From Ornstein–Uhlenbeck to circular diffusions

Eduardo García-Portugués • Department of Statistics, UC3M

A (time-homogeneous) diffusion process $\{X_t\}$ is a stochastic process in continuous time and continuous state space that satisfies the Stochastic Differential Equation (SDE)

$$\mathrm{d}X_t = \mu(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t, \tag{1}$$

where μ and σ are the drift and volatility functions, respectively, and $\{W_t\}$ is a standard Brownian motion. The drift and volatility functions typically depend on a parameter $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^p$. For example, the famous Ornstein–Uhlenbeck (OU) process is a particular case of the diffusion process (1) with $\mu_{\boldsymbol{\theta}}(x) = \alpha(b - x)$ and $\sigma_{\boldsymbol{\theta}}(x) = s$ for $x \in \mathbb{R}$, and $\boldsymbol{\theta} = (\alpha, b, s) \in \mathbb{R}^+ \times \mathbb{R} \times \mathbb{R}^+$. The OU process is a mean-reverting process, which means that it tends to return to its mean value. The left plot of Figure 1 shows some simulations of the OU process. The OU process is widely used in finance, biology, and physics, among other fields.

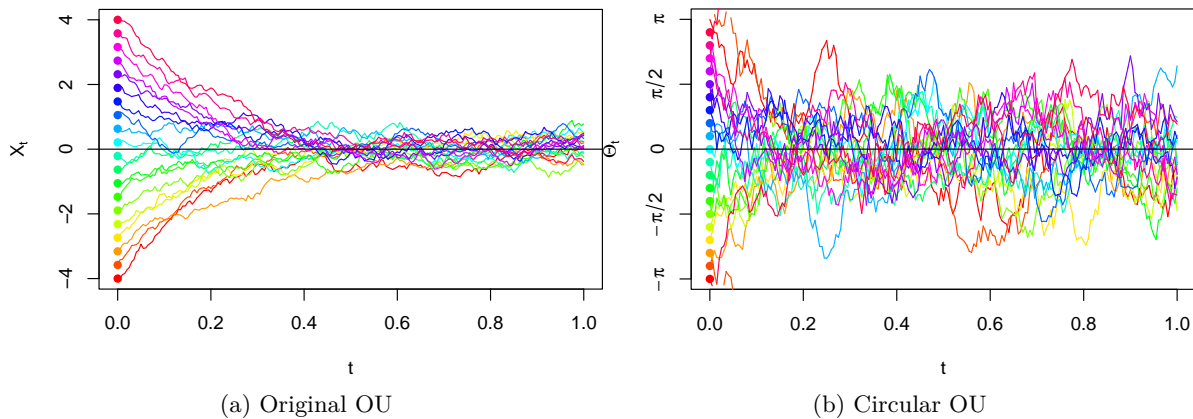(a) Original OU   (b) Circular OU

Figure 1: Simulated paths of the OU

One of the reasons why the OU is so popular is because it is relatively easy to estimate its parameters using maximum likelihood on a discretized sample path $\{X_{i\Delta}\}_{i=0}^n$, with $\Delta > 0$. In the general case (1), the maximum likelihood estimation of $\boldsymbol{\theta}$ is a very challenging problem because each of the n addends of the log-likelihood function is only known as the solution to a Partial Differential Equation (PDE) called the Fokker–Planck equation.

The OU process has been adapted to the circle $\mathbb{T}^1 = [-\pi, \pi)$, with $-\pi$ and π identified, by using the von Mises distribution. The right of Figure Figure 1 shows some simulated paths of this "circular" OU process, with paths "crossing" the boundary $-\pi \equiv \pi$. Unfortunately, this process is not nearly as

tractable as the original OU process because the Fokker–Planck equation does not have a closed-form solution. Circular diffusion processes have applications in biology (e.g., modeling the movement of ants), while their extensions to the torus $\mathbb{T}^2 = [-\pi, \pi) \times [-\pi, \pi)$ have applications in protein modeling (García-Portugués et al., 2019; García-Portugués & Sørensen, 2024; Golden et al., 2017).

This thesis proposal intends to exploit the very recent advances in the construction of diffusions in the torus with closed-form likelihood given in García-Portugués & Sørensen (2024) to create a new class of diffusion models on the circle (and potentially torus). This new class is to be more flexible than previous ones, while still retaining attractive properties such as known likelihood function, stationarity, and time-reversibility. Simulations will be used to explore the different dynamics that the class can model and to verify that maximum likelihood estimation works as intended. Real data applications are possible to a dataset of ant movements, and to the modeling of the evolution of a dihedral angle in molecular dynamics simulations.

References

García-Portugués, E., & Sørensen, M. (2024). A family of toroidal diffusions with exact likelihood inference. *arXiv:2409.02705*. https://doi.org/10.48550/arXiv.2409.02705

García-Portugués, E., Sørensen, M., Mardia, K. V., & Hamelryck, T. (2019). Langevin diffusions on the torus: Estimation and applications. *Statistics and Computing*, *29*(1), 1–22. https://doi.org/10.1007/s11222-017-9790-2

Golden, M., García-Portugués, E., Sørensen, M., Mardia, K. V., Hamelryck, T., & Hein, J. (2017). A generative angular model of protein structure evolution. *Molecular Biology and Evolution*, *34*(8), 2085–2100. https://doi.org/10.1093/molbev/msx137