

---

# Inference and Sampling of Point Processes from Diffusion Excursions (Supplementary Material)

---

Ali Hasan<sup>1,2</sup>   Yu Chen<sup>2</sup>   Yuting Ng<sup>1</sup>   Mohamed Abdelghani<sup>3</sup>   Anderson Schneider<sup>2</sup>   Vahid Tarokh<sup>1</sup>

<sup>1</sup>Department of Electrical and Computer Engineering, Duke University, Durham, North Carolina, USA

<sup>2</sup>Machine Learning Research, Morgan Stanley

<sup>3</sup>Department of Mathematics, University of Alberta, Edmonton, Alberta, Canada

We provide some technical results that supplement the main results in the text, including the calculation of the excursion length density, discussion about the connections between diffusion-based and intensity-based frameworks, and Lamperti transformation for formulating a constant variance diffusion process. For all analysis, we will continue to work with the SDE given by

$$dZ_t = \mu(Z_t, t)dt + \sigma(t)dW_t. \quad (1)$$

## A PROOFS

### A.1 PROOF OF EXCURSION LENGTH DENSITY VIA CHANGE OF MEASURE

We restate the main proposition here:

**Proposition 1** (Diffusion Excursion Density). *Let  $Z_t$  satisfy an SDE with drift  $\mu$  such that  $Z_t$  is recurrent at zero. Then the density of the excursion lengths  $\tau$  of  $Z_t$  is given by:*

$$p_Z(\tau) = p_e(\tau; \delta) \mathbb{E}_{\mathbb{Q}_\dagger^\delta} \left[ \exp \left( \int_0^\tau \mu(e_t, t; \theta) de_t - \frac{1}{2} \int_0^\tau \mu^2(e_t, t; \theta) dt \right) \right]. \quad (2)$$

*Proof.* The proof follows a straightforward change of measure argument. Performing the change of measure, we can write the Radon Nikodym derivative as  $\frac{d\mathbb{P}^\mu}{d\mathbb{Q}_\dagger^\delta}$  where  $\mathbb{Q}_\dagger$  is the excursion length density;  $\mathbb{Q}_\dagger^\delta$  is the corresponding length density of  $\delta$ -excursions. Under Novikov's condition, we can write the exponential martingale as

$$\frac{d\mathbb{P}^\mu}{d\mathbb{Q}_\dagger^\delta} = \exp \left( \int_0^{\tau \wedge T} \mu(e_t, t) de_t - \frac{1}{2} \int_0^{\tau \wedge T} \mu(e_t, t)^2 dt \right). \quad (3)$$

Now we have that

$$\begin{aligned} \mathbb{P}^{\mu, \delta}(\tau \in dt; \delta) &= \mathbb{E}_{\mathbb{Q}_\dagger^\delta} \left[ \frac{d\mathbb{P}^\mu}{d\mathbb{Q}_\dagger^\delta} \mathbb{1}(\tau \in dt) \right] \\ &= \mathbb{E}_{\mathbb{Q}_\dagger^\delta} \left[ \exp \left( \int_0^{\tau \wedge T} \mu(e_t, t) de_t - \frac{1}{2} \int_0^{\tau \wedge T} \mu(e_t, t)^2 dt \right) \mathbb{1}(\tau \in dt) \right] \\ &= \mathbb{E}_{\mathbb{Q}_\dagger^\delta} \left[ \exp \left( \int_0^{\tau \wedge T} \mu(e_t, t) de_t - \frac{1}{2} \int_0^{\tau \wedge T} \mu(e_t, t)^2 dt \right) \right] p_e(\tau; \delta) \end{aligned} \quad (4)$$

with the  $e_t$  being excursions of length  $\tau$ . This results in the final PDF of  $\tau$  as

$$p_Z(\tau; \delta) = p_e(\tau; \delta) \mathbb{E}_{\mathbb{Q}_\dagger^\delta} \left[ \exp \left( \int_0^\tau \mu(e_t, t; \theta) de_t - \frac{1}{2} \int_0^\tau \mu^2(e_t, t; \theta) dt \right) \right]. \quad (5)$$

□

## A.2 PROOF OF APPROXIMATION DISTANCE

*Proof.* First, by Gibbs and Su [2002, Theorem 4] note that the 1-Wasserstein distance between measures with support on  $[0, T]$  is bounded by the TV distance according to

$$W_1(p_\star, p_0) \leq T \frac{1}{2} \|p_\star - p_0\|.$$

Now using Pinsker's inequality, we can bound the TV distance in terms of the log of the Radon-Nikodym derivative.

$$T \|p_\star - p_0\| \leq T \sqrt{\frac{1}{2} \int dp_0 \left[ \int_0^T \mu dX_t - \frac{1}{2} \int_0^T \mu^2 dt \right]}.$$

Taking the supremum over all functions within the function space that satisfies the positive recurrence condition leads to the result.  $\square$

## B DETAILS OF EXCURSION LENGTH DENSITY GIVEN BY LAPLACE TRANSFORM

State-dependent excursion densities can be calculated by inverting the Laplace transform of the solution of a differential equation. In practice, this is difficult since the inverse Laplace transform is numerically unstable and intractable in general. We do not follow this approach in the paper but state the details of the approach here for completeness. We define the procedure in the following proposition based on Pitman and Yor [1999]:

**Proposition 2** (Pitman and Yor [1999]). *Let  $Z_t$  be the solution to the SDE in (1). Then define the speed and scale measures*

$$s'(x) = \exp\left(-\int_0^x 2\mu(u)/\sigma(u)du\right), \quad m(x) = \frac{2}{\sigma(x)s'(x)}. \quad (6)$$

Additionally, define the operator  $\mathcal{A}$  as

$$\mathcal{A} := \frac{1}{m(x)} \frac{d}{dx} \frac{1}{s'(x)} \frac{d}{dx}.$$

Then the Laplace transform of the hitting time distribution is given by solving the eigenvalue problem of

$$\mathcal{A}\phi(x) = \lambda\phi(x)$$

for the function  $\phi(x)$  with Wiener-Hopf factorization given by  $\phi_\pm(x)$ :

$$\mathbb{E}_x \left[ e^{(-\lambda\tau(Z_t))} \right] = \begin{cases} \phi_{\lambda,-}(x_0)/\phi_{\lambda,-}(x_1) & x_0 < x_1 \\ \phi_{\lambda,+}(x_0)/\phi_{\lambda,+}(x_1) & x_0 > x_1 \end{cases} \quad (7)$$

Then, using the strong Markov property, the total duration of the excursion from  $x_0$  to  $x_1$  and  $x_1$  to  $x_0$  (assuming  $x_1 > x_0$ ) is given by

$$\mathbb{E} \left[ e^{(-\lambda(\tau)(Z_t))} \right] = \frac{\phi_{\lambda,-}(x_0) \phi_{\lambda,+}(x_1)}{\phi_{\lambda,-}(x_1) \phi_{\lambda,+}(x_0)}. \quad (8)$$

Notably, the inversion of the Laplace transform numerically is unstable, leading to difficulties in recovering the density from an optimized  $\mu$  using this technique.

As the drift and the variance terms in the SDE defined above are only state-dependent, it is not straightforward to depict the general Ito process as in Eq 1. For the forward problem, a remedy is: first, start with deriving the first-hitting-time property of a simple SDE with only state-dependent drift and variance such as Brownian motion, linear drift Brownian motion, Bessel process, OU process [Borodin and Salminen, 1997, Part II], so that the eigenvalues and eigenfunctions can be solved easily with closed-form; then apply the Girsanov's theorem to extend the property for more general SDE. But for the backward problem, the Laplace transformation of the hitting time density does not usually have closed-form except for a few cases where the Laplace transform of the first-hitting-time density can be calculated in closed-form and can be decomposed into increasing eigenfunction and decreasing eigenfunction as in (7), solving SDE drift or variance is thus not straightforward.

Here we present one example of solving the drift function through the Laplace transform for the first-hitting-time problem (not yet the excursion problem). The example illustrates the difficulty of formulating the diffusion process via the Laplace transform from the modeling perspective as opposed to the proposed likelihood strategy in the main text. We use the OU process for the forward problem, which is already well-known. Next, we focus on the backward problem, i.e. recovering the drift of an SDE given the first-hitting-time density. Let the process start from  $x_0$  and the hitting boundary be a constant  $\alpha$ . The target first-hitting-time density is

$$p(t) = \frac{2|\alpha - x_0|}{\sqrt{2\pi}} (e^{2t} - 1)^{-\frac{3}{2}} \exp \left\{ 2t - \frac{(\alpha - x_0)^2}{2e^{2t} - 2} \right\}. \quad (9)$$

Next, we solve the corresponding SDE by setting  $a(x) = 1$  with constant variance. The Laplace transform of the density is,

$$\mathbb{E}_{x_0}[\exp(-\lambda H_\alpha)] = \begin{cases} \frac{\exp(x_0^2) D_{-\lambda}(-\sqrt{2}x_0)}{\exp(\alpha^2) D_{-\lambda}(-\sqrt{2}\alpha)} = \frac{\Phi_{-, \lambda}(x_0)}{\Phi_{-, \lambda}(\alpha)}, & \alpha > x_0 \\ \frac{\exp(x_0^2) D_{-\lambda}(\sqrt{2}x_0)}{\exp(\alpha^2) D_{-\lambda}(\sqrt{2}\alpha)} = \frac{\Phi_{+, \lambda}(x_0)}{\Phi_{+, \lambda}(\alpha)}, & x_0 > \alpha \end{cases} \quad (10)$$

where

$$D_{-\lambda}(x) := \frac{1}{\Gamma(\lambda)} e^{-\frac{x^2}{4}} \int_0^\infty t^{\lambda-1} e^{-xt - \frac{t^2}{2}} dt$$

which has the following properties

$$\begin{aligned} \frac{dD_{-\lambda}(x)}{dx} &= -\frac{x}{2} D_{-\lambda}(x) - \lambda D_{-\lambda-1}(x) = \frac{x}{2} D_{-\lambda}(x) - D_{-\lambda+1}(x) \\ \lambda D_{-\lambda-1}(x) &= -x D_{-\lambda}(x) + D_{-\lambda+1}(x) \\ \frac{d}{dx} e^{x^2} D_{-\lambda}(\sqrt{2}x) &= 2x e^{x^2} D_{-\lambda}(\sqrt{2}x) + e^{x^2} D'_{-\lambda}(\sqrt{2}x) \\ &= 2x e^{x^2} D_{-\lambda}(\sqrt{2}x) + \sqrt{2} e^{x^2} \left( -\frac{\sqrt{2}x}{2} D_{-\lambda}(\sqrt{2}x) - \lambda D_{-\lambda-1}(\sqrt{2}x) \right) \\ &= -\sqrt{2} \lambda e^{x^2} D_{-\lambda-1}(\sqrt{2}x) \\ \frac{d^2}{dx^2} e^{x^2} D_{-\lambda}(\sqrt{2}x) &= -\sqrt{2} \lambda \frac{d}{dx} e^{x^2} D_{-\lambda-1}(\sqrt{2}x) = 2\lambda(\lambda+1) e^{x^2} D_{-\lambda-2}(\sqrt{2}x) \end{aligned}$$

According to Eq (7) and (10), the corresponding function can be defined as,

$$\Phi_{+, \lambda}(x) := e^{x^2} D_{-\lambda}(\sqrt{2}x). \quad (11)$$

As defined above,  $D_{-\lambda}(x) > 0$ , and the derivative of  $\Phi_{+, \lambda}(x)$  is negative, so the function is decreasing satisfying the Laplace transformation decomposition. Taking this back to the infinitesimal generator eigenfunction in Eq 6, we have the equality below for  $\lambda \neq 0$

$$\begin{aligned} \lambda(\lambda+1) e^{x^2} D_{-\lambda-2}(\sqrt{2}x) - b(x) \sqrt{2} \lambda e^{x^2} D_{-\lambda-1}(\sqrt{2}x) &= \lambda e^{x^2} D_{-\lambda}(\sqrt{2}x) \\ \iff -\sqrt{2}x D_{-\lambda-1}(\sqrt{2}x) + D_{-\lambda}(\sqrt{2}x) - b(x) \sqrt{2} D_{-\lambda-1}(\sqrt{2}x) &= D_{-\lambda}(\sqrt{2}x). \end{aligned}$$

So the solution is  $b(x) = -x$ , and the SDE satisfies,

$$dZ_t = -Z_t dt + dW_t \quad (12)$$

which is a standard OU. This is a simple example that the Laplace transform can be derived with closed-form and the solution is achievable and is only state-dependent. In general, such convenience is not guaranteed and there may not have a state-dependent solution.

## C EXPERIMENTAL DETAILS

### C.1 DRIFT RECOVERY

The list of drifts  $\mu$  for the multi-dimensional drift recovery experiments is provided in Table 1. We first simulate the  $d$ -dimensional diffusion process then compute the times when each dimension has an excursion from the corresponding

Experiment	$\mu(X_t)$	
Cubic	$= -X_t^3$	$X_t \in \mathbb{R}^{10}$
Circle	$= \begin{cases} -X_{1,t} - X_{2,t} \\ -X_{2,t} + 5X_{1,t} \end{cases}$	$X_t \in \mathbb{R}^2$
Tanh	$= -\tanh(X_t)$	$X_t \in \mathbb{R}^{10}$
OU	$= -X_t$	$X_t \in \mathbb{R}^5$

Table 1: Table of drifts for multi-dimensional experiments.

axis. The parameter  $\delta$  is chosen to be zero, but due to discretization in simulation, an infinite number of excursions is not observed. We use  $\Delta_t = 0.01$  and simulate until terminal time  $T = 10$ . The time points where the diffusion crosses the axis are recorded. For the excursion estimator, we consider both positive and negative excursions which we simulate using a Bernoulli random variable with a probability of 0.5. We compare this to the bridge estimator that interpolates between zeros using Brownian bridges. The hyperparameters for both models are equal to ensure a fair comparison. We train for 200 epochs, with a learning rate of  $1 \times 10^{-3}$  using the Adam optimizer. The architecture is a 64 width, 6 depth multi-layer perceptron with `Softplus` activation function.

**Kernel Coefficient Recovery for Exogenous and History Dependent Drift** For these experiments, we consider an OU process influenced by either an exogenous process or the history by defining the drift as  $\mu = -x + w\varphi(\cdot)$  with  $\varphi(z) = \exp(-z/\sigma)$ . We choose  $\sigma = 1$  for the exogenous dependence and  $\sigma = 2$  for the history dependence. For the both experiments, we simulate using Euler-Maruyama with  $\Delta t = 0.05$  up to  $T = 50$  and observe the zero times. The exogenous signal is generated according to 100 samples from a uniform random variable as  $\mathcal{U}(0, 40)$  and sorted by the value. For these experiments, we only optimize over the estimated parameter  $\hat{w}$  due to the lack of identifiability if both components are unknown. We use the same parameters and only compare the standard SDE regression using Brownian bridges and the excursion based approach. The models are trained using the `AdamW` optimizer with a learning rate of  $1.0 \times 10^{-2}$ . Figure 3 illustrates a sample path with dependence on the times when it reaches zero.

## C.2 RENEWAL PROCESSES

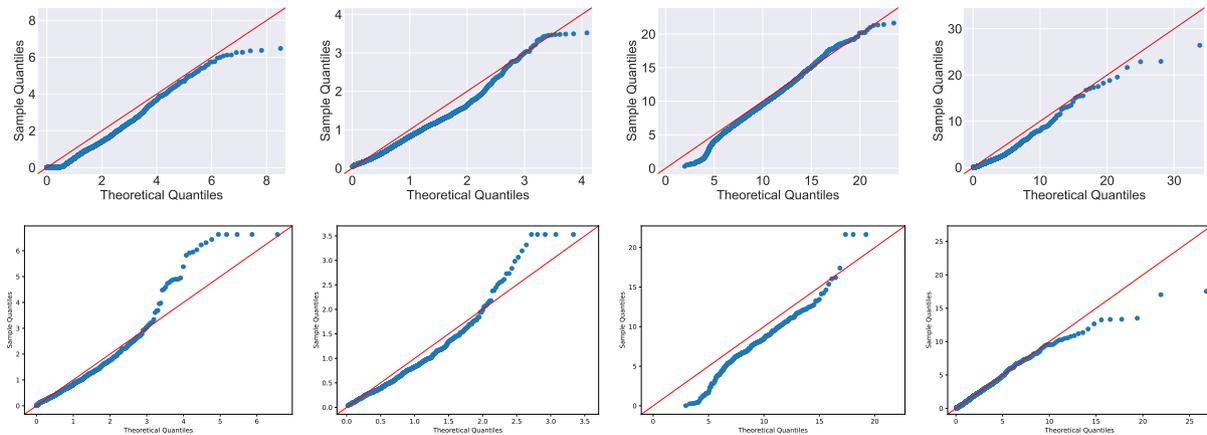


Figure 1: Comparison of QQ plots for Poisson process, Weibull, Gamma, and log-normal renewal processes. The top row are results from our excursion-based estimator. The bottom row are results from an intensity-based estimator. All are fit with the same architecture, optimization parameters and 200 samples from the specified renewal process.

We consider four canonical renewal processes with parameters described in Table 2. The architecture is a basic MLP with width 16 and 6 layers and the `Softplus` activation function. They were trained for 2000 epochs using the `AdamW` optimizer with learning rate of  $1 \times 10^{-3}$ . The learning rate for the  $\delta$  parameter was  $1 \times 10^{-2}$ . For reference on how the proposed estimator is performing compared to a standard deep learning-based point process estimator, we provide Figure 1 which compares the performance of an intensity-based model Shchur et al. [2019] to our excursion-based model.

Distribution	Parameters
Exponential	$\lambda = 1$
Gamma	$\alpha = 9, \beta = 1$
Log-Normal	$\mu = 0, \sigma = 1$
Weibull	$\lambda = 1, k = 1.5$

Table 2: Renewal distributions with the corresponding parameters for the experiments in main text.

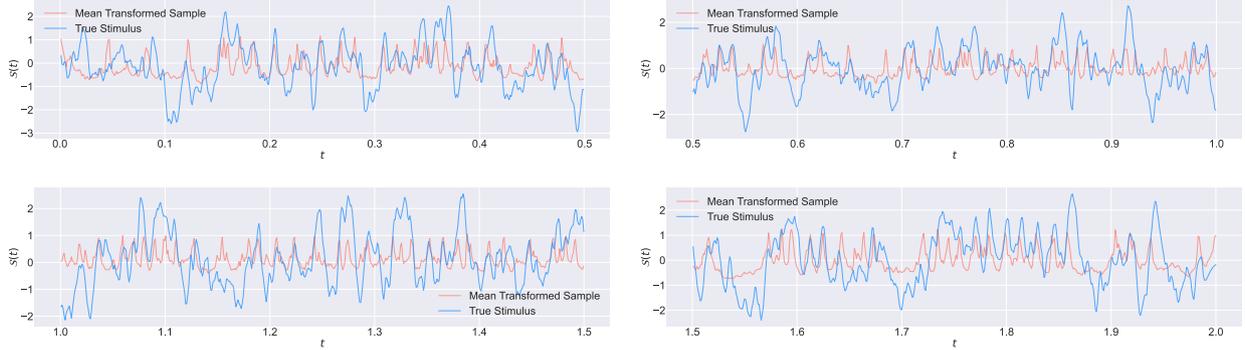


Figure 2: Scaled learned stimulus for all time points.

### C.3 REAL DATA

For the real data experiment, we consider a modification of the neural network that includes a positional encoding layer. This allows the network to learn higher frequency functions [Tancik et al., 2020]. We use a LeakyReLU activation function.

**Dataset Description** The neuroscience dataset is the same as [Tripathy et al., 2013], which is composed of *in vitro* whole-cell patch clamp recordings of mitral cells from mouse olfactory bulb slices. Spikes were recorded with 100 trials of a 2 seconds duration by simulating the neuron with repeated frozen noise current, which is generated by convolving a white-noise with an alpha function with  $\tau = 3$  ms.

**Transforming the Learned Stimulus** Since the proposed method models the continuous as excursions, we invoke a transformation such that the zeros are peaks. In particular, we compute  $\tilde{Z}_t = a(\mathbb{E}[\log Z_t]) + b$  where  $a, b$  are found using least squares. Note that due to the log, this will result in  $\tilde{Z}_t$  being either positive or negative. In that sense, we cannot recover the sign of the excursion unless we're provided additional information. However, the main components we want to match are the peaks of the true stimulus and the learned stimulus. Figure 2 shows that this is effectively accomplished.

## D FURTHER MODELING CONSIDERATIONS

### D.1 INCORPORATING HISTORY

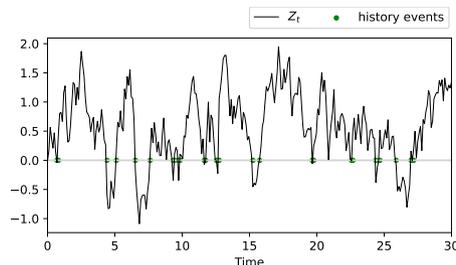


Figure 3: Example of a sample path affected by its history.

In practice, many point processes depend on the distribution of the history for a certain window. We can include this in the model by making the drift dependent on the history by, for example, considering the structure of the drift to be a recurrent neural network. Then the likelihood becomes a function of the history with the change of measure being the same as in (4). Specifically, we can change the drift function from  $\mu(Z_t, t)$  in (1) to  $\mu(Z_t, t, \mathcal{H}_t)$ , where  $\mathcal{H}_t$  represents the history of the process up to time  $t$ .  $\mathcal{H}_t$  can be the history of the whole path of  $Z_t$ , or it can be the history of discrete events  $\{s \mid dL^f(s) = 1, s < t\}$ , or subset of the history events. The challenge of the model extension raises from more complicated modeling of the drift function. We leave such extension as the future work.

To illustrate the point, here we consider a special example of history-dependent SDE, which mimics the intensity function of the Hawkes process,

$$dZ_t = \mu(t, \mathcal{H}_t)dt + dW_t \quad (13)$$

$$\mu(t, \mathcal{H}_t) = \mu_0 t + h(t - t_{h,1}) + h(t - t_{h,2}) \quad (14)$$

where  $\mu_0 t$  is the baseline drift with a constant slope, and the drift function depends on the last event  $t_{h,1}$  and the last second event  $t_{h,2}$ , so the diffusion process is history-dependent.  $h(\cdot)$  is the kernel function describing how past events affect the future drift. In the example, the process only depends on the past two events before  $t$ . For example, if  $h(t) = e^t \cdot \mathbb{I}(t \geq 0)$ , the past events will have a positive influence on the drift, and such influence decays as the events stay further back. As shown in the examples in Appendix F, the linear Hawkes process is not equivalent to this problem, even though both of them consider the additive history effects. An example of a sample path exhibiting this behavior is given in Figure 3.

## D.2 CONNECTION TO MORE GENERAL POINT PROCESSES

In this section, we illustrate the modeling connection between the diffusion-based framework and the intensity-based framework. The bridge between the two groups of methods is the conditional density of the next event given the history up to the current time. For the intensity-based model, the conditional density of the next event is [Daley and Vere-Jones, 2003],

$$g(t \mid t_1, \dots, t_{n-1}) = \lambda(t \mid t_1, \dots, t_{n-1}) \times \exp \left\{ - \int_{t_{n-1}}^t \lambda(u \mid t_1, \dots, t_{n-1}) du \right\}$$

for  $t \geq t_{n-1}$  and where  $\lambda$  is the conditional intensity function,  $t_i$  are timestamps,  $g$  is the density of the next event. Reversely, given the density  $g$ , the intensity function can also be derived. See Appendix F. For the diffusion-based model, suppose the point process is generated by an underlying SDE, where the events are a sequence of first-hitting times that reset the process to the initial level after every hit. Let  $g(\tau), \tau \in \mathbb{R}_+$  be the density of the next event, assuming the current event occurs at  $t = 0$  ( $g(\tau)$  can be different for every hitting time). For the next event, there exists an SDE with a time-varying drift function  $\mu(t)$  such that its first-hitting-time density is  $g$  (again,  $\mu(t)$  can be different for every hitting time). The equivalent model is the driftless diffusion, such as Brownian motion, but with a time-varying boundary (the boundary can be different for every hitting time). If the drift function or the boundary depends on the history, the model will include history dependency. We leave more detailed discussions and numerical examples, such as the renewal process and Hawkes process, in Appendix F.

## D.3 INTERPRETING THE LEARNED DRIFT

A byproduct of the proposed method is a parameterization of a diffusion that generates the data. In particular, the drift that is recovered can be subsequently analyzed using traditional methods from stochastic calculus. In general, the method should be used in cases where such interpretation is useful, such as in health care or in finance. The drift may provide clinically useful insight into the distribution of action potentials for a diseased versus healthy patient. The drift provides an understanding at a multiscale level, leading to potential therapies that correspond to modulating the drift function.

In a similar example as described in the introduction, bursty transcription could be investigated again at a finer, molecular level. The drift can correspond to some production rate or movement of the underlying molecules which can again lead to potential pathways for developing medications.

Finally, in a financial setting, arrivals of aks orders in a market may be related to excursions of the drawdown process of the perceived fair price. As the fair price reaches the running maximum, the drawdown process reaches zero, signaling an appropriate time to sell the financial instrument. Developing and analyzing the drift of the latent fair price could lead to better risk management for market makers or more effective trading strategies.

## E EXAMPLE APPLICATIONS

To motivate the proposed model, we include a few examples where the method may be appropriate and an intuitive interpretation is present. We study some of these experiments in greater detail by considering the representational capabilities of the method in describing the data.

**Fair Pricing from Bids and Asks of Illiquid Assets.** Suppose we observe sets of point processes for a given asset on an exchange. Denote the set of samples associated to bids as  $\mathbb{B} = \{t_i^{(b)}, b_i\}_{i=1}^N$  and the process for asks as  $\mathbb{A} = \{t_i^{(a)}, a_i\}_{i=1}^M$ . We assume the following properties of observations based on a latent fair price,  $Z_t$ :

1. Bids are generated when  $Z_t < \mathbb{E}\left[Z_t \mid Z_{t_i^{(b)}} = b_i\right] - \delta$  for  $t_i^{(b)} < t$ . That is, the fair price should not exceed the expected fair price following the last bid.
2. The fair price does not cross the expectation of the diffusion bridge between any two arrivals.

$Z_t$  satisfies an SDE with unknown drift  $\mu(Z_t, t)$ , we consider the excursions above and below the curves:

$$\begin{aligned} f_{\text{bid}} &= \mathbb{E}[Z_t \mid Z_{t_i^{(b)}} = b_i] & f_{\text{ask}} &= \mathbb{E}[Z_t \mid Z_{t_i^{(a)}} = a_i] \\ &= \int_{t_i^{(b)}}^t \mu(z, t) dt + b_i & &= \int_{t_i^{(a)}}^t \mu(z, t) dt + a_i \end{aligned}$$

Specifically, since  $a_i \geq b_i$ ,  $f_{\text{ask}}(t) \geq f_{\text{bid}}(t)$  for all  $t > 0$ . Qualitatively, the model suggests that a *new bid* occurs when the fair price exceeds the expected value conditioned on starting at the *last bid* price and the fair price  $Z_t$  being at least  $\delta > f_{\text{bid}}(t)$  within that interval with the opposite characterization for the process involving asks. This is characterized by a  $\delta$ -excursion above  $f_{\text{bid}}$  or below  $f_{\text{ask}}$ , leading to the interarrival time being the excursion length of the fair price above or below the curve. We illustrate this behavior and the generated point process in Figure 4.

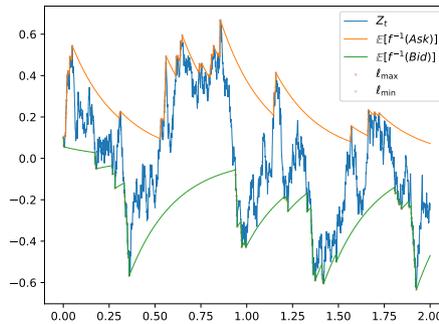


Figure 4: Examples of bids and asks generated according to the intersection of the fair price (blue) with the expected price from the last bid (green) and ask (orange).

**Heart Rate Variability** In Barbieri et al. [2005], the authors consider a renewal process for modeling heart rate variability based on a drifted Brownian motion. We consider a similar model but based on excursions reaching a minimum height. This allows us to make an interpretation of the latent path being a continuous path instead of discontinuous as in the first hitting time case. The continuous path should resemble the ensemble of cardiac action potentials that generate the electrical signal whose peaks are the arrival times we observe. Analyzing the drift that governs the excursions can then be used for qualitative reasoning about the distributions of action potentials.

**EDA Data** On another front considering electrophysiology, Subramanian et al. [2020] describes a point process structure that describes the spiking properties in electrodermal activity (EDA) data. The authors posit the existence of a latent Brownian motion with first crossing times that are distributed according to the observed data. The Brownian motion is seen to provide a description of the underlying electrophysiology related to the release of sweat that in turn generates the points in

the point process. We generalize this argument by considering a series of excursions above a threshold limit. This provides a continuous description relating to the release of sweat that generates the arrival of the point process. The data are in the form of the arrival time of the pulse times derived from EDA data for the control group of patients (EDA), and for the patients under the influence of the drug Propofol (EDA-P).

## F CONNECTION BETWEEN DIFFUSION-BASED MODEL AND INTENSITY-BASED MODEL

As introduced in section D.2, the conditional intensity of the next event relates the two groups of modeling methods of the diffusion-based method and the intensity-based method. Recall the relationship between the joint arrival time distribution and the intensity function is given by

$$\lambda(t_n|t_1, \dots, t_{n-1}) = \frac{p(t_n|t_1, \dots, t_{n-1})}{1 - \int_{t_{n-1}}^{t_n} p(s|t_1, \dots, t_{n-1})ds}.$$

Additionally, we have that  $p(t_n|t_{i < n}) = p(t_n, \dots, t_1)/p(t_{n-1}, \dots, t_1)$ . Numerically, when considering the proposed estimator, we must compute the ratio of two expectations given our observations. The variance of this becomes a bit high, but we can use common random numbers to reduce this. That is, we compute the same set of bridges for both expectations. This leads to only computing the expectation of the martingale portion multiplied by the original measure. Specifically,

$$\lambda(t_n|t_1, \dots, t_{n-1}) = \lambda(t_n|t_{n-1}) = \frac{p(t_n - t_{n-1}|t_{n-1})}{1 - \int_{t_{n-1}}^{t_n} p(t_n - t_{n-1}|t_{n-1})}.$$

Of course, if  $\mu$  does not depend on  $t$  then the conditional term is not needed. We then approximate the integral in the denominator by a Riemann sum.

On the other hand, the density of the next event can be derived from the intensity function.

$$p_n(t|t_1, \dots, t_{n-1}) = \lambda(t|t_1, \dots, t_{n-1}) \exp \left\{ - \int_{t_{n-1}}^t \lambda(u|t_1, \dots, t_{n-1})du \right\}. \quad (15)$$

Next, we provide a few examples. Let  $p(x, t|x_{t_{(n-1)}}, 0)$  be the transition density of the dynamics  $X_t$  start from the origin,  $g(u|x_{t_{(n-1)}}, 0)$  be the density of the first hitting time of the process. Consider the random walk from  $(x_{t_{(n-1)}}, 0)$  to  $(x_t, t)$  through first hitting point  $(S(u), u)$ . By marginalizing out the first hitting time  $(S(u), u)$  (Chapman–Kolmogorov equation), we have

$$p(x, t|x_{t_{(n-1)}}, 0) = \int_0^t p(x, t|S(u), u) \cdot g(u|x_{t_{(n-1)}}, 0)du. \quad (16)$$

The above is called the Fortet equation, which can be used to solve the density of the first hitting time [Sacerdote and Giraudo, 2013].  $g(u|x_{t_{(n-1)}}, 0)$  depends on the design of  $S(u)$ . In general condition, there exists such a boundary [Potiron, 2021], which is proved using a piecewise linear representation that is also the technique used by [Sacerdote and Zucca, 2003]. The approximation accuracy can be arbitrarily high (the error is  $O(h^2)$  where  $h$  is the knot distance) [Zucca and Sacerdote, 2009].

Take the target density  $g$  into Eq (16) to solve the boundary function  $S$ . If  $g$  includes the history of the process, such as the Hawkes process, the solution  $S$  becomes a history-dependent and time-varying boundary function.  $p(x, t|x_{t_{(n-1)}}, 0)$  and  $p(x, t|S(u), u)$  can be easily obtained using the basic property of Brownian motion. [Sacerdote and Zucca, 2003] provides an approximation solution to the problem.

## G LAMPERTI TRANSFORM

In this section, we show that if  $\sigma$  is not a constant, the process can be transformed into an equivalent process with constant  $\sigma$ . So model in the main text assumes the constant drift  $\sigma = 1$  without loss of generality. The transformation is described in Ait-Sahalia [2002]. Let  $Z_t$  represent the original process satisfying (1). Let  $Y_t := \gamma(X_t, t) = \int \frac{dx}{\sigma(X_t, t)}$ . From Ito's lemma,

$$dY_t = \mu_Y(Y_t, t)dt + dW_t$$

$$\mu_Y(Y_t, t) = \frac{\partial \gamma}{\partial t} (\gamma^{-1}(Y_t, t), t) + \frac{\mu(\gamma^{-1}(Y_t, t))}{\sigma(\gamma^{-1}(Y_t, t))} - \frac{1}{2} \frac{\partial \sigma}{\partial t} (\gamma^{-1}(Y_t, t), t)$$

has unit diffusion. To overcome identifiability issues, we assume that this transformation implicitly occurs, and we only wish to recover the unit variance process  $Y_t$ .

The multivariate Lamperti transform is described by Aït-Sahalia [2008]. Let  $\sigma(y)$  be a symmetric positive matrix for all  $y \in \mathbb{R}^d$ . Additionally let  $\sigma$  be differentiable everywhere and for all  $y$

$$\frac{\partial \sigma(y)}{\partial y_k} \sigma(y)^{-1} e_j = \frac{\partial \sigma(y)}{\partial y_j} \sigma(y)^{-1} e_k$$

where  $e_j$  is the  $j^{\text{th}}$  canonical basis vector of  $\mathbb{R}^d$ . Then the SDE given by Lipschitz  $\mu, \sigma$  is a reducible SDE with unit diffusion.

## References

- Yacine Aït-Sahalia. Maximum likelihood estimation of discretely sampled diffusions: a closed-form approximation approach. *Econometrica*, 70(1):223–262, 2002.
- Yacine Aït-Sahalia. Closed-form likelihood expansions for multivariate diffusions. *The Annals of Statistics*, 36(2):906–937, 2008.
- Riccardo Barbieri, Eric C Matten, AbdulRasheed A Alabi, and Emery N Brown. A point-process model of human heartbeat intervals: new definitions of heart rate and heart rate variability. *American Journal of Physiology-Heart and Circulatory Physiology*, 288(1):H424–H435, 2005.
- Andrei N Borodin and Paavo Salminen. Handbook of brownian motion-facts and formulae. *Journal of the Royal Statistical Society-Series A Statistics in Society*, 160(3):596, 1997.
- Daryl J Daley and David Vere-Jones. *An introduction to the theory of point processes: volume I: elementary theory and methods*. Springer, 2003.
- Alison L Gibbs and Francis Edward Su. On choosing and bounding probability metrics. *International statistical review*, 70(3):419–435, 2002.
- Jim Pitman and Marc Yor. Laplace transforms related to excursions of a one-dimensional diffusion. *Bernoulli*, pages 249–255, 1999.
- Yoann Potiron. Existence in the inverse shiryayev problem. *arXiv preprint arXiv:2106.11573*, 2021.
- Laura Sacerdote and Maria Teresa Giraudo. Stochastic integrate and fire models: a review on mathematical methods and their applications. *Stochastic biomathematical models*, pages 99–148, 2013.
- Laura Sacerdote and Cristina Zucca. Threshold shape corresponding to a gamma firing distribution in an ornstein-uhlenbeck neuronal model. *Scientiae Mathematicae Japonicae*, 58(2):295–306, 2003.
- Oleksandr Shchur, Marin Biloš, and Stephan Günemann. Intensity-free learning of temporal point processes. *arXiv preprint arXiv:1909.12127*, 2019.
- Sandya Subramanian, Riccardo Barbieri, and Emery N Brown. Point process temporal structure characterizes electrodermal activity. *Proceedings of the National Academy of Sciences*, 117(42):26422–26428, 2020.
- Matthew Tancik, Pratul Srinivasan, Ben Mildenhall, Sara Fridovich-Keil, Nithin Raghavan, Utkarsh Singhal, Ravi Ramamoorthi, Jonathan Barron, and Ren Ng. Fourier features let networks learn high frequency functions in low dimensional domains. *Advances in Neural Information Processing Systems*, 33:7537–7547, 2020.
- Shreejoy J Tripathy, Krishnan Padmanabhan, Richard C Gerkin, and Nathaniel N Urban. Intermediate intrinsic diversity enhances neural population coding. *Proceedings of the National Academy of Sciences*, 110(20):8248–8253, 2013.
- Cristina Zucca and Laura Sacerdote. On the inverse first-passage-time problem for a wiener process. *The Annals of Applied Probability*, pages 1319–1346, 2009.