2}{\partial y^2} q(x,y,t'), \tag{4}$$

$$q(a_x, y, t') = q(b_x, y, t') = q(x, a_y, t') = q(x, b_y, t') = 0, \tag{5}$$
$$0 < t' \leq t.$$

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:

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

$$\Pr\left(X(t) \in dx, Y(t) \in dy, \min_{t'} X(t') = a_x, \max_{t'} X(t') = b_x, \min_{t'} Y(t') = a_y, \max_{t'} Y(t') = b_y \,\middle|\, 0 < t' \leq t, X(0) = x_0, Y(0) = y_0, \theta\right). \tag{6}$$

The transition density (3) with less than all four boundaries has been used in computing first passage times [Kou et al., 2016, Sacerdote et al., 2016], with application to structural models in credit risk and default correlations [Haworth et al., 2008, Ching et al., 2014]. He et al. [1998] 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.

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 $a_1 = -\infty$ and $b_1 = \infty$, the method of images can be used to enforce the remaining boundaries. For either $a_1, a_2 = -\infty$ or $b_1, b_2 = \infty$, eigenfunction of the Fokker-Plank equation can be found in radial coordinates. Both of these techniques are used and detailed by He et al. [1998]. 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 at least 16 eigenvalue problems to evaluate the density function for a single observation, motivating the need for an efficient numerical method to solve (4) - (5).

It is still possible to approach the general problem by proposing a biorthogonal expansion in time and space (Risken [1989], sections 6.2), where the eigenfunctions for the differential operator are approximated as sinusoidal series 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 large and dense. 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 (trigonometric series) or introducing numerical diffusion (finite difference).

In this paper, we propose a solution to the general problem (4) - (5) which is obtained by combining a small-time analytic solution with a finite-element method. 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.

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

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

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 simpler diffusion equation:

$$\frac{\partial}{\partial t} p(x,y,t) = \mathcal{L} p(x,y,t), \quad (x,y) \in (0,1) \times (0,1) := \Omega. \tag{7}$$

subject to the constraints

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

with the differential operator $\mathcal{L}$ takes the form

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

where $\tau_x = \frac{\sigma_x}{(b_x - a_x)}$, $\tau_y = \frac{\sigma_y}{(b_y - a_y)}$. Note here that under this transformation ρ remains the same as in the original problem. We will call equation (**??**) the *normalized* problem and will consider its solution in terms of the diffusion parameters (τ_x, τ_y, ρ) without a loss of generality.

2.1 Eigenfunction Expansion

Following Section 6.2 of Risken [1989], we may use the biorthogonal decomposition of the solution as a sum of eigenfunctions and time-dependent coefficients determined by eigenvalues:

$$p(x,y,t) = \phi_\nu(x,y) e^{-\lambda_\nu t}, \tag{8}$$

where the eigenfunctions $\phi_\nu(x,y)$ satisfy the boundary conditions. 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 0, so that the solution behaves as expected (see section 6.3 of Risken [1989]).

Since we require $\phi_\nu(x,y)$ to be zero on the boundaries, we may represent the eigenfunction as a linear combination of sines

$$\phi_\nu(x,y) = \sum_{l=0}^{L}\sum_{m=0}^{M} c_{l,m,\nu} \sin(2\pi l x)\sin(2\pi m y) := \Psi(x,y)^T c_\nu,$$

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

$$\begin{aligned}\psi_{l,m}(x,y) &= \sin(2\pi l x)\sin(2\pi m y),\\ \Psi(x,y) &= (\psi_{0,0}(x,y),\ldots,\psi_{L,M}(x,y))^T,\\ c_\nu &= (c_{0,0,\nu},\ldots,c_{L,M,\nu})^T.\end{aligned}$$

The biorthogonal representation (8) leads to the eigenvalue problem

$$\mathcal{L}\phi_\nu = -\lambda_\nu \phi_\nu, \tag{9}$$

where $\mathcal{L}$ is the differential operator in the normalized Fokker-Planck equation. Applying $\mathcal{L}$ to ϕ_ν produces the linear system

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

where A is a constant matrix dependent on $\tilde{\theta}$. In the case where $\rho = 0$, A is diagonal because $\{\psi_{l,m}(x,y)\}_{l,m}$ are the eigenfunctions to $\mathcal{L}$. When $\rho \neq 0$, A is no longer diagonal and is in fact dense. This 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]. Substituting the linear representation of $\mathcal{L}\phi_\nu$ into the eigenvalue problem (9), we arrive to the system

$$\Psi(x,y)^T A^T c_\nu = -\lambda_\nu \Psi(x,y)^T c_\nu \quad \Leftrightarrow \quad A^T c_\nu = -\lambda_\nu c_\nu$$

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 $A^T c_\nu = -\lambda_\nu c_\nu$.

With all of the eigenpairs (c_ν, λ_ν), the approximate solution is then

$$p(x,y,t) \approx \sum_\nu \Psi(x,y)^T c_\nu e^{-\lambda_\nu t}.$$

Recalling equation (6), to compute the likelihood 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

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

such that

$$\frac{\partial^4}{\partial a_x \partial b_x \partial a_y \partial b_y} p(x,y,t) \approx \frac{\sum_k c_k p(x,y,t|a_x \pm \varepsilon, b_x \pm \varepsilon, a_y \pm \varepsilon, b_y \pm \varepsilon)}{(2\varepsilon)^4}.$$

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

2.2 Finite Difference

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

$$\dot{c}(t) = Ac(t) \Rightarrow \qquad c(t) = \exp(At)\, c(0), \tag{10}$$

which reduces to the eigenvalue decomposition of a matrix A. Here, $c(t)$ is a vector of coefficients representing the value of the solution in (**??**) on each point on a grid over Ω at time t, and the product $Ac(t)$ 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 the labeling $c_{l,m}(t) \to c_k(t)$ on the vector $c(t)$ to denote the approximation of the solution at point (x_l, y_m) on the grid,

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

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

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

It should be noted here that a regular grid approach with a constant h 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 (10) 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 $h \to 0$. The eigenvalue problem is therefore much cheaper to solve. The system matrix can be made even sparser on a regular grid by performing a 45° rotation which removes the mixing term $\rho\tau_x\tau_y \frac{\partial^2}{\partial x \partial y}$ from the problem PDE and preserves the boundaries.

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 truncation error becomes 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

$$p_{FD}(x_l, y_m, t | a_x) = p(x_l, y_m, t) + O(h^2; a_x).$$

Note here the residual term $O(h^2; a_x)$ a function of the paramter a_x. The finite difference approximation yields:

$$\begin{aligned} \frac{p_{FD}(x_l, y_m, t | a_x) - p_{FD}(x_l, y_m, t | a_x - \varepsilon)}{\varepsilon} &= \frac{p(x_l, y_m, t | a_x) - p(x_l, y_m, t | a_x - \varepsilon)}{\varepsilon} + \frac{O(h^2 | a_x) - O(h^2 | a_x - \varepsilon)}{\varepsilon} \\ &\approx \frac{\partial}{\partial a_x} p(x_l, y_m, t) + O(\varepsilon) + \frac{O(h^2 | a_x) - O(h^2 | a_x - \varepsilon)}{\varepsilon}. \end{aligned}$$

If the residual term $O(h^2 | a_x)$ is differentiable with respect to a_x,

$$\frac{O(h^2 | a_x) - O(h^2 | a_x - \varepsilon)}{\varepsilon} \to O(h^2) \text{ as } \varepsilon \to 0.$$

Otherwise

$$\frac{O(h^2|a_x) - O(h^2|a_x - \varepsilon)}{\varepsilon} \to O(h^2/\varepsilon^r) \text{ as } \varepsilon \to 0.$$

for some $r > 0$. Under this condition, as $\varepsilon \to 0$ with h fixed, the term $O(h^2/\varepsilon^r)$ begins to dominate the approximation. Hence, for a fixed h, we have a bound on how accurately we may represent the likelihood function provided the finite difference residual term is not differentiable with respect to the boundary parameters.

The finite value for h introduces yet another practical concern. On a regular grid, the delta function initial condition needs to be either rounded to the nearest grid point or represented as a weighted sum of delta functions on the four nearest grid points. Either approach introduces a numerical diffusion into the problem which, for a finite h, can bias the numerical solution. We will demonstrate this phenomenon in Section [].

2.3 Finite Element Method

We propose a numerical method which 1) maintains a functional approximation of the differential operator $\mathcal{L}$ while 2) imposing a computational burden comparable to or better than that of the finite difference approach. Further, our method explicitly minimizes the error associated with using a finite, smooth representation of the initial condition $\delta(\xi - \xi_0)\delta(\eta - \eta_0)$.

As such, our approach consists of two parts

i) a small-time analytic solution $q(x, y, \tau_\varepsilon)$ for the IC/BC problem,

ii) a family of orthonormal basis functions which represent $q(x, y, \tau_\varepsilon)$ parsimoniously.

By combining *i*) and *ii*), we can efficiently find a weak solution to the PDE (**??**) via the finite element method [Shaidurov, 2013]. Convergence of our method to the strong solution under the $L^2(\bar{\Omega})$ norm is guaranteed as long as the family of basis functions we propose is complete in the Hilbert space under $L^2(\bar{\Omega})$ [Salsa, 2016].

The small-time solution is derived by considering the fundamental solution $G(\xi, \eta|\tau, \xi_0, \eta_0)$ for the unbounded problem in (**??**), which is the bivariate Gaussian density with mean and covariance determined by the initial condition and the diffusion parameters [Stakgold and Holst, 2011]. We can then find a small enough t_ε such that $G(\xi, \eta|\tau, \xi_0, \eta_0)$ is numerically zero on three of the four boundaries of $\bar{\Omega}$. The zero-condition on the remaining boundary is enforced by suitably moving the source term for the fundamental solution to some suitable (ξ_0', η_0'). The small-time solution therefore takes on the analytic form

$$q(\xi, \eta, \tau_\varepsilon) = G(\xi, \eta|\tau_\varepsilon, \xi_0, \eta_0) - G(\xi, \eta|\tau_\varepsilon, \xi_0', \eta_0').$$

The construction of the orthonormal basis functions is also motivated by the fundadmental solution for the unbounded problem (**??**): before performing Gram-Schmidt orthogonalization, the finite family of basis functions $\{\tilde{\psi}_k(x, y|x_k, y_k, \rho, \sigma)\}_k^K$ are of the form

$$\tilde{\psi}_k(x, y|x_k, y_k, \rho, \sigma) = N\left((x, y)^T \middle| (x_k, y_k)^T, \quad \begin{pmatrix} \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 \end{pmatrix}\right) x(1-x)y(1-y).$$

Essentially, the collection $\{\tilde{\psi}_k(x, y|x_k, y_k, \rho, \sigma)\}_k^K$ is composed of fundamental solutions to a heat diffusion problem tuned by σ and ρ, tapered such that their support is on $\bar{\Omega}$, and centered along some grid over $\bar{\Omega}$ and are still smooth.

The advantage of these elements is that they better resolve the fundamental solution for the unbounded problem by taking into account ρ in the covariance of each kernel [**if not prove, demonstrate**]. By performing Gram-Schmidt orthogonalization under the $L^2(\Omega)$ norm, we arrive at a family of orthonormal functions $\{\psi_k(x, y|x_k, y_k, \rho, \sigma)\}_k^K$ which can better resolve small-time solutions having a large correlation coefficient. [This needs a lot more work].

3 Estimation

Our numerical method is specifically designed for computational efficiency for repeatedly evaluating the density function (6). The method is therefore particularly suited for performing maximum likelihood estimation of parameter values with respect to models involving systems of the type (1) - (2). In such a scenario, the maximum likelihood estimator (MLE) for the true parameters θ are those which maximize the (log-)likelihood function conditional on the observed data. MLEs are especially useful in practical settings when they exhibit *consistency*, *i.e.*: the MLEs gets closer to the true parameter θ as more data is collected and included in the likelihood (assuming the model and its parameters remain constant).

In this section, we prove that the MLE based on the **approximate** (weak) finite element solution to the governing Fokker-Planck equation (4) is consistent. We define X as the random variable sampled from the distribution corresponding to the true density (6):

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

where we have explicitly made the density F dependent on the model parameters θ.

For a family of basis functions $\{\tilde{\psi}_l(x,y|x_k,y_k,\rho,\sigma)\}_{l=0}^{k}$, let $q_k(x,y,t|\theta)$ be the weak solution to the governing Fokker-Planck equation (4) - (5). The approximation to the true pdf is then

$$\frac{\partial^4}{\partial a_x \partial b_x \partial a_y \partial b_y} q_k(x,y,t).$$

The distribution corresponding to this density is denoted as $F_k(\theta)$. Let X_k be the random variable corresponding to this distribution:

$$X_k \sim F_k(\theta).$$

We prove Lemma 1, which states that the distribution of random variables sampled according to the approximate solution converge in density to random variables sampled from the true solution:

Lemma 1. *For X_k and X defined above, $X_k \xrightarrow{d} X$ as $k \to \infty$.*

Proof. **do not read below; I'm working on it tonight** The weak solution $q_k(x,y,t)$ has the property that for any $C^\infty(\Omega)$ function $v \in L^2(\Omega)$,

$$\lim_{k\to\infty}(q_k, v) = (q, v),$$

where $(v,u) := \int_{a_x}^{b_x}\int_{a_y}^{b_y} v(x,y,t)u(x,y,t)dxdy$. and q is the strong solution.

Recalling equation (3), the strong solution has the probabilistic interpretation

$$\begin{aligned} q(x,y,t) &= \Pr\left(X(t) \in dx, Y(t) \in dy, a_x < X(t') < b_x, a_y < Y(t') < b_y | 0 < t' \le t, X(0) = x_0, Y(0) = y_0, \theta\right) \\ &= \Pr\left(X(t) \in dx, Y(t) \in dy, \min_{t'} X(t') \ge a_x, \max_{t'} X(t') \le b_x, \min_{t'} Y(t') \ge a_y, \max_{t'} Y(t') \le b_y \,\middle|\, 0 < t' \le t, X(0) = x_0, Y(0) = y_0, \theta\right) \end{aligned}$$

□

Lemma 2. *The maximum likelihood estimator is consistent as $n \to \infty$ and $k \to \infty$:*

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

.

Proof. **It's a bit sloppy. Also, we do not need asymptotic efficiency to prove equation (11)** By Lemma 1

$$X_k \xrightarrow{d} X \text{ as } k \to \infty.$$

Next, given Theorem 4.1 in Singler [2008], 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 Casella and Berger [2002] to guarantee that, for data $X_k \sim F_k(\theta)$,

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

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

$$\Pr_X\left(\left|\hat{\theta}_{n,k}(X) - \theta\right| \geq \varepsilon\right) \leq \frac{\mathrm{E}_X\left[(\hat{\theta}_{n,k}(X) - \theta)^2\right]}{\varepsilon^2}.$$

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,

$$\mathrm{E}_{X_k}\left[(\hat{\theta}_{n,k}(X_k) - \theta)^2\right] \to \mathrm{E}_X\left[(\hat{\theta}_{n,k}(X) - \theta)^2\right] \text{ as } k \to \infty$$

by the portmanteau lemma. Finally, we can show that

$$\mathrm{E}_{X_k}\left[(\hat{\theta}_{n,k}(X_k) - \theta)^2\right] \to 0 \text{ as } n \to \infty, \tag{11}$$

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

$$\Pr_X\left(\left|\hat{\theta}_{n,k}(X) - \theta\right| \geq \varepsilon\right) \leq \frac{\mathrm{E}_X\left[(\hat{\theta}_{n,k}(X) - \theta)^2\right]}{\varepsilon^2} < \delta$$

□

References

George Casella and Roger L Berger. *Statistical inference*, volume 2. Duxbury Pacific Grove, CA, 2002.

Wai-Ki Ching, Jia-Wen Gu, and Harry Zheng. On correlated defaults and incomplete information. *arXiv preprint arXiv:1409.1393*, 2014.

Helen Haworth, Christoph Reisinger, and William Shaw. Modelling bonds and credit default swaps using a structural model with contagion. *Quantitative Finance*, 8(7):669–680, 2008.

Hua He, William P Keirstead, and Joachim Rebholz. Double lookbacks. *Mathematical Finance*, 8(3):201–228, 1998.

Steven Kou, Haowen Zhong, et al. First-passage times of two-dimensional brownian motion. *Advances in Applied Probability*, 48(4):1045–1060, 2016.

Bernt Oksendal. *Stochastic differential equations: an introduction with applications*. Springer Science & Business Media, 2013.

Hannes Risken. *The Fokker-Planck Equation: Methods of Solution and Applications*. Springer-Verlag, 1989.

Laura Sacerdote, Massimiliano Tamborrino, and Cristina Zucca. First passage times of two-dimensional correlated processes: Analytical results for the wiener process and a numerical method for diffusion processes. *Journal of Computational and Applied Mathematics*, 296:275–292, 2016.

Sandro Salsa. *Partial differential equations in action: from modelling to theory*, volume 99. Springer, 2016.

Vladimir Viktorovich Shaidurov. *Multigrid methods for finite elements*, volume 318. Springer Science & Business Media, 2013.

John R Singler. Differentiability with respect to parameters of weak solutions of linear parabolic equations. *Mathematical and Computer Modelling*, 47(3):422–430, 2008.

Ivar Stakgold and Michael J Holst. *Green's functions and boundary value problems*, volume 99. John Wiley & Sons, 2011.