UNIVERSITÀ DEGLI STUDI DI PADOVA

Dipartimento di Fisica e Astronomia "Galileo Galilei"

Corso di Laurea Magistrale in Physics

Master's Thesis

Statistical Inference and Model Discrimination in Stochastic Dynamical Systems

Supervisor
Prof. Sandro Azaele

Candidate
Alice Bertelli

Co-supervisor
Dr. Javier Aguilar Sánchez

Academic Year 2025/2026

Contents

Introduction

Many current challenges, which directly influence the well-being and health of society - including climate change, epidemic spreading, and financial instability - can be effectively investigated through the quantitative frameworks provided by physics, mathematics, and statistics [1][2][3]. At the same time, these tools also play a central role in addressing fundamental open problems in science, such as ecological and microbial population dynamics, as well as gene expression and regulation mechanisms [4][5]. Specific events and dynamics can be predicted using data-driven methods or model-driven analysis. In the first case, one relies on a large amount of data, which is analyzed to build statistical models. This methodology offers great flexibility and requires little prior knowledge of the process under analysis; however, results are difficult to interpret and require very large datasets, which are not always available [6]. In contrast, the model-based method relies on the prior construction of a model that can efficiently describe the process under study and on the subsequent adaptation of the model to the available data. This requires a more theoretical approach and has limited application flexibility; however, it requires little data and offers excellent interpretability of results [7][8]. Often, the best choice depends on the specific case study and involves both methodologies [9].

In this thesis project, I will adopt a model-based approach, where data are used to refine the model. Specifically, I will focus on stochastic models, which represent a powerful tool for describing processes affected by uncertainty. Stochastic processes, indeed, can capture the randomness of the underlying dynamics in addition to a deterministic trend [10]. Unlike deterministic models, their dynamics is not entirely determined by initial conditions and takes into account variability and noise. These characteristics make them ideal for describing real-world processes, such as those mentioned above [2][5][4]. The strength of these models lies in their ability to represent the distribution of the possible outcomes of a process, making them useful for describing complex systems, characterized by interactions between different elements [7][8].
When combined with inference methods such as Maximum Likelihood Estimation (MLE), these models make it possible to analyze data within a probabilistic framework, characterizing its statistical properties, while preserving a mechanistic interpretation of the underlying processes [11].

By simultaneously employing stochastic modeling and statistical inference, one is thus able to efficiently characterize a dynamical system. However, there are still several limitations associated with this method of analysis.
First, the fact that different stochastic models fit the same dataset means that it is not always

clear which is the best stochastic description [12]. Note that the use of different models leads to different interpretations of the data.
Another problem is the estimation of the parameters that characterize stochastic models. The importance of correct parametric inference lies in the fact that model parameters encode the essential features of the underlying process: an approximate estimate of the parameters may lead to an erroneous interpretation of the dynamical system [11]. It is important to remember that the purpose of a good model and optimal parametric inference is to provide a faithful representation of the system while enabling robust prediction and interpretation [13][6][7].

Although the focus is on a model-based approach, it must be acknowledged that the study of a dynamical system must eventually involve adaptation to the available data [6]. Data collection is therefore a fundamental step and the way in which data are collected influences subsequent analysis, as it deeply changes the information they can provide [14]. As will be shown, both parametric inference and the ability to distinguish between two different stochastic processes depend critically on the observed sample, which is characterized by the number of points, the dataset size, and the time interval between the data points, commonly referred to as the sampling time interval [13][12].
Furthermore, data availability is often limited by several factors, including the cost of data collection, experimental constraints, the time required to acquire data, storage and computational resources, and restrictions imposed by privacy regulations (for example, in the case of health-related data) [15].
These limitations can make statistical analysis more difficult and less reliable [12]. The solution could be to optimize the dataset itself within the experimental limits. It will be shown that, assuming a fixed sample size, the dataset can be optimized by identifying a range for the sampling time interval that minimizes the parametric inference error or increases the probability of correctly identifying the origin of a dataset [12].

In light of this, the aim of this work will be to identify an optimal sampling protocol to ensure the efficiency of parametric inference and model discrimination for a stochastic dynamical system. In order to address and clarify these issues, the thesis is structured as follows.
The first part of this thesis reviews the framework developed in [12], where the problem of parametric inference and model discrimination for stochastic differential equations (SDEs) is studied. In the second part of the thesis, the problem of parametric inference is extended to the context of jump processes: the aim is to examine how the outcomes from the previous section apply in this context and to identify the differences and similarities in both the methods and the results. Then, the tools developed for inference and model distinguishability, in both of the previous sections, are applied to the analysis of two models of gene expression, introduced in [4], which will be described later. Finally, the last Chapter shows how the framework developed for jump processes can be exploited to study chemical reactions models.
Specifically, parametric inference will rely on Maximum Likelihood Estimation (MLE), while model distinguishability will be assessed by comparing the likelihoods of competing models

given the available data. Consequently, the evaluation of the likelihood represents a central challenge of this work. For both stochastic differential equations (SDEs) and jump processes, deriving the exact likelihood requires analytical knowledge of the process's propagator [12]. Since a closed-form solution is often unavailable, likelihood approximations must be introduced. A primary objective of this thesis is to detail a possible solution within the extended framework of jump processes and discuss its limitations at large sampling intervals.

Chapter 1

Inference and model discrimination in continuous stochastic processes

In this first chapter, we address parametric inference and model discrimination for continuous-state, continuous-time stochastic processes with a single degree of freedom, described by stochastic differential equations (SDEs).
Stochastic differential equations are the natural extension of ordinary differential equations and are used to describe systems subject to fluctuations. They are characterized by a drift term, which encodes the deterministic behavior, and a term that models the system's noise:

$$dX_t = A(X_t)\,dt + B(X_t)\,dW_t \tag{1.1}$$

where $A(\cdot)$ is the drift function, $B(\cdot)$ is the diffusion function and W_t denotes the Wiener process, which models Gaussian white noise with $\langle dW_t\, dW_s\rangle = \delta(t-s)\,dt$. This allows us to describe both the average behavior of the system, governed by the deterministic part of the equation, and its random fluctuations over time, captured in the diffusive term [8].

1.1 Stochasticity in population dynamics

SDEs are often used to describe population dynamics in ecological, socio-economic contexts and epidemiological dynamics. In ecology, the term 'population' refers to a group of individuals of the same species, while in other contexts like epidemics, refers to a group of individuals sharing a common condition, for example, related to health (e.g. the population of infected individuals) [16] [17]. In this context, the drift term takes on a specific meaning, namely describing the growth or decline of the population under study, and the diffusive term encodes demographic or environmental variability [12].
The noise characterizing a dynamic system can have a general dependence on the state and time $B(x,t)$, but we will focus on homogeneous systems where $A(x)$ and $B(x)$ are not explicitly dependent on time. The homogeneous noise function can be additive (i.e. $B(x) = D$ constant), where the noise intensity remains constant throughout the process, or multiplicative, where the diffusion function depends on the system's state itself (e.g. $B(x) = Dx^2$) [8].

An example of additive noise process can be represented by the Ornstein-Uhlenbeck (OU). This process is characterized by the diffusion coefficient D, which defines the intensity of the noise, the autocorrelation time (process relaxation time) $\tau = k^{-1} > 0$ and the mean value μ [18]:

$$dX_t = k(\mu - X_t)\,dt + D\,dW_t. \tag{1.2}$$

Although it is one of the simplest stochastic models, characterized by linear drift and constant diffusion, the OU model provides a highly accurate and mathematically tractable description of population dynamics. This local approximation is particularly effective when the initial state of the population, X_0, is close to the deterministic equilibrium ($X_0 \sim \mu$), allowing the model to capture essential diffusive fluctuations without the need for nonlinear terms [19]. Furthermore OU statistical properties can be computed analytically [10]. The probability density for the process to go from $X(0)$ to $X(t)$, known as propagator, is given by

$$p(X(t)|X(0) = x_0; \Theta) = \mathcal{N}(x;\, \mu + (x_0 - \mu)\, e^{-kt}, \frac{D^2}{2k}(1 - e^{-2kt})) \tag{1.3}$$

where $\mathcal{N}(x; \mu_t, \sigma_t^2)$ is the normal probability density function evaluated at time t and $\Theta = (k, \mu, D)^T$ [10]. Then, the stationary probability density function for the process is a Gaussian:

$$p(x) = \mathcal{N}(x; \mu, \frac{D^2}{2k}).$$

However, empirical population abundance distributions are usually asymmetric with respect to the mean, so that, at steady state, they are better described by Gamma distributions [12]. We therefore introduce a set of processes characterized by multiplicative fluctuations - such that the diffusion function depends on the state of the system - and by a stationary Gamma distribution. These processes are defined by the solutions to the family of SDEs given by

$$dX_t = kX_t^{\theta}(\mu - X_t)\,dt + DX_t^{\frac{\theta+1}{2}}\,dW_t, \tag{1.4}$$

where the values of $\theta < \frac{2k\mu}{D^2}$ define specific processes, as we will discuss below. The stationary distribution associated with this equation can be derived from the corresponding Fokker-Plank equation [18] and is given by [4]

$$p(x) = \frac{(\frac{2k}{D^2})^{\frac{2k\mu}{D^2}-\theta}}{\Gamma(\frac{2k\mu}{D^2} - \theta)}\, x^{\frac{2k\mu}{D^2}-\theta-1} e^{-\frac{2kx}{D^2}}. \tag{1.5}$$

Depending on the choice one makes for the θ value, the parameters take different meanings and the diffusion function encodes different fluctuation behaviors [12].

In particular, this framework allows us to distinguish between two types of stochasticity that are commonly encountered in population dynamics: demographic stochasticity, described by the SDE with $\theta = 0$, and environmental stochasticity, modeled through the SDE with $\theta = 1$. These two dynamics describe fluctuations with different natures [12].
Demographic stochasticity arises from the intrinsic randomness of birth and death events at

individual level and has a stronger impact in small populations. In contrast, environmental stochasticity is generated by random variations in environmental conditions; it affects all individuals simultaneously by modifying the parameters that govern the dynamics of the population and can therefore be interpreted as parametric noise [20] [21].

Eq. 1.4 with $\theta = 0$ becomes the Cox-Ingersoll-Ross model (CIR):

$$dX_t = k(\mu - X_t)\,dt + D\sqrt{X_t}\,dW_t. \tag{1.6}$$

Since the drift term is linear in X_t, all its statistical properties can be derived analytically [12] [10]. The time-dependent distribution associated with Eq. 1.6 is [22]

$$p(x,t) = \lambda\, e^{-\lambda(x_0 e^{-kt}+x)}[\frac{x}{x_0 e^{-kt}}]^{\frac{\delta}{2}} I_\delta(2\lambda\sqrt{x_0 x e^{-kt}}), \tag{1.7}$$

where I is the modified Bessel function of the first kind, $\lambda = \frac{2k}{D^2(1-e^{-kt})}$ and $\delta = \frac{2k\mu}{D^2}$.
The square root diffusion function in Eq. 1.6 stems from the continuous approximation of a birth-death Markov process, so is a consequence of the birth-death fluctuations on the population abundance (demographic noise) (Appendix A) [23].

Eq. 1.4 with $\theta = 1$ represents the stochastic logistic growth:

$$dX_t = k\,X_t(\mu - X_t)\,dt + DX_t\,dW_t. \tag{1.8}$$

This environmental noise model can be derived by allowing fluctuations in the parameters of the deterministic logistic growth model (Appendix B) [23] [18]. Notice that the non-linear drift term and the state-dependent diffusion function makes the Fokker-Plank equation associated with the SDE analytically unsolvable, so that there is no closed-form propagator for 1.8 [18].

1.1.1 Milstein method

The Milstein method is a powerful numerical method to solve numerically SDEs. It achieves a higher order of convergence than the Euler-Maruyama method, which means that the approximate trajectory has higher accuracy and is closer to the analytic one. In particular,

$$\mathbb{E}[|X(T) - X_n|] \leq C(\Delta t)^\gamma,$$

where $X(t)$ represents the real trajectory at time t and X_n is the n-th point of the numerical trajectory computed with a time step Δt, C is a constant. The order of convergence γ is $\gamma = 0.5$ for the Euler-Maruyama method and $\gamma = 1$ for the Milstein method [24]. Notice that the approximation becomes better with the simulation time step Δt decrease.

The numerical solution is obtained by iterating the Milstein update formula

$$X_{n+1} = X_n + A(X_n)\Delta t + B(X_n)\Delta W_n + \frac{1}{2}B(X_n)B'(X_n)(\Delta W_n^2 - \Delta t),$$

which is derived by approximately solving the integral form of the SDE:

$$X(t+\Delta t) = X(t) + \int_t^{t+\Delta t} A(X_s)\, ds + \int_t^{t+\Delta t} B(X_s)\, dW_s.$$

In Euler-Maruyama approximation, both integrals would have been approximated using $A(X_s) \simeq A(X_t)$ and $B(X_s) \simeq B(X_t)$. Instead, to obtain the Milstein scheme, one should expand the diffusion function B up to the first order in Δt, then

$$B(X_s) \simeq B(X_t) + B'(X_t)(X_s - X_t).$$

Using Euler-Maruyama locally, one has $X_s - X_t \simeq A(X_t)(s-t) + B(X_t)(W_s - W_t) \sim B(X_t)(W_s - W_t) + o(\Delta t)$, knowing $dW_t^2 = dt$, so that the equation above becomes

$$B(X_s) \simeq B(X_t) + B'(X_t)B(X_t)(W_s - W_t).$$

Then solving the Itô integral

$$\int_t^{t+\Delta t} (W_s - W_t)\, dW_s = \frac{\Delta W_t^2}{2} - \frac{\Delta t}{2}$$

one gets

$$X_{t+\Delta t} = X_t + A(X_t)\Delta t + B(X_t)\Delta W_t + \frac{1}{2}B(X_t)B'(X_t)(\Delta W_t^2 - \Delta t),$$

which leads to the recursive Milstein scheme [25].

Graphs in Figure 1.1 are constructed simulating $n = 5000$ trajectories for each process with the Milstein method, as it provides higher accuracy than the Euler–Maruyama scheme while maintaining a comparable computational cost [24]. Moreover, even though the Milstein approximation is not always necessary (e.g. for the OU process), using a single numerical scheme across all models ensures consistency and facilitates comparison of the results. In the case of the OU process, the Milstein method coincides with the Euler–Maruyama method, since the diffusion coefficient is constant.
For the OU and CIR processes, the histograms are constructed from the trajectories values at $T = 10\,s$ and the corresponding analytical time-dependent distributions (1.3)(1.7) evaluated at $T = 10\,s$ are superimposed. For the environmental noise process (EN), trajectories were simulated for a sufficiently long time to reach stationarity, and the stationary Gamma distribution was superimposed on the histogram.

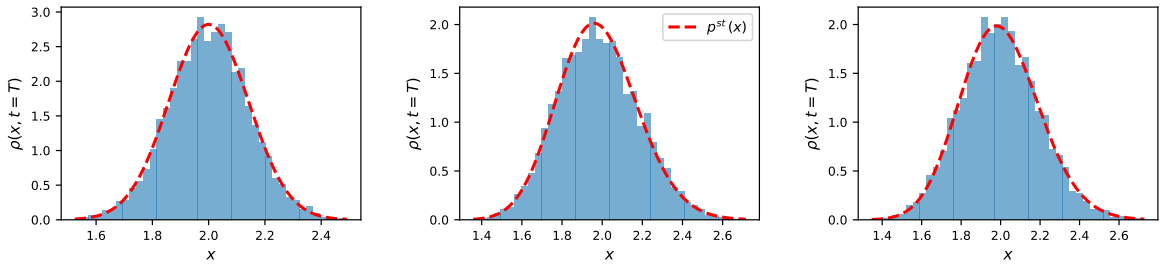

Figure 1.1: The histogram on the left is the OU process, the one on in the middle is the EN (stochastic logistic growth) and the one on the right is the CIR model. The parameters used for the simulations are $D = 0.2$, $k = 1$ and $\mu = 2$.

1.1.2 Autocorrelation time

The autocorrelation time is the characteristic timescale over which a fluctuating system retains a memory of its previous states before its dynamics become entirely randomized by noise.
We will use it frequently as the characteristic normalization factor for each process, in the plots that will follow.
The autocorrelation time is defined as

$$\tau = \int_0^\infty dt\, C(s),$$

where $C(s)$ is the normalized autocorrelation function, and so quantifies the time-scale of the autocorrelation exponential decay. Since we assumed to observe stationary trajectories, we are interested in computing the stationary autocorrelation time, which depends only on the interval s and not on the time t

$$C(s) = \frac{\mathbb{E}[X_s X_0] - \mu^2}{\sigma^2},$$

where μ and σ are mean and variance of the stationary distribution [12].
For the OU and CIR processes, the autocorrelation time can be computed analytically as [26][18][27]

$$\tau = \frac{1}{k}.$$

With the same calculation, one can find the autocorrelation time of the EN process, provided that the SDE is linearized near the stationary state μ within the regime of small fluctuations, obtaining (Appendix C)

$$\tau \sim \frac{1}{\mu k}.$$

Nevertheless, normalization values in the following plots are obtained computationally (procedure in Appendix C) using the time series obtained from the simulations. This is useful for τ_{EN}, and is also performed for the other processes for consistency. Note that we are assuming the autocorrelation is known and, therefore, it will not depend on the sampling protocol.

1.2 Inference from time-series data

When modeling a stochastic dynamical system, its properties are determined by the functional form of the drift and diffusion terms, as well as by the values of the model parameters [18]. Inference aims at estimating these parameters and identifying the model that best describes the observed stochastic dynamics from experimental data. Since different stochastic processes can share the same stationary distribution, temporal correlations are essential to distinguish between competing models. By analyzing the dynamics of a time-series it is possible to distinguish between two models and identify the one that better fits data.
In particular, we will discuss how the choice of a sampling protocol for the time-series affects the distinguishability of models and the estimation of parameters.
Specifically, since data collection occurs at discrete times, we assume that we have M data points describing the temporal dynamics of a population, obtained under time-invariant sampling conditions. Since data are collected at constant time intervals Δt_s, the samples used for inference consist of a time series $\{x_i\}_i$ associated with equally spaced time points $t_i = t_1 + (i-1)\,\Delta t_s$, where $x_i = X(t_i)$ [12]. Given this framework, the sampling protocol is defined by the values of the dataset size M and the sampling time interval Δt_s.
Notice that, although the observations are collected at discrete times, our goal is to describe the underlying dynamics through a continuous-time stochastic model.
Throughout this work, we initialize trajectories at $X_0 = \mu$, which corresponds to the peak of the stationary distribution. Since the stationary distributions considered here are strongly peaked around μ, the transient is very short, and this initialization is effectively equivalent to drawing X_0 from the stationary distribution. Consequently, the process rapidly behaves as if stationary, and we adopt this initialization as a practical approximation throughout the simulations.

1.3 Parametric inference

In this section, I will focus on parametric estimation, assuming prior knowledge of the model and estimating its parameters. We therefore consider a model

$$dX_t = A(X_t^{\vec{\Theta}}, \vec{\Theta})dt + B(X_t^{\vec{\Theta}}, \vec{\Theta})dW$$

where the functional forms of $A(*)$ and $B(*)$ are known and the goal is to estimate $\vec{\Theta} = (\mu, k, D)$. One of the most commonly used inference methods to estimate the parameters of a model from a time series of data is the maximum likelihood estimation (MLE) [12]. The maximum likelihood estimation principle states that the desired set of parameters $\vec{\Theta}$ is the one that makes the observed data more likely. This means maximizing the likelihood function, which represents the probability of observing a particular dataset given the model parameters [28]. Since we work with time series, we are interested in maximizing the path likelihood which quantifies the probability of the time series, given a fixed set of parameters. Since X_t is a Markov process, the joint probability of an entire path $\{x_{t_1}, x_{t_2}, \ldots, x_{t_M}\}$ factorizes as a product of transition probabilities: knowing the current state x_{t_i}, the future state $x_{t_{i+1}}$ is

independent of all past states. Therefore, the probability of generating the observed path is simply the probability of the initial value times the product of the probabilities of each individual transition, giving [12]

$$L(\vec{x}|\vec{\Theta}) = p_{t_1}(x_{t_1}|\vec{\Theta}) \prod_{i=1}^{M-1} p_{t_i,t_{i+1}(x_{t_{i+1}}|x_{t_i},\vec{\Theta})}, \tag{1.9}$$

where $p_{t_1}(x_{t_1}|\vec{\Theta})$ is the time-dependent probability density at time t_1 and $p_{t_i,t_{i+1}(x_{t_{i+1}}|x_{t_i},\vec{\Theta})}$ is the propagator of the process, which is the solution of the Fokker-Plank equation associated with equation 1.1 [12]:

$$\partial_t p_{s,t}(x|y) = -\partial_x[A(x)p_{s,t}(x|y)] + \partial_x^2[\frac{B^2(x)}{2}p_{s,t}(x|y)]. \tag{1.10}$$

Since throughout this work we operate in the stationary regime, the initial distribution coincides with the stationary distribution $p_{\text{stat}}(x_{t_1}|\vec{\Theta})$ and the propagator $p_{t_i,t_{i+1}}(x_{t_{i+1}}|x_{t_i},\vec{\Theta})$ depends only on the time interval $\Delta t = t_{i+1} - t_i$. Notice that, since equation 1.10 does not always admit an analytical solution, a closed-form expression for the propagator is often unavailable.

As seen in Section 1.1, the OU and CIR possesses analytical forms for their propagators (Eq. 1.3) (Eq. 1.7). This enables us to conduct the parameter estimation by numerically maximizing the exact likelihood [1] (Eq. 1.9), derived from the propagator of each process.

The goal is now to identify an optimal sampling protocol, which requires analyzing how the error in parametric inference varies with the sampling time interval Δt_s and the dataset size M. Once I got the parameter estimation, I can compute the relative error with respect to the true value (of our "artificial" data set) as

$$\varepsilon_{\hat{k}} = \frac{|\mathbb{E}[\hat{k}] - k|}{k}$$

and build a plot of the relative error versus the sampling time Δt_s. In our case, the aim is to build the relative error trend for D, μ and k, parameters of the OU and CIR processes. To compare different sampling intervals, samples corresponding to larger Δt_s were obtained by subsampling the original computational trajectory generated using the Milstein method, so as to avoid the randomness of different trajectories affecting the comparison. For this subsampling procedure to be meaningful, the sampling time interval must be significantly larger than the simulation time step, i.e., $\Delta t_s >> \Delta t$.

The following graphs (1.2, 1.3, 1.4) show both the comparison between the error trends for the OU and CIR processes as Δt_s varies, and the comparison between OU (and CIR) trends realized for different values of the data set size M.

[1] Notice that the likelihood form in Eq. 1.9 is the exact form one should use when discrete data are available, as it exploits the propagator between two points of the process.

The simulations of the processes are performed with $D = 0.2$, $k = 1$, $\mu = 2$ and $M = 1000$.

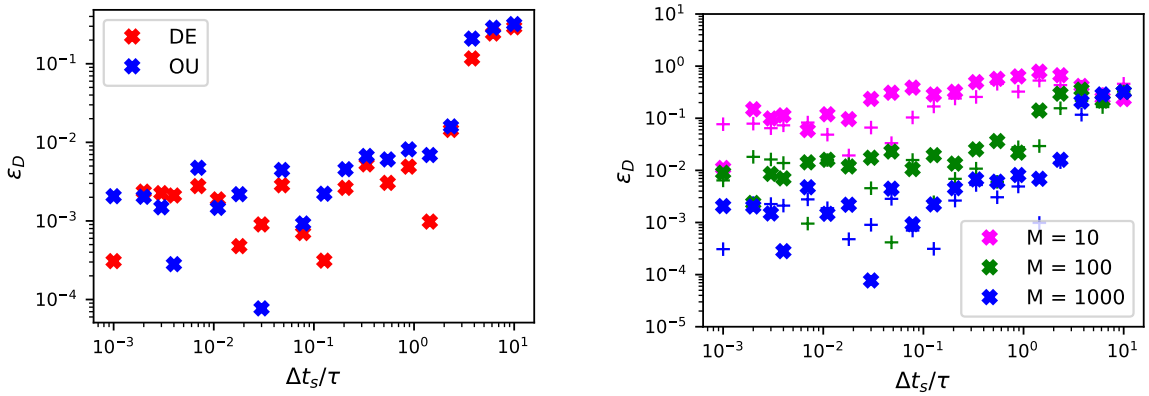

Figure 1.2: Plots of the relative error on the estimation of the diffusion coefficient D with respect to the sampling time interval Δt_s normalized by the autocorrelation time of the process τ. Left: comparison of the trends for 2 different processes (OU (Eq. 1.2) and CIR (Eq. 1.6)). Right: comparison of the trends for the 2 processes using different values of M, where crosses indicate CIR model and plus indicate OU model.

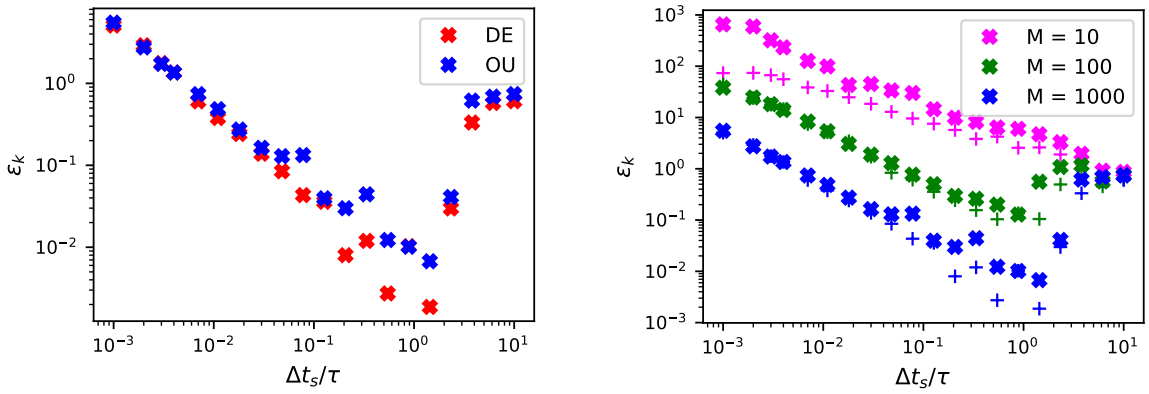

Figure 1.3: Plots of the relative error on the estimation of the coefficient k with respect to the sampling time interval Δt_s normalized by the autocorrelation time of the process τ. Left: comparison of the trends for 2 different processes (OU (Eq. 1.2) and CIR (Eq. 1.6)). Right: comparison of the trends for the 2 processes using different values of M, where crosses indicate CIR model and plus indicate OU model.

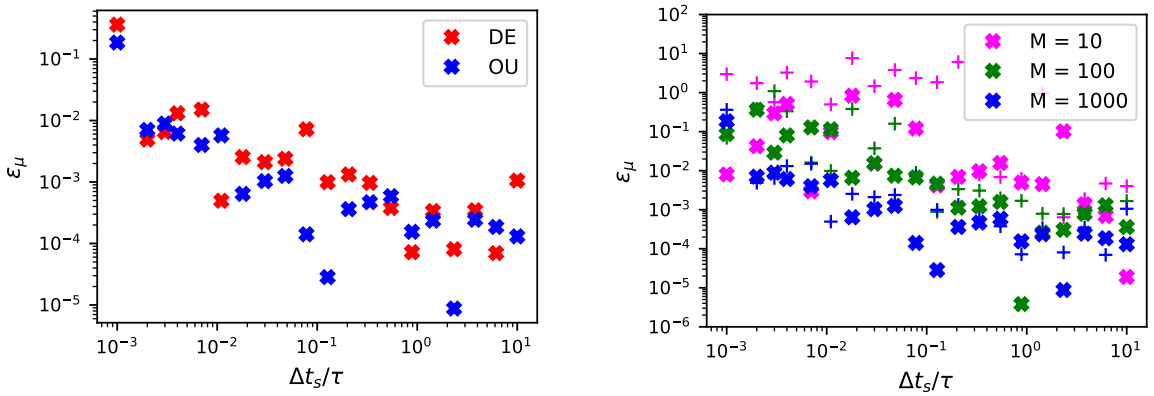

Figure 1.4: Plots of the relative error on the estimation of the coefficient μ with respect to the sampling time interval Δt_s normalized by the autocorrelation time of the process τ. Left: comparison of the trends for 2 different processes (OU (1.2) and CIR (1.6)). Right: comparison of the trends for the 2 processes using different values of M, where crosses indicate CIR model and plus indicate OU model.

The graphs show that the error depends on the sampling protocol: the estimation error follows a clear trend as the sampling interval varies, and this trend is influenced by the number of data points in the sample. In particular, the observed behavior depends on the parameter under consideration, but remains consistent across the two models [2] when the same parameter is examined and the sampling time is normalized by the autocorrelation time of each process. A similar effect is observed as M varies, which is particularly evident in the right panel of Figure 1.3, where the relative minimum disappears as M increases.

Once a trend in the error for each estimated parameter has been identified, the focus shifts to determining the sampling protocol that minimizes this error. As expected, increasing the dataset size M leads to an overall reduction in the errors, because of an increase in the quantity of information. However, although large datasets are preferable, they are not always available. The question we must therefore ask is: "to what extent does a small dataset compromise the information it contains?". Furthermore, once a maximum dataset size has been established, how should sampling be designed to maximize the information the dataset provides?

As we deduce from figures above, the answers to these questions depend on the parameter under consideration.

Looking at the trend that describes the error for parameter D, we observe that while the error is minimum for values $\Delta t_s < \tau$, at $\Delta t_s \sim 2\tau$, there is an abrupt increase. This is due to the role of D on the system's stochasticity, which dominates at short time scales, where data are highly correlated. Notice that, after the decorrelation of data, the dependence on M is less pronounced, as the error already reached high values because of the sampling time interval.

The parameter k achieves the best result for values $\Delta t_s \sim \tau$. This is because k governs the relaxation dynamics of the system and its estimation requires resolving temporal correlations,

[2] Although the parameters k and D are shared nominally between the two models, they possess distinct physical dimensions and operate on different scales. Consequently, while their dimensionless relative errors exhibit similar scaling trends in the plots, the corresponding absolute errors can be very different. The choice of relative error is instrumental here to ensure a normalized, scale-independent comparison, but it should not mask the fact that the actual physical uncertainty affects each model differently.

which are lost for large Δt_s - causing an abrupt increase in the error trend - but it also influences the drift, which is obscured by noise at very small Δt_s values.
Finally, one can notice, even if it is less pronounced, that the μ parameter is optimized for high values of the sampling time interval, this is due to its role as the mean value in the stationary distributions of the models, which is better studied with uncorrelated data [12].

The method used above to estimate parameters from OU and CIR models works as long as a closed-form solution for the propagator is available. When we move to the EN model, described in Section 1.1, the likelihood computation becomes the real challenge, as Eq. 1.9 is no longer applicable.

In the following I describe the alternative solutions developed in [12] and compare the resulting trends with the one obtained using MLE with the exact likelihood.

For the estimate of D, one can avoid the computation of the likelihood using the moment-based estimation of the diffusion function, which relies on the relationship between the second moment of the process increments and the diffusion function $B(\cdot)$ [3] [12][29]:

$$B^2(x) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \mathbb{E}[(X_{t+\Delta t} - x)^2 | X_t = x].$$

The empirical quadratic variation estimator of the diffusion function reads [12]:

$$\hat{B}_i = \sqrt{\frac{1}{\Delta t_s \, M_i} \sum_{j \in J_i} (x_{t_{j+1}} - x_{t_j})^2} \tag{1.11}$$

where M_i is the number of points into the i-th bin of the discretized x space, $J_i = \{j : x_{t_j} \in [x_i, x_i + \Delta x)\}$ and Δt_s is the sampling time interval of the dataset (as defined in Section 1.2). Note that Eq. 1.11 is an approximation for the previous formula, which becomes worse as the sampling time increases and the dataset size decreases.
The estimate of the diffusion parameter D can be easily obtained assuming a specific functional form for $B(x)$. For OU, $B(x) = D$ is constant, so $\hat{D}$ evaluated in the i-th bin is $\hat{B}_i$. Since this expression no longer depends on x_i, we can average over all bins, replacing the sum over J_i with an average over the entire time series, to obtain

$$\hat{D}_{OU} = \sqrt{\frac{1}{\Delta t_s} \mathbb{E}[(x_{t+\Delta t_s} - x_t)^2]},$$

where $\mathbb{E}[\cdot]$ denotes the average over the time series. With the same procedure, substituting the functional formula $B(x) = Dx$ in 1.11 and evaluating $\hat{D}$ in the i-th bin centered in x_i, for model EN one obtains

$$\frac{\hat{B}_i}{x_i} = \sqrt{\frac{1}{\Delta t_s \, M_i} \sum_{j \in J_i} \frac{(x_{t_{j+1}} - x_{t_j})^2}{x_{t_j}^2}}.$$

Taking again the average over all bins, since D is a constant parameter which no longer

[3] Where we are assuming the process to be at steady-state.

depends on x, one gets

$$\hat{D}_{EN} = \sqrt{\frac{1}{\Delta t_s}\mathbb{E}[\frac{(x_{t+\Delta t_s} - x_t)^2}{x_t^2}]},$$

where $\mathbb{E}[\cdot]$ is the average over the time-series.
The drift function can be inferred using the same method, since [29]

$$A(x) = \lim_{\Delta t \to 0} \frac{1}{\Delta t}\mathbb{E}[(X_{t+\Delta t} - x)|X_t = x].$$

However, in this case, estimating the individual parameters becomes more challenging, as the drift function depends on multiple parameters rather than a single one.
Therefore, it is more convenient to estimate the parameters of the drift function using maximum likelihood estimation (MLE) and an approximate form of the likelihood. This allows us to estimate the parameters - in our case, μ and k - jointly.

A possible solution, shown in [12], is to assume a continuous-time stochastic process and adopt a path-inference approach for the construction of the likelihood. Under this assumption, one can use Girsanov's theorem to relate path probabilities and obtain the following equation for the likelihood [30] [12]:

$$L_t = \frac{d\mathbb{P}}{d\mathbb{V}} = e^{tl_t} = \exp[\int_0^t \frac{A(x_s, \vec{\Theta})}{B^2(x_s)}dx_s - \frac{1}{2}\int_0^t (\frac{A(x_s, \vec{\Theta})}{B(x_s)})^2 ds] \tag{1.12}$$

where $\mathbb{P}$ is the measure for the general stochastic process in equation (1.1) and $\mathbb{V}$ is the measure for the diffusion process $dX_t = B(X_t)dW_t$. Note that the first integral in the exponent of Eq. 1.12 is a stochastic integral defined in the Itô sense. One can find the maximum likelihood estimator by directly maximizing the normalized log-likelihood, which is monotonically related to the likelihood [28]:

$$\nabla_{\vec{\Theta}} l_t = 0. \tag{1.13}$$

This is convenient both computationally, due to higher numerical stability, and analytically, as it simplifies the computation of derivatives by transforming products into sums.
Let us define $A(x) = \vec{\Theta} \cdot \vec{g}(x)$, where $\vec{\Theta} = (\Theta_1, \Theta_2)$ and $\vec{g}(x) = (g_1(x), g_2(x))$. By solving equation 1.13, one obtains [12]

$$\begin{cases} \Theta_1 = \frac{\langle g_2^2 \rangle_B [g_1]_B - \langle g_1 g_2 \rangle_B [g_2]_B}{Z} \\ \Theta_2 = \frac{\langle g_1^2 \rangle_B [g_2]_B - \langle g_1 g_2 \rangle_B [g_1]_B}{Z} \end{cases} \tag{1.14}$$

where $Z = \langle g_1^2 \rangle_B \langle g_2^2 \rangle_B - \langle g_1 g_2 \rangle_B^2$, $\langle g \rangle_B = \frac{1}{T}\int_0^T \frac{g(x_t)}{B^2(x_t)}\,dt$ and $[g]_B = \frac{1}{T}\int_0^T \frac{g(x_t)}{B^2(x_t)}\,dx_t$.
Notice that, in practice, we observe data at discrete times, so the integrals in 1.14 must be discretized in order to be efficiently computed, i.e.

$$\langle g \rangle_B \sim \frac{1}{T}\sum_{i=1}^{M} \frac{g(x_{t_i})}{B^2(x_{t_i})}\,\Delta t_s,$$

$$[g]_B \sim \frac{1}{T}\sum_{i=1}^{M}\frac{g(x_{t_i})}{B^2(x_{t_i})}\,(x_{t_{i+1}} - x_{t_i}),$$

where $T = (M-1)\Delta t_s$ [12].
Note that

- for the OU process $\Theta_1 = k\mu$, $\Theta_2 = k$, $g_1(x) = 1$, $g_2(x) = -x$ and $B(x) = D$,
- for the EN process $\Theta_1 = k\mu$, $\Theta_2 = k$, $g_1(x) = x$, $g_2(x) = -x^2$ and $B(x) = Dx$.

The following graphs (1.5, 1.6, 1.7) show both the comparison between the error trends for the OU and EN processes as Δt_s varies, and the comparison between OU (and EN) trends realized for different values of the data set size M.
The simulations of the processes are performed with $D = 0.2$, $k = 1$, $\mu = 2$ and $M = 1000$.

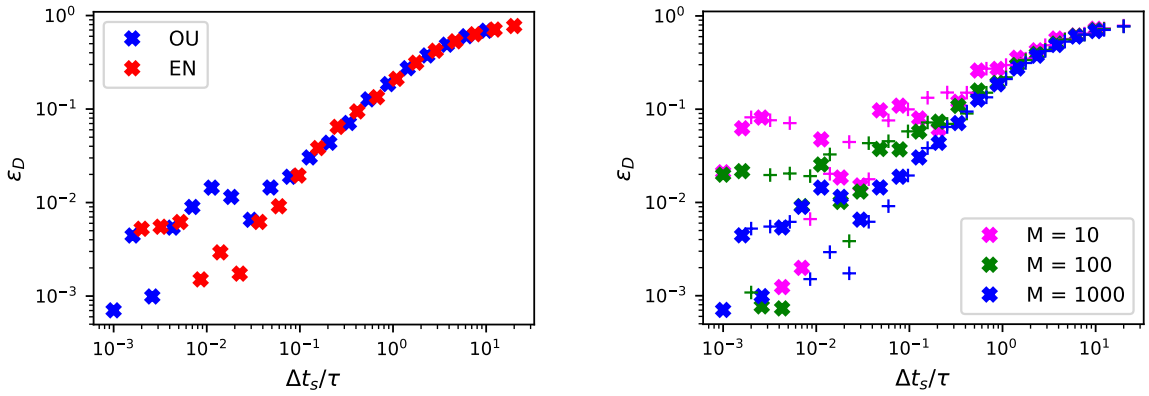

Figure 1.5: Plots of the relative error on the estimation of the diffusion coefficient D with respect to the sampling time interval Δt_s normalized by the autocorrelation time of the process τ. The estimate is computed using moment-based estimation. Left: comparison of the trends for 2 different processes (OU (1.2) and EN (1.8)). Right: comparison of the trends for the 2 processes using different values of M, different symbols indicate different models.

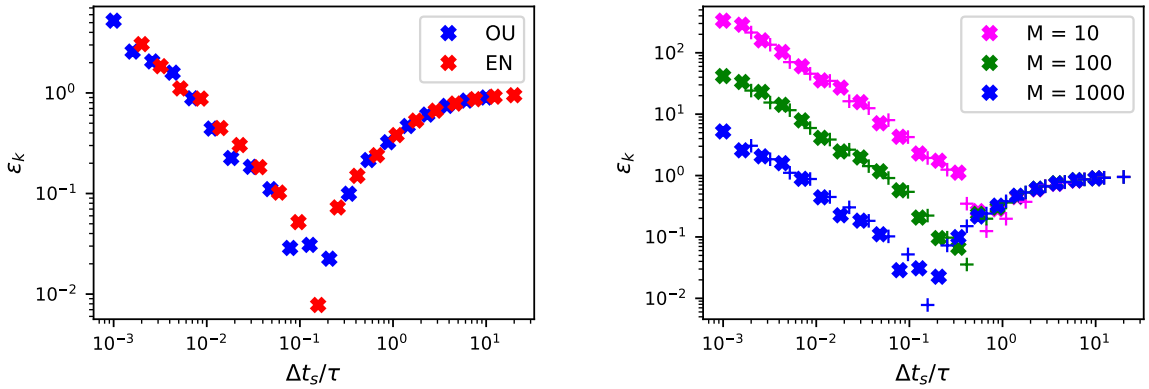

Figure 1.6: Plots of the relative error on the estimation of the coefficient k with respect to the sampling time interval Δt_s normalized by the autocorrelation time of the process τ. The estimate is computed using path-inference approach. Left: comparison of the trends for 2 different processes (OU (1.2) and EN (1.8)). Right: comparison of the trends for the 2 processes using different values of M, different symbols indicate different models.

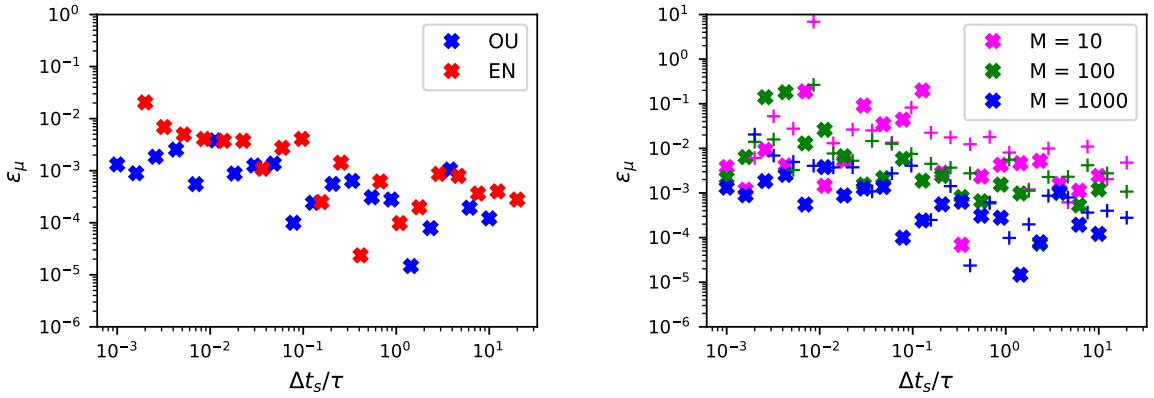

Figure 1.7: Plots of the relative error on the estimation of the coefficient μ with respect to the sampling time interval Δt_s normalized by the autocorrelation time of the process τ. The estimate is computed using path-inference approach. Left: comparison of the trends for 2 different processes (OU (1.2) and EN (1.8)). Right: comparison of the trends for the 2 processes using different values of M, different symbols indicate different models.

From the graphs above, we observe again a strong dependence on the sampling protocol. Error trends do not change between models when sampling time is normalized by the autocorrelation time, but are very specific to the parameter under consideration. The dependence on M is particularly evident in the right panel of figure 1.6 and becomes particularly pronounced when Δt_s takes on small values. This is due to the fact that, at a high sampling frequency, the collected data are strongly correlated and many observations are redundant. Under these conditions, a larger number of samples significantly affects the estimation error, as it allows for a longer section of the trajectory to be sampled, thereby increasing the amount of information that can be derived from it [31]. Recall that increasing the number of data M at a fixed Δt_s, means to increase the total time of the observed trajectory T, since $T = \sum_{i=1}^{M} \Delta t_i = M\Delta t_s$. If we compare the results obtained from the analysis previously conducted using the exact likelihood, we notice that overall the trends are very similar. The most obvious difference can be observed in D and k error trends. In both plots, the error is regularly increasing from small/intermediate values to high values of the sampling time interval Δt_s, while in the previous comparison the error showed an abrupt increase at $\Delta t_s \sim 2\tau$. This is likely due to the approximation error caused by the discretization in Girsanov's formula, which becomes more significant for large values of the sampling interval: the trend in Figure 1.5 and Figure 1.6 are therefore influenced not only by the actual reduction in information contained in the data, but also by the approximation error coming from the method we used to estimate the likelihood. Regarding the parameter μ, as before, the error decreases with higher values of the sampling time interval. Given that the process is in a stationary regime, the estimator of μ mathematically converges to the sample mean, rendering it robust to discretization errors; consequently, the estimate does not deteriorate, even when the likelihood is approximated via Girsanov's theorem [30].

Both analysis - using the exact likelihood method and the path-inference approach - lead to the conclusion that a perfect sampling time should be heterogeneous, based on the parameter

one wants to estimate. Since this is an impractical solution, a compromise could be represented by Δt_s equal to the autocorrelation time [12].

1.4 Model discrimination

When studying dynamical systems, one often aims to identify a model that provides a coherent description of the underlying process. This task is frequently challenging; indeed, when sampling a stochastic process, the resulting dataset may be compatible with multiple models. As previously mentioned, the ability to correctly identify a specific model depends on the nature of the dataset [12].

Let's define the probability P to correctly identify the model generating data and observe its trend with respect to a Δt_s variation, for different values of M. Using n_{traj} paths of synthetic data, P is computed as

$$P = \frac{n_{\text{correct}}}{n_{\text{traj}}}$$

where n_{correct} is the number of times the model is been correctly identified. A correct identification is counted when the true model likelihood value is higher than the competitive model likelihood one.
The discrimination between two models is made by simulating and subsampling one of the two models (which will be defined as the true model), forgetting about the simulation parameters, and then performing parametric inference on the datasets for both models. In this way, the analysis is non-trivial and unbiased by the knowledge of the true simulation parameters (Figure 1.8).

In Figure 1.9 I perform model discrimination between CIR (1.6) and OU (1.2). The likelihoods for the two models are computed through the analytic propagators (1.7)(1.3).
The plot of P_{CIR} represents the probability for the likelihood of the CIR model (being CIR the true model) to be higher of the one for the OU model, the same meaning is consistently applied to P_{OU}. So the index $i = \text{OU, CIR}$ in P_i indicates the true model for each plot. Specifically, P_{CIR} and P_{OU} denote the conditional probabilities of correct assignment when the data are generated by the CIR and the OU model respectively, i.e.

$$P_i = P(\text{correct}|i).$$

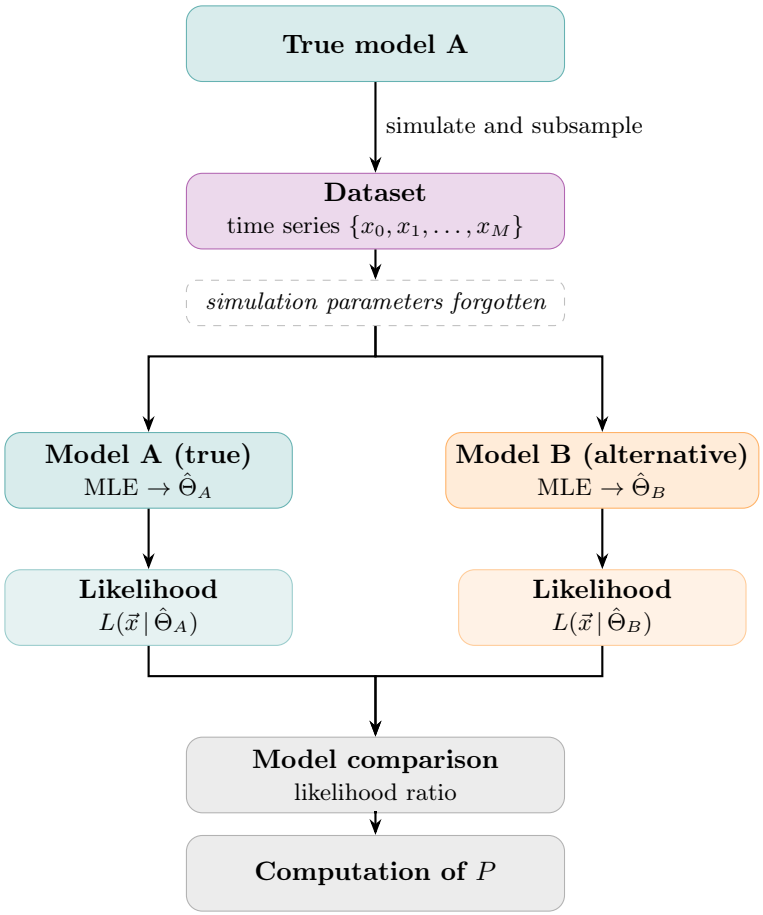

Figure 1.8: Model discrimination scheme. A dataset is generated and subsampled from the true mode. Both candidate models are fitted independently via MLE, and the resulting likelihoods are compared.

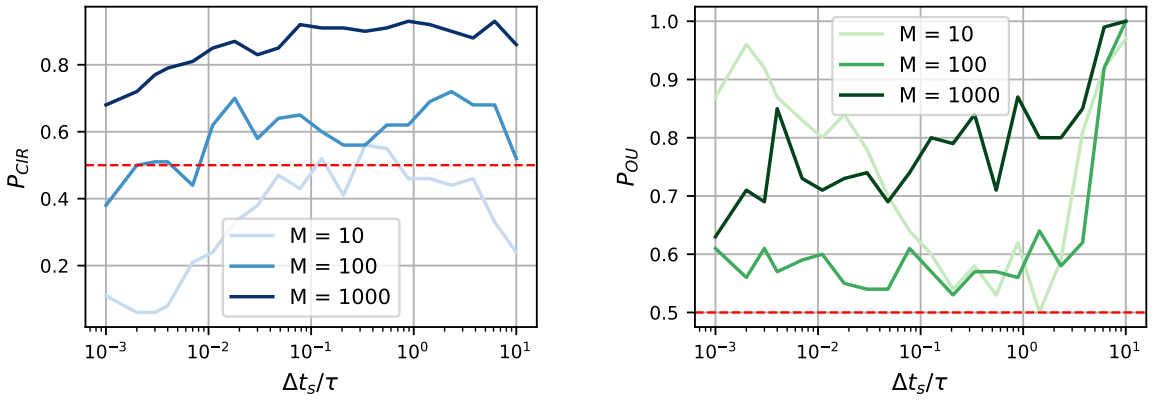

Figure 1.9: Left: plot of the probability P_{CIR} to identify correctly the CIR model with respect to the sampling time interval Δt_s normalized by the autocorrelation time of the process τ. Right: plot of the probability P_{OU} to identify correctly the OU model with respect to the sampling time interval Δt_s normalized by the autocorrelation time of the process τ.

It is immediately clear that the results exhibit a bias for small values of M. Observing the case with $M = 10$, we see that for sampling times shorter than τ, the OU model is preferred regardless of which model is the true one. In the left graph, the distinguishability drops below 0.5, indicating that the incorrect model is selected more frequently than the correct one. In the right plot, the distinguishability is high for high sampling frequencies, but this does not reflect a choice guided by the information content of the data, but rather by the absence of such information, which favors the simpler model.

The bias favoring OU for small values of the sample size can be explained by analyzing the analytical structure of the likelihood. Since the log-likelihood is computed as the sum of the logarithm of the transition probability step by step, it acts accumulating penalties when data deviate from the expected trend [32].

In OU model, the propagator variance does not depend on the state X_t and so the penalty for stochastic fluctuations is limited and uniform along the whole trajectory.

In CIR model, on the other hand, since the diffusion function depends on $\sqrt{X_t}$, the penalty term presents an inversely proportional dependence on X_t [30]. The stationary state dynamics of our trajectories is evolving near 0 [4] and so fluctuations, in this region, are highly penalized. Notice the penalty is reduced for higher values of Δt_s, because the conditional variance of CIR grows with Δt_s [33]:

$$\mathrm{Var}(X_{t_{i+1}}|X_{t_i}) = \mu\frac{D^2}{2k}(1-e^{-k\Delta t_s})^2 + X_{t_i}\frac{D^2}{k}(e^{-k\Delta t_s} - e^{-2k\Delta t_s}) \xrightarrow{\Delta t_s\to\infty} \mu\frac{D^2}{2k}.$$

When dealing with small values of M, a single noisy point makes the log-likelihood of CIR drop. However, for high values of M these extreme single fluctuations are statistically balanced [34] [35].

Note that the distinguishability trends remain asymmetric for the two different models at every value of M.

In general, distinguishability increases as the sampling time interval increases. In fact, at high

[4] $x_0 = \mu = 2$

sampling frequencies, the total length of the observed trajectory is very short, proportional to M: $T = \Delta t_s\, M$ [12]. Therefore, the algorithm cannot distinguish between the two models because the information contained in the data is limited.
The asymmetry between the 2 graphs at high M values is led by the abrupt increase of P_{OU} at $\Delta t_s > \tau$, which is instead absent when looking at the left graph, where the distinguishability stabilizes at $\sim 0.7/0.9$. As before, this is due to the lower strength of the CIR model [32]. In particular, when Δt_s become higher than the autocorrelation time, the uncorrelated data begin to describe the stationary distributions of the 2 models. While the Gaussian of the OU model simply ignores the asymmetry of CIR data, the Gamma distribution of the CIR tries to fit the symmetry of the OU process, causing the parameters to reach extreme values, which highly penalize the likelihood [36]. Recall indeed that for the Gamma to reach symmetry the shape $\alpha = \frac{2k\mu}{D^2}$ should go to infinity [30].

To quantify the overall performance of the model discrimination procedure, we define the unbiased discrimination probability P, which does not depend on the generating model and represents the actual probability to distinguish the two competing models. Under the assumption of equal prior probabilities for both competing models, P can be computed as the arithmetic mean of the individual discrimination probabilities:

$$P = P_{CIR}\,\mathcal{P}(CIR) + P_{OU}\,\mathcal{P}(OU) = \frac{P_{CIR} + P_{OU}}{2},$$

where $\mathcal{P}(CIR) = \mathcal{P}(OU) = 0.5$ are the prior probabilities representing the degree of confidence one assigns to each model before analyzing the data. The choice of considering a symmetric prior probability is guided by the absence of a reason to prefer a candidate over the other.

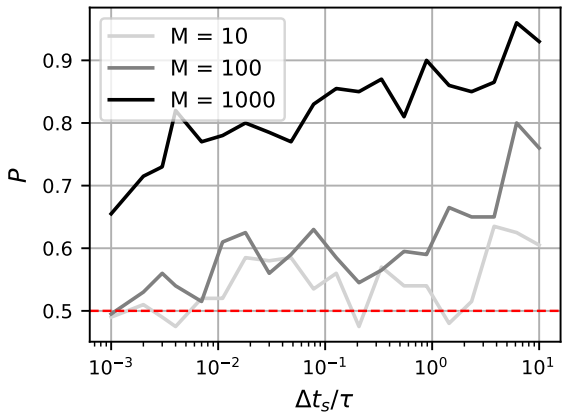

Figure 1.10: Unbiased discrimination probability P to distinguish CIR from OU, defined as $\frac{P_{CIR}+P_{OU}}{2}$. The dashed red line is the lower limit for P, i.e. the limit representing indistinguishability, where the probability of choosing the correct model is reduced to a random choice.

Looking at Figure 1.10 we can conclude that the 2 models are easily distinguishable for high values of both the sampling time and the dataset size.

Performing the same analysis for OU and EN is more complicated because for EN the analytic propagator is not available. As we did before for parametric inference, one could use the Girsanov theorem to compute the likelihood. Anyway, this method introduces consistent approximation errors as Δt_s increases. A possible solution is presented in [12] and not treated in this work, where I chose to focus on the applicability of this framework to Jump processes.

This concludes the chapter on continuous-state and continuous-time stochastic processes. We have revisited some of the topics previously introduced in [12]. In particular, we identified an optimal sampling protocol for parametric inference and discussed the process and results for model distinguishability between 2 simple SDE models. Furthermore, we commented on the comparison between the error trend obtained through the maximization of the exact likelihood and the one obtained through the maximization of an approximate likelihood. Building on this framework, we now turn to inference via maximum likelihood estimation (MLE) in the context of jump processes.

Chapter 2

Jump processes

Jump processes are commonly used to describe systems involving a countable set of elements, such as individuals in population dynamics, particles, genes, or proteins. These processes are defined by transition rates between discrete states and are governed by Master Equations which determine the time evolution of the probability distribution:

$$\frac{dp_n(t)}{dt} = \sum_{n'} [w_{n,n'} p_{n'}(t) - w_{n',n} p_n(t)], \tag{2.1}$$

where $w_{n,n'} \equiv w(n' \to n)$ denotes the transition rate per unit time from state n' to state n [1] Consequently, the first term in the sum represents the probability flux into state n, while the second term accounts for the flux out of n [27]. Jump processes can reproduce deterministic dynamics in an appropriate limit [27], while capturing intrinsic randomness through discrete, individual jump events. In this framework the transition rates provide a self-consistent description of the system's thermodynamics and kinetics, as they determine uniquely both the macroscopic average behavior and the microscopic fluctuations around it. In contrast, in SDEs, drift and noise are mathematically independent terms that originate from different modeling assumptions. Consequently, in SDEs, a fluctuation-dissipation relation is needed to phenomenologically link the noise amplitude to the drift in a physical context [18] [37]. Furthermore, the Fokker-Plank equation, whose solution defines the distribution of a continuous stochastic variable, is obtained through a Kramer-Moyal expansion of the Master Equation, truncated at the second order (Appendix A) [27].

In the context of jump processes, parametric inference aims at estimating the transition rates per unit time $w_{n,n'}$ for all n and n' states, which define the dynamics of the system.

To perform parametric inference, we again use Maximum Likelihood Estimation (MLE). Having discrete datasets, the exact likelihood of a jump process has the same structure as Eq.

[1] We will mainly use arrow notation, as it is more readable. Subscript notation will be used for brevity only on a few occasions.

1.9, with the continuous variable x_{t_i} replaced by the discrete state n_i:

$$L(\vec{n}|\vec{\Theta}) = p_{t_1}(n_1|\vec{\Theta}) \prod_{i=1}^{M-1} p_{t_i,t_{i+1}}(n_{i+1}|n_i,\vec{\Theta}), \tag{2.2}$$

where $p_{t_i,t_{i+1}}(n_{i+1}|n_i,\vec{\Theta})$ is the propagator, obtained as the solution of the Master Equation (Eq. 2.1) evaluated at $n = n_{i+1}$, with initial condition $p_n(t_i) = \delta_{n,n_i}$. However, even in this case a closed-form for the propagator doesn't always exist [27] [18]. So, once again the aim is to address an approximation that can be applied in the context of jump processes.

Formally, a trajectory for a jump process is characterized by a sequence of discrete states and associated times at which transitions occur. Let $t_0 = 0$ be the initial time and $\{t_i\}_{i\geq 1}$ the sequence of jumping times. The path of the system is defined as a stochastic step function with $n(t) = n_i$ for $t \in [t_i, t_{i+1}[$, where n_i is the state of the system during the waiting time between the i-th and the $(i+1)$-th jump.
Assuming the entire path of $n(t)$ is known, i.e. all the jumps of the process in a given time window $[t_0, t_J]$, where J is the number of jumps from time t_0 to time t_J, one can compute a path likelihood of the jump process as follows:

$$L_{[t_0,t_J]} = \prod_{i=0}^{J-1} e^{-\Delta t_i W(n_i)}\, w(n_i \to n_{i+1}), \tag{2.3}$$

where $W(n) = \sum_m w(n \to m)$ is the total exit rate from state n, which defines the reciprocal of the average time between jumps. The exponential part $e^{-\Delta t_i W(n_i)}$ is the probability for the system to remain in the same state n_i for a time $\Delta t_i = t_{i+1} - t_i$ and $w(n_i \to n_{i+1})$ is the rate for the specific transition occurred at time t_{i+1} (Appendix D).

This is equivalent to having access to the full output of a Gillespie simulation for the process [38]. However, for real-world experimental data, this scenario is rarely feasible. Indeed, jumping times are stochastic and inherently unpredictable; hence, the sampling schedule cannot be dynamically adapted to the process. Consequently, observation times almost never coincide with the exact instances at which jumps occur. Therefore, when considering a constant sampling interval Δt_s, the total number of detected jumps, denoted as J^s (with $J^s \leq J$), will generally differ from the true underlying dynamics. The exact jumping times are approximated by the multiples of Δt_s at which a change in state is observed. This leads to an approximation of the path likelihood:

$$\tilde{L}_{[t_0,t_{J^s}]} = \prod_{i=0}^{J^s-1} e^{-n_{s,i}\,\Delta t_s W(n_i)}\, w(n_i \to n_{i+1}), \tag{2.4}$$

where $n_{s,i}$ is the number of sampling points collected while the system remains in state n_i. Here $\{n_i\}$ denotes the sequence of observed states, which may differ from the true hidden sequence due to undetected jumps. Therefore, the time spent in state n_i is approximated by $n_{s,i}\,\Delta t_s$. Notice that this approximation becomes exact as $\Delta t_s \to 0$ (Appendix F). This

becomes evident when looking at the trajectory of the jump process compared to the paths reconstructed using our approximation in Figure 2.1. Then it is clear that the estimation of the likelihood from collected data will inevitably show a significant dependence on the sampling protocol, as we will see in the following section.

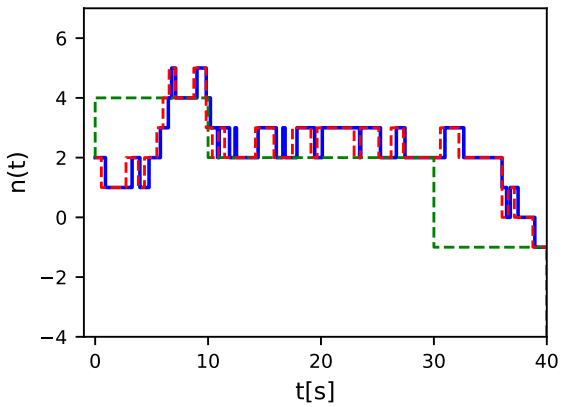

Figure 2.1: Comparison between the real trajectory and two of the analyzed sample trajectories. The red one belongs to the sample with $\Delta t_s = 0.379$. The green one belongs to the sample with $\Delta t_s = 10.0$.

In the following section I will focus on a very simple jump process to illustrate the general framework of parametric inference approximation, which consists in using the approximate path likelihood $\tilde{L}_{[t_0,t_{J^s}]}$ in Eq. 2.4 for MLE.

2.1 Parametric inference for a simple Poissonian jump process

Consider a Poissonian jump process with rates $w(n \longrightarrow n+i)$. In particular, define

$$w(n \longrightarrow n+j) = \begin{cases} k_+ & \text{if } j = 1, \\ k_- & \text{if } j = -1. \end{cases} \tag{2.5}$$

with an exponential waiting time between jumps given by (Appendix E)

$$f(t) = (k_+ + k_-)e^{-(k_+ + k_-)t} \tag{2.6}$$

To perform parametric inference I apply the maximization of the likelihood in Eq. 2.3 (MLE) with respect to the rates. In particular, let us compute the rate k_+ by maximizing the log-likelihood:

$$\partial_{k_+} l_{[t_0,t_J]} = 0$$

where l is the log-likelihood i.e.

$$l_{[t_0,t_J]} = \frac{1}{J}\sum_{i=0}^{J-1}(-\Delta t_i W(n_i) + \log(w(n_i \to n_{i+1}))) \tag{2.7}$$

where the total exit rate is

$$W(n) = \sum_j w(n \to n+j) = w(n \to n+1) + w(n \to n-1) = k_+ + k_-. \tag{2.8}$$

Deriving with respect to k_+, one obtains

$$\partial_{k_+} l_{[t_0,t_J]} = \frac{1}{J}\sum_{i=0}^{J-1}(-\Delta t_i) + \frac{1}{J}\sum_{i=0}^{J_+-1}\frac{1}{k_+}$$

where in the second sum J_+ is the number of positive jumps: after deriving, the non-zero terms in the summary are those with $w(n \to n+1) = k_+$.
Maximizing one obtains

$$\hat{k}_+ = \frac{J_+}{\sum_{i=0}^{J-1}\Delta t_i}. \tag{2.9}$$

where $\Delta t_i = t_{i+1} - t_i$ are the times between 2 consecutive jumps [2].
Then, the relative error for the parameter estimation is given by

$$\varepsilon_{k_+} = \frac{|\mathbb{E}[\hat{k}_+] - k_+|}{k_+},$$

where the $\mathbb{E}[\cdot]$ is the average over different realizations (multiple trajectories).
If we had access to the entire path of the dynamics, i.e. to all trajectory jumps, the estimator $\hat{k}_+$ would be unbiased, so that $\mathbb{E}[\hat{k}_+] \simeq k_+$ and the relative bias would vanish. The relative fluctuations of single estimates, $(\hat{k}_+ - k_+)/k_+)$, would instead be centered around zero with variance approximately $1/\mathbb{E}[J_+]$ [38].

As already said, this is not the case for real datasets, which are constructed by sampling M points from the trajectory at fixed Δt_s, so that the number of jumps, which can be detected from the sample, J^s does not correspond exactly to the actual number of jumps. This means the number of positive jumps in the estimator 2.9 is approximated by J^s_+, which represents the number of positive jumps detected from the sample and depends on Δt_s. Furthermore, the term $\sum_{i=0}^{J-1}\Delta t_i$ in 2.9 is approximated by $M\Delta t_s$. This procedure corresponds to directly maximizing the approximate path likelihood described in Eq. 2.4.
In the following graphs we show the relative error trend with respect to the sampling time interval Δt_s, where the estimate $\hat{k}_+$ is computed through the approximation introduced above:

$$\tilde{\hat{k}}_+ = \frac{J^s_+}{M\Delta t_s} \tag{2.10}$$

The x axis of graphs in figures 2.2 and 2.3 represents the sampling time normalized by the autocorrelation time. Notice that in this specific case the autocorrelation time τ of the process is given by the average value between two consecutive jumps [18]:

$$\tau = \frac{1}{k_+ + k_-}.$$

[2] Notice that $\sum_{i=0}^{J-1}\Delta t_i = t_J$, assuming $t_0 = 0$

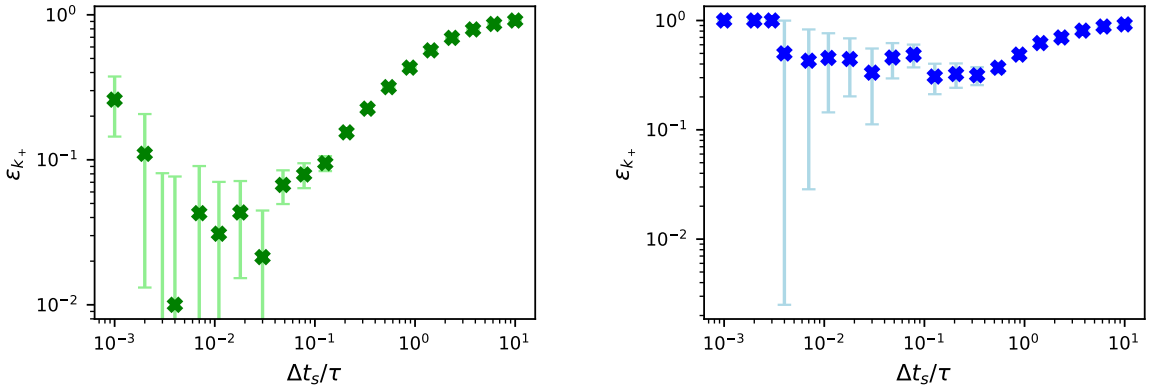

Figure 2.2: Plot of the relative error k_+ with respect to the sampling time Δt_s. Left: the number of points for each time series is $M = 1000$. Right: the number of points for each time series is $M = 10$. The error bars are computed as $\frac{\sigma_{\hat{k}_+}}{\sqrt{n_{\text{traj}}}}$, where n_{traj} is the number of trajectories used to infer $\hat{k}_+$.

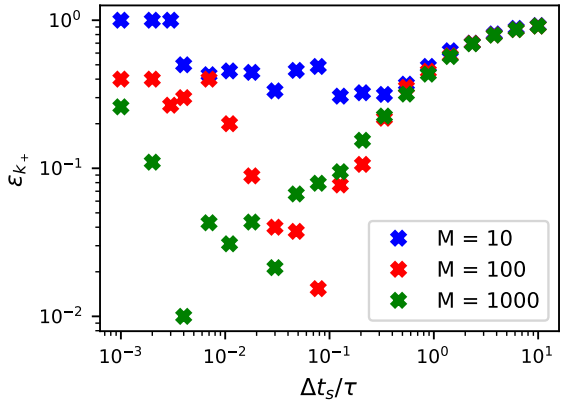

Figure 2.3: Comparison of plots of the relative error computed using different sample sizes.

Observing the graph above, one can notice that the relative error, for medium-high values of Δt_s, increases as the sampling frequency decreases. This is expected because, moving to bigger values of Δt_s, we lose information about the actual number of jumps. In particular, for k_+ one needs to compute the number of positive jumps J_+ (see equation 2.9), but when $\Delta t_s > \tau$ we are not able anymore to detect all jumps. In figure 2.1 I compared 2 different time series (the red and the green dashed lines) to the real trajectory of the jumping process (blue solid line). The red sample is built through a short Δt_s, and it follows almost exactly the real trajectory. The green sample is built with a very high Δt_s, and follows the original trend only on average, missing a lot of jumps. Therefore, the information gained from the last one is highly limited, because between two consecutive points of the sample we have a "blind spot", where we don't see the actual dynamics of the process.

If one admits a constant observational time T_{obs}, when increasing the sampling frequency $\Delta t_s < \tau$, we would expect the trend to flatten out, maintaining low values, because a higher sampling frequency allows us to detect the real number of jumps. But this is not our case

because M is fixed, then, as the sampling interval decreases, the total observational time T_{obs} also reduces. This setup implies two major consequences:

- The trajectory portion I am sampling is shorter: I lose information about the whole dynamics (see graph 2.4). Consequently, in the limit of high-frequency sampling ($\Delta t_s << \tau$), the relative error trend increases again. This behavior arises because the drastically shortened observation window T_{obs} scales down to a point where the system rarely transitions. As a result, the total number of observed upward jumps, J_+, becomes critically small. The resulting lack of statistical data prevents an accurate estimation of the transition rate k_+, thereby driving up the corresponding relative error.

- The error bars on the estimated parameter increase [3], leading to larger statistical fluctuations in the observed trend. Specifically, the estimator is (Eq. 2.10)

$$\hat{k}_+ = \frac{J_+}{T_{\text{obs}}}.$$

The number of positive jumps follows a Poisson distribution with rate k_+ (Appendix G), so

$$\text{Var}(\hat{k}_+) = \frac{\text{Var}(J_+)}{T_{\text{obs}}^2} = \frac{k_+ T_{\text{obs}}}{T_{\text{obs}}^2} = \frac{k_+}{T_{\text{obs}}}.$$

Then the error bars, computed in the graph as

$$\frac{\sigma_{\hat{k}_+}}{\sqrt{n_{\text{traj}}}},$$

grows with the sampling frequency.

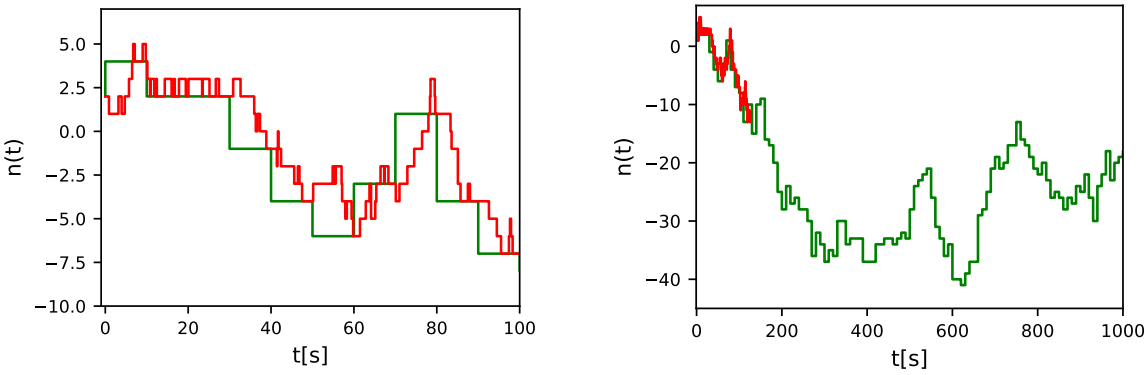

Figure 2.4: Compare between two of the samples used for the analysis. The red trajectory belongs to a sample with sampling time interval $\Delta t_s = 1.129$. The green trajectory belong to a sample with sampling time interval $\Delta t_s = 10.0$.

Notice that when the number of data points M in the time series decreases, the minimum for the relative error moves to bigger values of Δt_s (figure 2.3). This happens because the sampled trajectory gets shorter, meaning we need the interval between samples to be larger for

[3] The growth in error bars is real, but it is not that extreme; it is highlighted by the logarithmic scale

the statistics to be good enough; otherwise, the resulting statistics for J_+ are simply too poor to reliably capture the dynamics. For $M = 10$, the rightward shift of the minimum causes it to merge with the error growth driven by the approximation bias, causing an increase in the minimum value of the relative error. Finally, observe that for $\Delta t_s \lesssim 3 \cdot 10^{-3}$, the error reaches exactly 1. Mathematically, this can only happen if the estimate $\hat{k}_+$ evaluates to 0, which occurs when no positive jumps are observed within the sampled window.

In conclusion, parameter estimation for a jump process via Maximum Likelihood Estimation relies on the path likelihood, which can be exactly computed provided the full trajectory is observed. When only discrete observations are available, the solution we propose consists of reconstructing the path from the sample as shown in Figure 2.1. This procedure yields estimates that depend heavily on the specific sampling protocol.
We can define an optimal sampling protocol, which is the one parametrized by a sampling time interval $\Delta t_s << \tau$, so that the approximate path is sufficiently similar to the real one, and a dataset size M sufficiently large to get enough statistics $\frac{T_{\text{obs}}}{\tau} = (M-1)\frac{\Delta t_s}{\tau} > 1$. Consequently, once we fix a $\Delta t_s << \tau$, the focus must shift on the trajectory duration T_{obs}, rather than the number of sampling points M, whose lower limit is just a consequence of the limit on T_{obs}.

Chapter 3

Two different models for gene expression

Thus far, Chapter 1 has explored the framework underlying parametric inference and model distinguishability for SDEs, while Chapter 2 has introduced the theoretical approximation to be employed for jump processes when the exact process propagator is unavailable. We now turn our attention to applying these theoretical tools to a concrete biological phenomenon.

The specific function of any given cell is dictated by the proteins it contains and gene expression is a fundamental process responsible for protein synthesis within a cell. In particular, the expression of a gene relies on discrete and random biochemical reactions that drive the production of mRNAs and proteins. These stochastic effects are of profound biological significance: because DNA is present in low copy numbers, statistical fluctuations do not average out, leading seemingly identical cells to exhibit unique characteristics [39].

To investigate these dynamics, we will analyze two stochastic models of gene expression, which are efficiently described in [4]. In this section, we apply the aforementioned theoretical frameworks to perform parametric inference and assess the distinguishability of these two models. This will allow us to determine the sampling protocol under which they can be effectively distinguished, thereby providing a robust mathematical criterion for selecting the most appropriate model to describe experimental data.
We begin by briefly outlining the dynamics of the two models.

3.1 Burst-like model for gene expression

We study a simple model of gene expression that aims to characterize the evolution of the number of proteins of a single type, denoted by x. Gene expression, which leads to variations in x, occurs through two main processes: transcription into mRNA molecules with a rate k_1 and translation of the mRNA into proteins with a rate k_2. Degradation rates for mRNA and proteins are defined by γ_1 and γ_2. Since protein production has been experimentally observed to occur in bursts [4], we can describe gene expression with a jump Markov process defined by two events: a birth event occurring at a rate k_1 with a jump size distributed according to the

geometric distribution

$$g(I) = \frac{1}{b}\left(\frac{b}{1+b}\right)^I$$

where $b = \frac{k_2}{\gamma_1}$; and a death event occurring at rate $\gamma_2 x$, where x is the number of proteins, with unitary jump amplitude. Formally, one can write

$$w(x \longrightarrow x + j) = \begin{cases} k_1\, g(j) & \text{if } j = I > 0, \\ \gamma_2 x & \text{if } j = -1 \end{cases} \tag{3.1}$$

This jump process can be approximately described through an effective Langevin-like equation [4]. In this framework, degradation is treated as a continuous process, an approximation that holds well in the high protein number regime, while protein production retains its discrete, burst-like description:

$$\dot{x}(t) = \delta - \gamma_2 x(t) + \Lambda(t), \tag{3.2}$$

where δ represents the basal protein production and $\Lambda(t)$ is a non-Gaussian white noise term modeling the protein bursts. In particular, $\Lambda(t)\Delta t = \sum_{k=1}^{n(t)} I_k$, where $n(t)$ denotes the number of burst events within the time interval $[t, t+\Delta t]$ and $\{I_k\}_{k=1,\dots,n(t)}$ is the sequence of positive jump amplitudes (Appendix I, Appendix J) [4]. Consequently, the deterministic drift accounts for the exponential degradation of proteins, while the stochastic term models the burst-like birth events. In this continuous limit, the amplitudes $\{I_k\}$ are drawn from the continuous exponential approximation of the geometric distribution, namely $f(I) = \frac{1}{b}e^{-I/b}$ (Appendix H). The propagator associated with the above SDE can be computed analytically as [4]

$$\begin{aligned} p(x,t) =& e^{-k_1 t}\delta(x-\xi_t) + \Theta(x-\xi_t)\frac{k_1\gamma_1}{k_2\gamma_2}(e^{\gamma_2 t}-1)e^{-k_1 t} \\ &\times \exp[-\frac{\gamma_1}{k_2}e^{\gamma_2 t}(x-\xi_t)] \\ &\times\ {}_1F_1(\frac{k_1}{\gamma_2}+1, 2; \frac{\gamma_1}{k_2}(e^{\gamma_2 t}-1)(x-\xi_t)), \end{aligned} \tag{3.3}$$

where ${}_1F_1$ is the confluent hypergeometric function and $\xi(t) = x_0 e^{-\gamma_2 t} + \frac{\delta}{\gamma_2}(1 - e^{-\gamma_2 t})$ is the solution to the deterministic part of equation 3.2 [4].

3.1.1 Simulation strategies for the burst-like model

To validate the parametric inference framework described in Chapter 2 and to perform the model distinguishability analysis against the competing model we are going to introduce in the following section, synthetic trajectories of the first model are generated using two distinct numerical strategies.

For the initial validation of the parametric inference framework, we simulate the exact microscopic birth-death process described by the transition rates in Eq. (3.1). This is achieved via the Gillespie algorithm [40]. In this framework, the state space is strictly discrete ($x \in \mathbb{N}$) and protein counts vary by integer steps: positive jumps represent geometrically distributed

bursts of production, while negative single-step jumps model individual degradation events. Testing our inference framework on these discrete trajectories allows us to evaluate its performance directly on the fundamental, microscopic description of the system.

When moving to the model distinguishability analysis, the need for a change in methodology arises: the competing model is not a jump process but a continuous one (to be precise, a CIR process). To ensure a fair comparison, the likelihood is computed using the exact analytical propagators in both cases. For our first model, the analytical propagator is derived from its effective Langevin-like approximation (Eq. 3.2). Consequently, to maintain mathematical consistency during the discrimination analysis, we must employ a numerical simulation that directly reflects this coarse-grained approximation, establishing a symmetrical comparison with the second model. Between consecutive bursting events, the protein concentration decays deterministically according to the continuous exponential drift. At exponentially distributed waiting times, the trajectory undergoes a positive, continuous jump sampled directly from the exponential distribution $f(I)$. This algorithm generates trajectories that perfectly match the mathematical assumptions under which the analytical propagator is derived, thereby allowing for a fair and coherent model comparison.

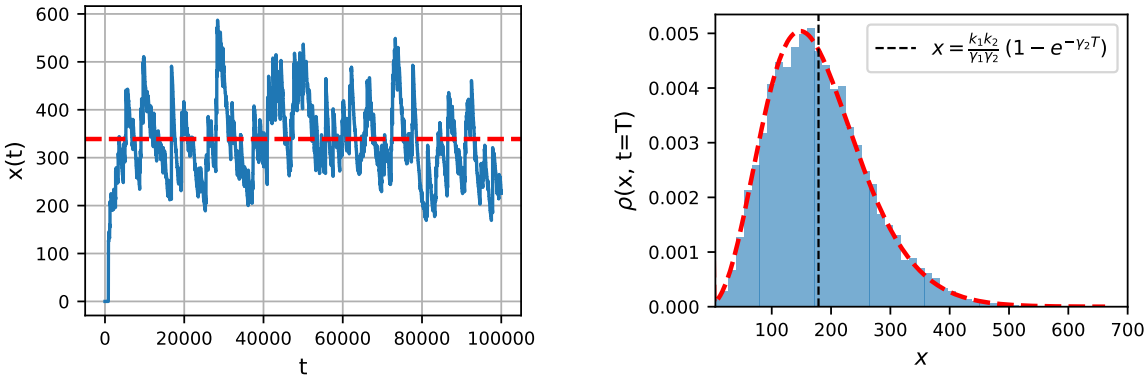

Figure 3.1: Jump process simulated through Gillespie algorithm from Eq. 3.1. The left plot shows a trajectory of the process (blue) and the average stationary value $\langle x \rangle = \frac{k_1 k_2}{\gamma_1 \gamma_2}$ with $\delta = 0$ (red dashed line). The right plot shows the histogram of x values (blue) and the distribution of the process (3.3) (red dashed).

3.2 Gaussian noise model for gene expression

An alternative model can be introduced if we consider Gaussian multiplicative white noise. This can be viewed as a coarse-grained approximation of the first model: it is obtained when information about the discrete and burst-like nature of protein production is lost.
Assume that mRNA degradation is faster than protein degradation $\gamma_1 >> \gamma_2$, in this case mRNA translates almost immediately: we can neglect the burst-like dynamics and assume a constant flux of proteins given by $\langle \Lambda(t) \rangle = \lambda = \frac{k_1 k_2}{\gamma_1}$ (Appendix J) [4]. Furthermore, given the large number of proteins, the Central Limit Theorem can be applied and fluctuations can be assumed proportional to $\sqrt{x}$. Then, the stochastic evolution of the protein concentration $x(t)$

can be described by the SDE [4]:

$$\dot{x}(t) = \lambda - \gamma_2 x(t) + \sqrt{Dx(t)}\eta(t), \tag{3.4}$$

where $\lambda \equiv \frac{k_1 k_2}{\gamma_1}$, $D \equiv \frac{\gamma_2 k_2}{\gamma_1}$ and $\eta(t)$ is a Gaussian white noise. The variance of this noise is state-dependent: $\langle\langle \sqrt{Dx(t)}\eta(t)\sqrt{Dx(t')}\eta(t')\rangle\rangle = 2Dx(t)\delta(t-t')$.
Notice that if we evaluate this variance at the stationary mean concentration $\langle x\rangle = \frac{k_1 k_2}{\gamma_1 \gamma_2}$, we find that $2D\langle x\rangle = 2k_1(k_2/\gamma_1)^2$, which exactly matches the variance of the burst-like noise (Appendix J).

Equation 3.4 is associated with the Fokker-Plank equation

$$\frac{\partial\, p(x,t)}{\partial t} = -\frac{\partial}{\partial x}[(\lambda - \gamma_2 x)p(x,t)] + D\frac{\partial^2}{\partial x^2}[x\, p(x,t)],$$

whose solution is given by [26]

$$p(x,t) = c\, e^{-(u+v)}\, (\frac{v}{u})^{\frac{q}{2}}\; I_q(2\sqrt{uv}), \tag{3.5}$$

where $c = \frac{\gamma_2}{D(1-e^{-\gamma_2 t})}$, $u = c\, x_0\, e^{-\gamma_2 t}$, $v = c\, x$, $q = \frac{\lambda}{D} - 1$ and I is the modified Bessel function of the first kind.
The stationary solution,

$$p^{\mathrm{st}}(x) = \frac{1}{\Gamma(\frac{\lambda}{D})}(\frac{\gamma_2}{D})^{\frac{\lambda}{D}}\, x^{\frac{\lambda}{D}-1}\, e^{-\frac{\gamma_2}{D}x},$$

is equal to the one we would have obtained solving the Master Equation for the jump process (Appendix K) [4]:

$$\frac{\partial p(x,t)}{\partial t} = -\frac{\partial}{\partial x}[(\delta - \gamma_2 x)p(x,t)] + k_1 \int_0^x f(x-y)\, p(y,t)\, dy - k_1\, p(x,t). \tag{3.6}$$

Note that the two models describe fundamentally different microscopic dynamics: the first model accounts for discrete burst-like production events and the second is a continuous coarse-grained approximation driven by Gaussian multiplicative noise. Despite that, both the discrete model and the continuous one yield the exact same stationary probability distribution when their macroscopic parameters are matched. Therefore, to analyze the microscopic mechanisms occurring inside the cell, one needs to look at the temporal dynamics.

The simulation of the Gaussian noise model is performed exactly from the process solution [22].

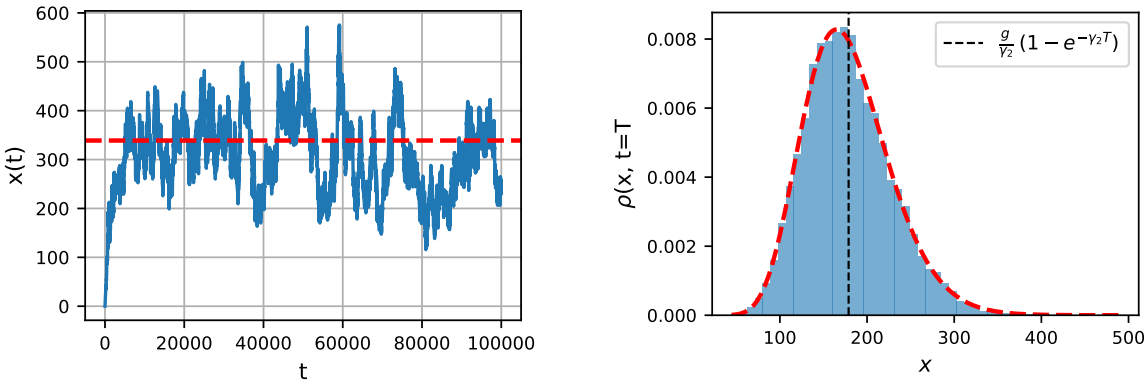

Figure 3.2: Gaussian noise model (3.4). The left plot shows a trajectory of the process (blue) and the average stationary value $\langle x \rangle = \frac{k_1 k_2}{\gamma_1 \gamma_2}$ (red dashed line). The right plot shows the histogram of x values (blue) and the distribution of the process (3.5).

Notice that Eq. 3.4 is the same demographic noise model we introduced in Chapter 1 as CIR process, with the mapping $k \to \gamma_2$, $\mu \to \frac{\lambda}{\gamma_2}$ and $D \to \sqrt{2D}$. So the parameter estimation can be performed in the same way as for the SDE in Eq. 1.6, using the true propagator of the process. The dependence of the relative errors from the sampling protocol is the one already discussed in Chapter 1.

3.3 Autocorrelation

An optimal sampling time interval for distinguishability is necessarily defined with respect to the autocorrelation times of the processes. Otherwise, the dependence of results from the values of model parameters would make the entire analysis inconclusive.
To correctly study the processes autocorrelation times, in this section we compute analytically the autocorrelation function for both processes and compare it to numerical results.
The autocorrelation function of the Gaussian noise model can be computed analytically starting from the SDE

$$dX_t = (\lambda - \gamma_2 X_t)dt + \sqrt{2DX_t}dW,$$

by employing standard conditional expectation techniques for Markov processes [18]. Taking the expectation value, we get

$$d\mathbb{E}[X_t] = (\lambda - \gamma_2 \mathbb{E}[X_t])dt$$

where we used $\mathbb{E}[dW] = 0$. Let us call $\mathbb{E}[X_t] = k$ and solve for k

$$\frac{d}{dt}k = \lambda - \gamma_2 k$$

$$e^{\gamma_2 t}\frac{d}{dt}k + \gamma_2 k e^{\gamma_2 t} = \lambda e^{\gamma_2 t} \qquad \Rightarrow \qquad \frac{d}{dt}(k e^{\gamma_2 t}) = \lambda e^{\gamma_2 t}$$

Integrating and substituting $k = \mathbb{E}[X_t]$, we get:

$$\mathbb{E}[X_t] = \frac{\lambda}{\gamma_2}(1 - e^{-\gamma_2 t}) + x_0 e^{-\gamma_2 t}$$

with $t_0 = 0$.
Given $s \leq t$:

$$\mathbb{E}[X_t|X_s = y] = \frac{\lambda}{\gamma_2}(1 - e^{-\gamma_2(t-s)}) + y\, e^{-\gamma_2(t-s)},$$

where $\mathbb{E}[X_t|X_s = y] = \int_0^\infty x\, p(x,t|y,s)\, dx$. So,

$$\begin{aligned}\mathbb{E}[X_t X_s] &= \int_0^\infty dy\; y\; P_4(y,s) \int_0^\infty x\, p(x,t|y,s)\, dx = \\ &= \int_0^\infty dy\; y\; P_4(y,s)[\frac{\lambda}{\gamma_2}(1 - e^{-\gamma_2(t-s)}) + y\, e^{-\gamma_2(t-s)}],\end{aligned}$$

where $p(y,s|x_0 = 0,0) = \frac{c_s^a}{\Gamma(a)} y^{a-1} e^{-c_s y}$ is the time dependent distribution for the Gaussian noise model imposing $x_0 = 0$ in Eq. 3.5, with $a = \frac{\lambda}{D}$, $b = \frac{\gamma_2}{D}$ and $c_s = \frac{b}{1-e^{-\gamma_2 s}}$. Using $\int_0^\infty dy\; y\; \frac{c_s^a}{\Gamma(a)} x^{a-1} e^{-c_s y} = \mathbb{E}[X_s] = \frac{a}{b}(1 - e^{-\gamma_2 s})$ with $x_0 = 0$ and solving the integral in the second term, we got

$$\mathbb{E}[X_t X_s] = \frac{a^2}{b c_s}(1 - e^{-\gamma_2(t-s)}) + \frac{a(a+1)}{c_s^2} e^{-\gamma_2(t-s)}.$$

At steady-state $c_s \xrightarrow{s\to\infty} b$. Substituting a, b with λ, γ_2 and imposing $t - s = \Delta t$ one obtains

$$\mathbb{E}[X_{t+\Delta t} X_t] = \frac{\lambda^2}{\gamma_2^2} + \frac{\lambda D}{\gamma_2^2} e^{-\gamma_2 \Delta t},$$

where $\frac{\lambda D}{\gamma_2^2}$ is the stationary value for the variance of X_t. Then, the autocorrelation function for the Gaussian noise model at stationarity is given by

$$C(\Delta t) = \frac{\mathbb{E}[X_{t+\Delta t} X_t] - \mathbb{E}[X_t]^2}{\mathbb{E}[X_t^2] - \mathbb{E}[X_t]^2} = e^{-\gamma_2 \Delta t}.$$

For the burst-like non-Gaussian model, we can perform the same calculation starting from

$$dX_t = (\delta - \gamma_2 X_t)dt + dK_t,$$

where $K_t = \sum_{k=1}^{n(t)} I_k$, with $n(t)$ number of burst events in $[0,t]$ and $\{I_k\}$ independent identically distributed burst amplitudes. Taking the expectation value,

$$d\mathbb{E}[X_t] = (\delta - \gamma_2 \mathbb{E}[X_t])dt + \lambda\, dt,$$

where we used $\mathbb{E}[dK_t] = \lambda\, dt$. Solving the equation for $\mathbb{E}[X_t]$ we get

$$\mathbb{E}[X_t] = \frac{\delta + \lambda}{\gamma_2}(1 - e^{-\gamma_2 t}) + x_0 e^{-\gamma_2 t}$$

and

$$\mathbb{E}[X_t | X_s = y] = \frac{\delta + \lambda}{\gamma_2}(1 - e^{-\gamma_2 (t-s)}) + y\, e^{-\gamma_2 (t-s)}.$$

Then, using the tower rule

$$\mathbb{E}[X_t X_s] = \mathbb{E}[X_s\ \mathbb{E}[X_t | X_s]],$$

and defining $t - s = \Delta t$, at steady-state, the 2-point correlation function becomes

$$\mathbb{E}[X_{t+\Delta t} X_t] = \mathbb{E}[X_t^2]\, e^{-\gamma_2 \Delta t} + (\frac{\delta + \lambda}{\gamma_2})^2\, (1 - e^{-\gamma_2 \Delta t}).$$

Finally, the correlation function is again

$$C(\Delta t) = \frac{\mathbb{E}[X_{t+\Delta t} X_t] - \mathbb{E}[X_t]^2}{\mathbb{E}[X_t^2] - \mathbb{E}[X_t]^2} = e^{-\gamma_2 \Delta t}.$$

The analytical expression for the autocorrelation time for both gene expression models is given by

$$\tau = \frac{1}{\gamma_2}.$$

This is the normalization value used for the x-axis plots in section 3.5. So at $x = 1$ we are observing the results for the sampling time interval Δt_s being equal to the autocorrelation time of the process τ.

In figure 3.3, we compare the analytical results for the stationary autocorrelation function with the numerical plots, each computed through [12]

$$C = \frac{\mathbb{E}[X_t X_{t+\Delta t}] - \mathbb{E}[X_t]^2}{\mathrm{Var}(X_t)},$$

where Δt are multiples of the simulation time step.

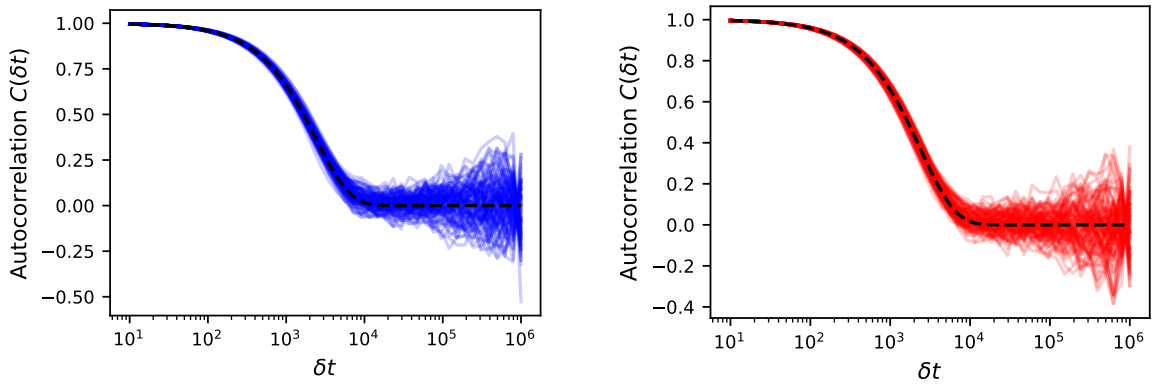

Figure 3.3: Plot of numerical autocorrelation function for 100 trajectories (blue/red) and theoretical plot of the analytical stationary autocorrelation function (black dashed line). On the left: simulation for the Gaussian noise model. On the right: simulation for the burst-like process.

3.4 Parametric inference for the jump model of gene expression

Now we focus on parametric inference for the burst-like model of gene expression presented in section 3.1.

Although the underlying biological process is governed by four parameters, the model in Eq. 3.1 is a single-variable approximation that tracks only the protein concentration. Because the mRNA dynamics are left implicit, we can directly estimate only the two explicit parameters governing the macroscopic dynamics (the burst frequency k_1 and the degradation rate γ_2). The remaining two parameters (k_2 and γ_1) are structurally unidentifiable on their own and can only be inferred as a composite ratio, representing the mean burst size ($b = k_2/\gamma_1$).

The estimation of model parameters is performed through MLE. To compute the likelihood for the burst-like model two distinct approaches can be adopted. The exact likelihood can be derived again using the analytical propagator (Eq. 3.3) of the Langevin-like formulation of the process, in Eq. 3.2, as done in Chapter 1 for OU, EN and CIR processes. This is computationally inconvenient because of the closed-form propagator shape, which is highly complex and prone to generating numerical overflow.
The alternative relies on the jump process definition of the dynamics, in Eq. 3.1, to construct the path likelihood as described in Chapter 2. In the following analysis, we will focus on this alternative approach to demonstrate a first application of the theory presented in Chapter 2.

Then, let us compute the likelihood exploiting Eq. 2.7 and adapting it to describe the gene expression model where the total exit rate is defined by

$$W(x) = \sum_{I=1}^{\infty} w(x \to x + I) \; + \; w(x \to x - 1) = k_1 + \gamma_2 x.$$

The log-likelihood takes the form

$$l_{[t_0,t_J]} = \frac{1}{J}\sum_{i=0}^{J-1}(-\Delta t_i(k_1+\gamma_2 x_i) + \log(w(x_i \to x_{i+1})). \tag{3.7}$$

Deriving with respect to k_1 one gets

$$\partial_{k_1} l_{[t_0,t_J]} = \frac{1}{J}\sum_{i=0}^{J-1}(-\Delta t_i) + \frac{1}{J}\sum_{i=0}^{J_+-1}\frac{1}{k_1},$$

where, in the second summand, the derivative of the terms are non-zero when in Eq. 3.7 $w(x \to x+j) = w(x \to x+I)$, so we sum up to the number of positive jumps J_+. Finally, imposing the maximization we obtain the estimate of k_1 as

$$\hat{k}_1 = \frac{J_+}{\sum_{i=0}^{J-1}\Delta t_i}. \tag{3.8}$$

To estimate γ_2 we perform similar computations obtaining:

$$\partial_{\gamma_2} l_{[t_0,t_J]} = \frac{1}{J}\sum_{i=0}^{J-1}(-\Delta t_i x_i) + \frac{1}{J}\sum_{i=0}^{J_--1}\frac{1}{\gamma_2}$$

$$\hat{\gamma}_2 = \frac{J_-}{\sum_{i=0}^{J-1}\Delta t_i x_i}. \tag{3.9}$$

Finally, we compute the parameter $b = \frac{k_2}{\gamma_1}$:

$$\partial_b l_{[t_0,t_J]} = \frac{1}{J}\sum_{i=0}^{J_+-1}\frac{1}{k_1\frac{1}{b}(\frac{b}{1+b})^{I_i}}(-\frac{k_1}{b^2}(\frac{b}{1+b})^{I_i} + \frac{k_1}{b}\,I_i\,(\frac{b}{1+b})^{I_i-1}\,\frac{1}{(1+b)^2}) = \frac{1}{J}\sum_{i=0}^{J_+-1}(-\frac{1}{b}+\frac{I_i}{b(1+b)})$$

$$\hat{b} = \frac{1}{J_+}\sum_{i=0}^{J_+-1} I_i - 1 \tag{3.10}$$

which is, as expected, the experimental counterpart of $b = \langle I\rangle - 1$, as the mean of the geometrical distribution $g(I)$ is $\langle I\rangle = b+1$.

As stated in Chapter 2, the direct application of this analytical framework to experimental data is unfeasible, as the real waiting times Δt_i and number of transitions J remain hidden. Therefore, to perform parameter estimation, we must restore the approximation, consisting of replacing waiting times Δt_i by multiples of the sampling time $n_{s,i}\,\Delta t_s$ and the number of transitions J by the observed number of transitions J^s (see Chapter 2). Consequently, the estimators approximate to

$$\tilde{\hat{k}}_1 = \frac{J^s_+}{(M-1)\Delta t_s} \tag{3.11}$$

$$\tilde{\hat{\gamma}}_2 = \frac{J^s_-}{\sum_{i=0}^{J^s-1} n_{s,i}\,\Delta t_s\,x_i} \tag{3.12}$$

$$\tilde{\hat{b}} = \frac{1}{J_+^s} \sum_{i=1}^{J_+^s} I_i - 1 \tag{3.13}$$

where $\{x_i\}$ identifies the time series of the sampled points and $\{I_i\}$ are the observed bursts. From these estimates, we can plot the relative errors $\varepsilon_{\hat{k}} = \frac{|\hat{k}-k|}{k}$, where $k = k_1, \gamma_2, b$, with respect to Δt_s as we did in the previous chapter.

In the following graphs, the trends for the relative error of $\hat{k}_1$ are plotted with respect to the sample time interval Δt_s, normalized by the average value between two consecutive jumps Δt_{jumps} which is given by

$$\Delta t_{\text{jumps}} = \frac{1}{k_1 + \gamma_2 \langle x \rangle},$$

where $\langle x \rangle = \frac{k_1 k_2}{\gamma_1 \gamma_2}$.

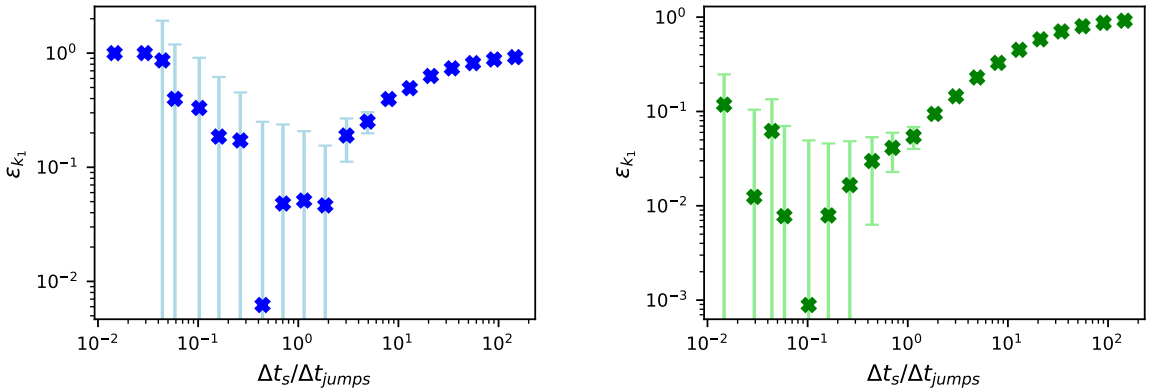

Figure 3.4: Plot of the relative error of k_1 with respect to the sampling time Δt_s. Left: the number of points for each sample is $M = 10$. Right: the number of points for each sample is $M = 1000$. The error bars are computed as $\frac{\sigma_{\hat{k}_1}}{k_1 \sqrt{n_{\text{traj}}}}$.

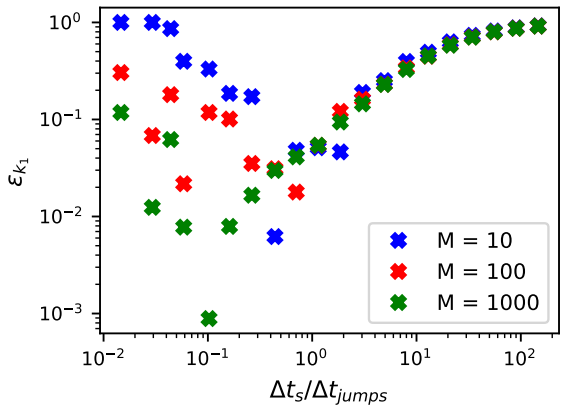

Figure 3.5: Comparison of plots for the relative error of k_1 with respect to the sampling time Δt_s, computed using different sample sizes M.

As expected, the trend of the error is the same as the one observed and discussed in the previous chapter. Again we observe that the error increases with $\Delta t_s > \Delta t_{\text{jumps}}$ and also for

$\Delta t_s << \Delta t_{\text{jumps}}$.

The same trend can be noticed for the error of γ_2.

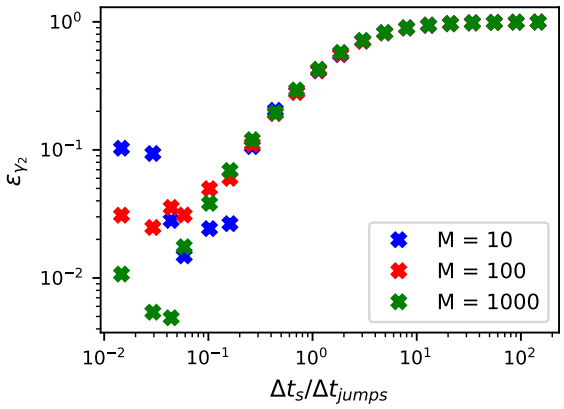

Figure 3.6: Plots for the relative error of γ_2 with respect to the sampling time Δt_s, for different values of the sample size M.

The difference between the error trends of γ_2 and k_1 stems from the difference in their respective total event rates. While the frequency of positive jumps is governed by the constant rate k_1, the frequency of negative jumps averages $\gamma_2\langle x\rangle = bk_1$. Since $b > 1$, degradation events are significantly more frequent. As a consequence, the critical lower bound of the observation window Δt_s, below which no events are recorded and the estimation error explodes, is much smaller for γ_2.

Finally, the error trend for the burst size estimate $\hat{b}$ in Figure 3.7 exhibits a slightly distinct behavior, as it relies on averaging the amplitudes of positive jumps. At large sampling intervals ($\Delta t_s >> \Delta t_{\text{jumps}}$), the error systematically increases, eventually exceeding 100% due to undetected events. On the other hand, at very fine sampling resolutions ($\Delta t_s \ll \Delta t_{\text{jumps}}$), the estimate becomes highly unstable. Holding M constant while decreasing Δt_s drastically shortens the total observation window $T_{\text{obs}} = (M-1)\,\Delta t_s$ and leads to fewer observed jumps, causing the error bars to increase. Indeed, for $\Delta t_s/\Delta t_{\text{jumps}} \lesssim 10^{-1}$, several trajectories record no jumps at all, effectively shrinking the number of trajectories relevant to the analysis. In extreme cases, such as when $M = 10$ and $\Delta t_s/\Delta t_{\text{jumps}} \lesssim 10^{-2}$, the number of valid trajectories drops to zero, making the estimation impossible.

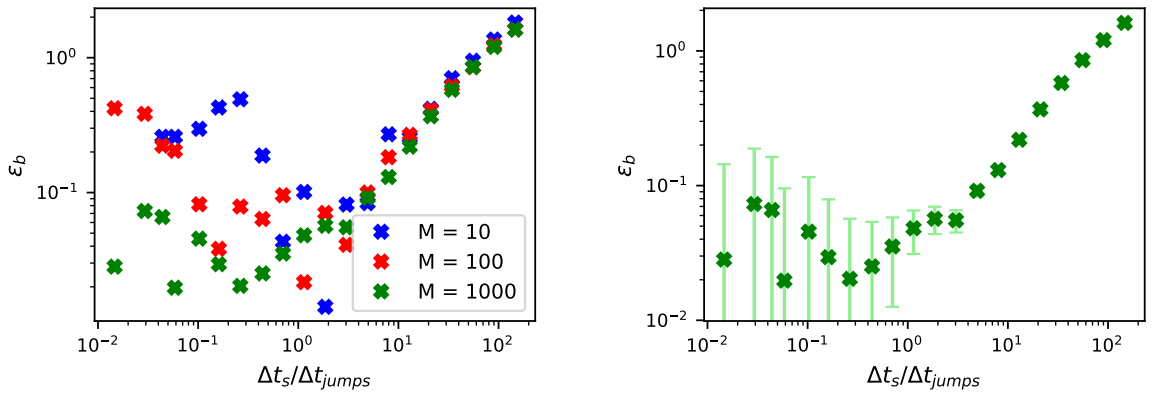

Figure 3.7: Plot of the relative error of $\hat{b}$ with respect to the sampling time Δt_s. Left: plots realized for different values of the sample size. Right: the number of points for each sample is $M = 1000$. The error bars are computed as $\frac{\sigma_{\hat{b}}}{b\sqrt{n_{\text{traj}}}}$.

3.5 Model distinguishability

Although the two analyzed models yield the same macroscopic dynamics (same stationary distribution), they are fundamentally different at the microscopic level. Understanding the microscopic dynamics of a system is crucial to link the observed process to its underlying physical causes. For example, assuming a burst-like protein production implies that proteins are synthesized in bulk over short periods. This could be decisive in pharmacological studies. Indeed, under these conditions, an effective drug aimed at modulating a specific protein must not only lower the mean of the protein distribution but also suppress sudden stochastic fluctuations. This is a critical requirement that would be entirely overlooked if the system were described using the continuous Gaussian noise model [41].

To address this need to distinguish the models at the microscopic level, the aim of this section is to establish a range of values for the sampling parameters allowing us to identify data generated by one of the two models.
Note that while likelihood approximations can give good results for parameter estimation, comparing likelihoods across models with the same stationary distribution requires huge precision [12]. Therefore, in this section, model discrimination is performed comparing the two models likelihoods computed with Eq. 1.9, since the exact propagator can be analytically derived for both models.
We evaluated the probability P of correctly identifying the data-generating model. Data were generated from the models mapping the parameters defining the models as $f \equiv \frac{k_1 k_2}{\gamma_1}$ and $D \equiv \frac{\gamma_2 k_2}{\gamma_1}$.
In Figure 3.8 we observe the results for model discrimination probability with respect to variations of the sampling protocol.

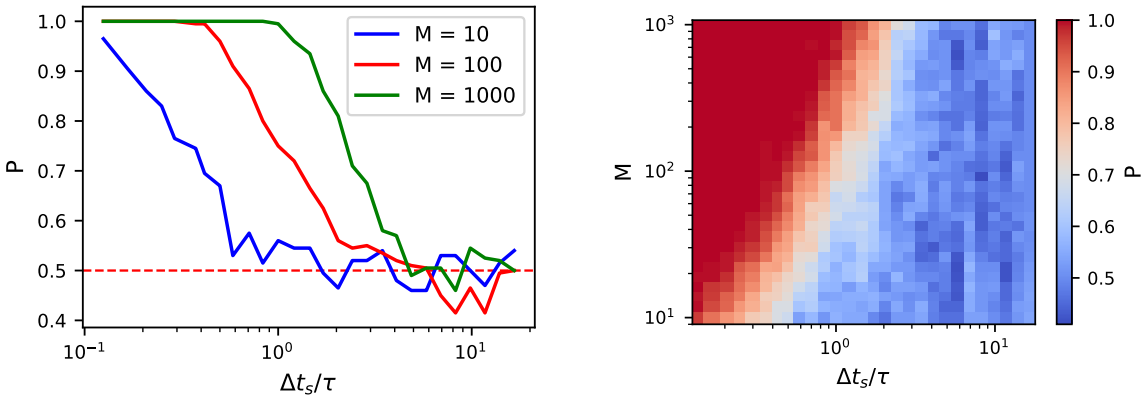

Figure 3.8: Model discrimination probability. Left: P trend with respect to the sampling time interval Δt_s, with fixed values of the sample size M. Right: heatmap of the discrimination probability P, with respect to variations of both Δt_s and M. The x-axis is normalized by the autocorrelation time τ for the two processes (see Section 3.3).

From figure 3.8, we can observe that the probability of correctly identifying the model decreases with the sampling frequency. This behavior is attributed to the decorrelation of the sampled data points. In fact, the two competing models share the same stationary distribution and, therefore, can be distinguished only through their dynamics. As the sampling intervals increase, the propagator relaxes toward the stationary distribution and the dynamics begin to disappear. For values of $\Delta t_s > \tau$, the data are weakly correlated. In this range of values, the number of data points in the sample makes the difference: for large M statistical noise is reduced and the inference procedure is able to detect the small differences still present between the two distributions. Finally, for values of $\Delta t_s >> \tau$, the probability drops to 0.5, making the two models completely indistinguishable: we move from a time series to random sampling within a stationary distribution common to both processes.

In this chapter, we investigated two alternative stochastic formulations to describe the dynamics of gene expression: a non-Gaussian Markov process capturing burst-like protein production, and a multiplicative Gaussian noise model structured as a Cox-Ingersoll-Ross (CIR) process.
When their macroscopic parameters are appropriately matched, both frameworks yield identical stationary distributions and a shared characteristic relaxation timescale $\tau = 1/\gamma_2$. This implies that static experimental data are insufficient to identify the underlying microscopic mechanisms, making the analysis of temporal dynamics mandatory.
By implementing the high-frequency sampling approximations developed in Chapter 2, we derived the Maximum Likelihood Estimators (MLE) for the jump model. The parametric inference analysis revealed again two distinct physical limitations bound to the sampling protocol, the same retrieved in Chapter 2. Nevertheless, it is worth emphasizing that, within certain sampling time regimes, the likelihood approximation presented here works very well for parametric inference despite the extreme simplification it entails.
Finally, we performed model distinguishability exploiting the two models propagators as done

in Chapter 1 and we successfully mapped the operational boundaries within which the two processes can be correctly identified.
These results provide a rigorous and systematic guideline for optimizing experimental design, establishing information about the sampling requirements to differentiate microscopic-detailed descriptions (the jumping process) from coarse-grained models (the SDE process).

Chapter 4

Chemical reactions

Let us finally shift our focus to systems where the framework developed in Chapter 2 is fully exploited. In particular, in the absence of useful alternatives, the approximate path likelihood in Eq. 2.4 (Chapter 2) will also be used to conduct model distinguishability analysis, probing its strengths and limitations across simple chemical reaction networks.
In what follows, after a brief overview of the context, we will investigate model discrimination across two different cases.

A chemical system is fully characterized by a set of $\rho = 1...N_\rho$ chemical reactions between its N_X species $\{X_i\}_{i=1...N_X}$. A set of reactions can be represented by

$$\sum_{i=1}^{N_X} \nu_{i,\rho}^{+} X_i \underset{k_\rho^-}{\overset{k_\rho^+}{\rightleftharpoons}} \sum_{i=1}^{N_X} \nu_{i,\rho}^{-} X_i,$$

where $\nu_{i,\rho}^{\pm}$ is the stoichiometric coefficient for species X_i in the forward and backward reaction ρ and $k_\rho^{\pm}$ are the forward and backward rate constants [27]. Every single reaction can be considered as a distinct, essentially instantaneous physical event which modifies the population of involved molecules. Note that, even if every single reaction were a deterministic event, when considering a chemical system, we do not take into account microscopic positions and velocities of the molecules or their interactions with the environment. As a consequence, molecular populations in a chemically reacting system are characterized as integer variables that evolve stochastically, driven by reaction rates [40]. In particular, for chemical reactions, the transition rates can be derived from the stochastic mass-action law, which computes all possible reactant combinations [40]:

$$w(\vec{n} \to \vec{n} \pm \vec{\delta}_\rho) = k_\rho^{\pm} \prod_{i=1}^{N_X^\rho} \prod_{m=0}^{\nu_{i,\rho}^{\pm}-1} (n_i - m), \tag{4.1}$$

where $\vec{n}$ is the system state, $\vec{\delta}_\rho$ is the vector updating the state for reaction ρ, N_X^ρ is the subset of species involved in reaction ρ. The rate constant $k_\rho^{\pm}$ depends on the temperature and volume of the system, and on the activation energy for the reaction [42].

Understanding the reactions that govern a given dynamics is essential for predicting how

macroscopic outcomes change when temperature, volume, or initial conditions are altered. In pharmacology, for example, a detailed understanding of the model governing molecular dynamics is necessary in order to alter it in a controlled manner [43]. Then, the model discrimination framework is particularly interesting when applied in this real-world context, allowing us to define a sampling protocol capable of distinguishing between two different microscopic processes leading to the same macroscopic outcome.
To study chemical reactions we can use jump processes and perform inference using the framework introduced in Chapter 2. This is particularly useful when a closed-form expression for the process propagator is not available and the exact likelihood cannot be computed.

4.1 Simple vs. Autocatalytic Reactions

A compelling first example of non-trivial distinguishability is the macroscopic behavior shared by two microscopically different reactions:

$$A \underset{k_-}{\overset{k_+}{\rightleftharpoons}} B \qquad (M1),$$

where A molecules can spontaneously transform into B; and the autocatalytic reaction

$$2A \underset{k_-}{\overset{k_+}{\rightleftharpoons}} A + B \qquad (M2),$$

where B production is triggered by an $A - A$ encounter. For both models, let us focus on the dynamics of just one species (A) and define $n_A = n$ and $n_B = N - n$.
The reaction rates are derived through the mass-action law in Eq. 4.1. For M1, they are defined by

$$w(n \to n \mp 1) = \begin{cases} k_+ \, n \\ k_- \, (N - n) \end{cases}$$

For M2 reaction rates are non-linear and read

$$w(n \to n \mp 1) = \begin{cases} k_+ \, n(n-1) \\ k_- \, n(N - n) \end{cases}$$

As a reference value for the sampling time in the plots that follow, we will use the average time between two consecutive jumps of the molecular dynamics at steady-state. This is the most natural choice, as it weights the temporal scales present in the process according to the stationary distribution.
Let us derive an analytical expression for the average time between consecutive jumps for each model.

- For M1 we can define a total jumping rate at steady-state as

$$w^* = k_+ \, n^* + k_- \, n_B^*,$$

where n^* and n_B^* are the population values at steady-state

$$n^* = N\frac{k_-}{k_+ + k_-}$$

and

$$n_B^* = N\frac{k_+}{k_+ + k_-}.$$

Thus, one obtains $\Delta t_{\text{jumps}} = \frac{1}{w^*} = \frac{k_+ + k_-}{2Nk_+k_-}$.

- For M2 the total jumping rate at steady-state is defined by

$$w^* = k_+ n^*(n^* - 1) + k_- n^*(N - n^*).$$

Since we are at steady-state and the system is closed, we assume $k_+ n^*(n^* - 1) = k_- n^*(N - n^*)$. Then

$$n^* = \frac{k_- N + k_+}{k_+ + k_-},$$

$$n^*(n^* - 1) = \frac{k_- N + k_+}{k_+ + k_-}\,\frac{k_-(N-1)}{k_+ + k_-}$$

and

$$n^*(N - n^*) = \frac{k_- N + k_+}{k_+ + k_-}\,\frac{k_+(N-1)}{k_+ + k_-}$$

$$w^* = \frac{k_- N + k_+}{k_+ + k_-}\,\frac{2k_+k_-(N-1)}{k_+ + k_-}.$$

So finally, the time interval between jumps at steady-state is $\Delta t_{\text{jumps}} = \frac{1}{w^*}$.

Although the macroscopic outcome of both models may appear identical, their behavior at the stochastic level diverges significantly. The non-linearity in the autocatalytic process acts as an amplifier for the chemical fluctuations. Given that autocatalytic reactions are fundamental drivers of cellular processes, discriminating the correct mechanistic model is essential: relying only on the macroscopic outcome would obscure the true noise profile that dictates biological decision-making and cell fate [44] [45] [46].

Due to the non-linearity introduced by M2, a closed-form propagator for molecule A dynamics cannot be analytically derived from the Master Equation. Therefore, to determine the discrimination probability we compare the two likelihood functions computed from the approximate formula in Eq. 2.4.
The likelihood is constructed from samples drawn from the simulations of the two models and rate constants estimated for each model via MLE (see Figure 1.8). Parameter estimation is performed using the same framework presented in Chapter 2. In particular, the exact parameter estimates are defined by

$$\hat{k}_- = \frac{J_+}{\sum_{i=0}^{J-1} \Delta t_i (N - n_i)} \qquad\qquad \hat{k}_+ = \frac{J_-}{\sum_{i=0}^{J-1} \Delta t_i \, n_i}$$

for M1,

$$\hat{k}_- = \frac{J_+}{\sum_{i=0}^{J-1} \Delta t_i \, n_i(N - n_i)} \qquad \hat{k}_+ = \frac{J_-}{\sum_{i=0}^{J-1} \Delta t_i \, n_i(n_i - 1)}$$

for M2, where $J_\pm$ is the number of jumps $n \to n \pm 1$. In the approximate estimates that we used, the exact number of transitions J is replaced by the observed one J^s and the waiting times $\{\Delta t_i\}$ are approximated by multiples of the sampling time $\{n_{s,i}\Delta t_s\}$:

$$\tilde{\hat{k}}_- = \frac{J^s_+}{\sum_{i=0}^{J^s-1} n_{s,i}\Delta t_s(N - n_i)} \qquad \tilde{\hat{k}}_+ = \frac{J^s_-}{\sum_{i=0}^{J^s-1} n_{s,i}\Delta t_s \, n_i}$$

for M1,

$$\tilde{\hat{k}}_- = \frac{J^s_+}{\sum_{i=0}^{J^s-1} n_{s,i}\Delta t_s \, n_i(N - n_i)} \qquad \tilde{\hat{k}}_+ = \frac{J^s_-}{\sum_{i=0}^{J^s-1} n_{s,i}\Delta t_s \, n_i(n_i - 1)}$$

for M2.

Notice the jump amplitude for both M1 and M2 is 1 because each reaction event increase or decrease the number of A molecule by a single element. When the trajectory is subsampled with a sampling time larger than Δt_{jumps}, the number of jumps we detect from the sample is underestimated, so the amplitude between two consecutive points of the time series can be $\Delta n > 1$. When this is the case, the algorithm assumes that $|\Delta n|$ jumps of amplitude 1 occurred at equally distributed times. However, this algorithm still does not account for cases where both positive and negative jumps have occurred between one sampling point and the next. In practice, the algorithm assumes that the trajectory is taking the minimum possible path between two sampling points (see Figure 4.1).

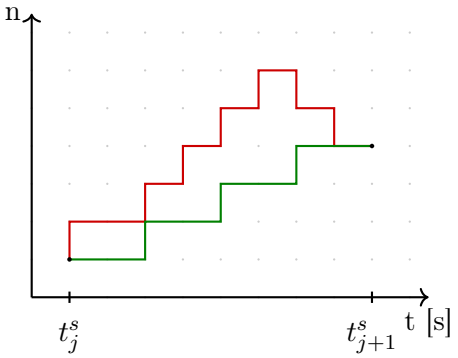

Figure 4.1: Schematic representation of the intermediate jump imputation. The red line represents a true unobserved stochastic trajectory, while the green line shows the uniform jump approximation assumed by the algorithm to connect the two sampled points at t^s_j and t^s_{j+1}.

Define P_i as the probability of correctly identifying the generating model when model Mi was used to generate the analyzed synthetic data. Define P as the discrimination probability not depending on the generating model, since it is computed as the mean $\frac{P_1+P_2}{2}$.

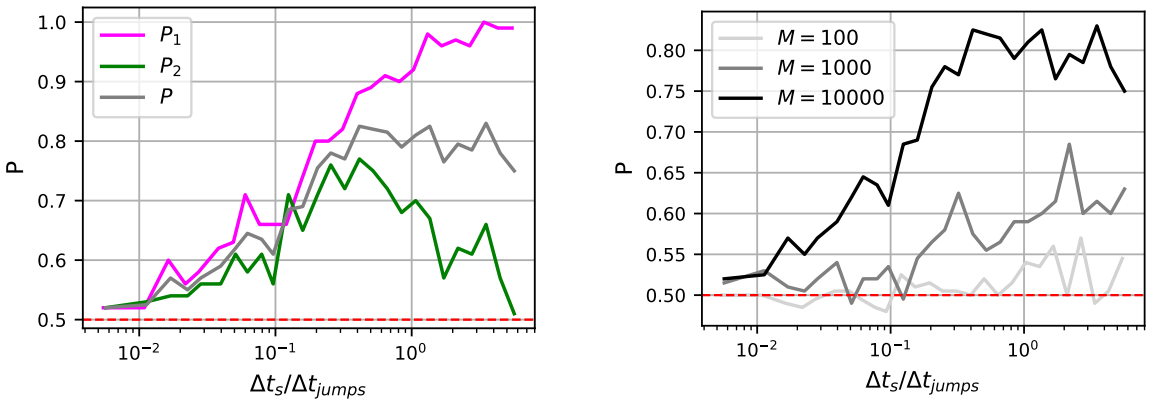

Figure 4.2: Discrimination probability trend with respect to variations of the sampling time Δt_s. The subscript $i = 1, 2$ of P_i refers to the generating model, while P is computed as the average of P_1 and P_2. In the left graph $M = 10000$ for each trend. The right graph represents P with respect to the sampling time Δt_s, for different fixed values of the sample size M.

Looking at the left plot in Figure 4.2, for high sampling frequencies, the discrimination probabilities stabilize slightly above 0.5, making the two models nearly indistinguishable. For this sampling time, indeed, the total observational time $T_{\text{obs}} = (M-1)\Delta t_s$ is very short. Up to this time a tiny number of jumps, given by $\frac{T_{\text{obs}}}{\Delta t_{\text{jumps}}} = (M-1)\frac{\Delta t_s}{\Delta t_{\text{jumps}}} \sim 10$, have occurred and the dynamics is hardly recognizable. For higher values of the sampling time interval ($\frac{\Delta t_s}{\Delta t_{\text{jumps}}} \sim 10^{-1} - 10^{0}$), the generating model can be correctly identified. When $\frac{\Delta t_s}{\Delta t_{\text{jumps}}}$ reaches values around 10^0, we notice a strong bias: results depend strongly on the generating model. In particular, the linear model (M1) is preferred, data are almost always correctly identified. On the other hand P_2 decreases, until the values reach a point where model M2 can no longer be identified. This behavior can be attributed to the subsampling procedure: when $\Delta t_s \gtrsim \Delta t_{\text{jumps}}$ the true number and timing of jumps are no longer available from the sample. This approximate information favors M1, causing the observed bias (Appendix L). This biased behavior is a consequence of the fact that, for $\Delta t_s > \Delta t_{\text{jumps}}$, the approximate paths the algorithm is considering (see Figure 4.1) tend to hide the sharp increases and decreases typical of a non-linear model, smoothing them into a steady trend more characteristic of a linear one. The indistinguishability observed for small values of Δt_s is even more visible for smaller values of the sample size M. Indeed, in the right plot of Figure 4.2, the unbiased probability P reaches smaller maximum values as M decreases. For $M = 100$, it barely move from the indistinguishable condition and for $\frac{\Delta t_s}{\Delta t_{\text{jumps}}} \sim 10^{-2}$ it stabilizes to 0.5, as no jumps are detected.

As a conclusion, to optimize the discrimination between models, the best choice is to sample the trajectories with $\Delta t_s < \Delta t_{\text{jumps}}$, but with values high enough to detect $\gtrsim 100$ jumps, accordingly with the sample size. Notice that in this context the distinguishability is more limited and requires larger values of M (Figure 4.3). Indeed, while in Chapter 3 for $M = 100$ we managed to find a Δt_s region where the two gene expression models were distinguishable, here $M = 100$ does not allow it.

Furthermore, the analysis just concluded highlights the intrinsic limitations of the approximation developed in Chapter 2 and applied here to unit-amplitude jump models. The observed discrimination bias is a direct consequence of these methodological limitations: in its current formulation, the algorithm is unable to reconstruct trajectories other than the minimal path between two consecutive sampling points. Consequently, when non-linear models are considered, this approach performs optimally only within a strictly limited regime.

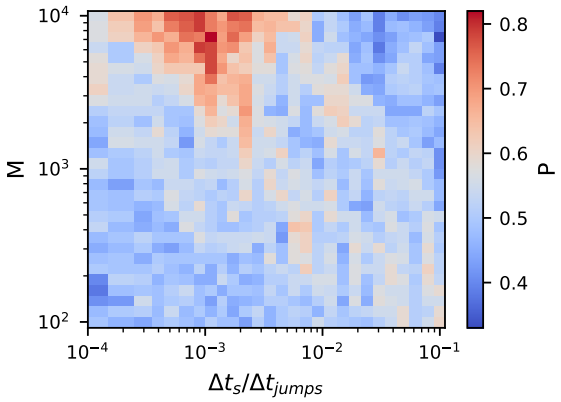

Figure 4.3: Heatmap of model distinguishability P. The red zone defines the sampling parameter sets $(\Delta t_s, M)$ allowing discrimination between the two models.

Conclusions

In this work, we focused on the problem of parametric inference by adopting a model-based approach and exploiting Maximum Likelihood Estimation, as well as the issue of model discrimination, which defines the probability of identifying the model that best describes a specific dataset. These questions were addressed and answered within the context of stochastic processes. We investigated the limits of dataset interpretability that arise from the discrete nature of real-world data, by applying parametric inference methods and model distinguishability frameworks to synthetic datasets under our control. Varying the dataset parameters (e.g. dataset size M and sampling time interval Δt_s) we were able to define a range for these parameters that maximizes data interpretability, allowing for an optimized inference of parameters and model discrimination.
The tool used to conduct these analyses is the likelihood function, which for a discrete set of data can be computed exactly (Eq 1.9 for SDEs and Eq. 2.2 for jump processes) only when a closed-form propagator of the process is available. Since this is not always the case, in this work we reviewed the approximate framework introduced in [12] to handle SDEs and developed an approximate framework for jump processes.

Chapter 1 focused on the review of some of the methods applied in [12]. We obtained some of the results highlighted in [12], such as the fact that the error in parameter estimation heavily depends on their physical nature: D estimation is optimized for small values of the sampling time $\Delta t_s < \tau$, k estimation error is minimized for $\Delta t_s \sim \tau$, while μ is better estimated for uncorrelated data $\Delta t_s > \tau$. We also demonstrated - in addition to what was already stated in the original paper - that the approximate method developed in [12] that exploits the Girsanov theorem for the likelihood computation works well for parameter estimation as the error trends obtained from the maximization of the exact likelihood and the approximate likelihood are very similar, suggesting that the approximation only slightly affects inference and the error variability is in fact due to the physical nature of the parameters.

In Chapter 2 we addressed the issue of inference for jump processes when a closed-form propagator is not available. We developed an approximate framework and defined its limitations in parameter estimation. In particular, Maximum Likelihood Estimation applied to the approximate likelihood in Eq. 2.4 works well in a range limited by a right bound defined by the average time between jumps $\Delta t_s < \Delta t_{\text{jumps}}$ and a left bound defined by the total observational time $T_{\text{obs}} = (M-1)\Delta t_s$ which therefore sets a lower limit for the dataset size M.

Finally, we analyzed biological and chemical applications for the inference frameworks we have described.
In Chapter 3 we studied two models for gene expression developed in [4]: one defined through jump rates and subsequently approximated through an SDE with non-Gaussian noise, the other described by an SDE with Gaussian noise. We exploited the jump model to apply the framework developed in Chapter 2 where the limitations of this approximate framework have been confirmed. We performed model distinguishability between the first model described through the non-Gaussian SDE and the second one defined by the Gaussian SDE. Discrimination was carried out through the comparison of the two exact likelihoods for the two models as an analytical closed-form propagator was available for both of them. Since the models share the same steady-state distribution, the results of model discrimination outlined how the two models remain distinguishable as long as data are temporally correlated, defining the optimal range for the sampling time as $\Delta t_s < \tau$, therefore proving that discrete fluctuations driven by Poisson noise are distinguishable from white noise fluctuations in a wide range of sampling times.
Finally, we considered a simple chemical network and we applied the approximate framework developed in Chapter 2 for jump processes to perform model distinguishability. We defined two models with the same macroscopic outcome and compared them using model discrimination performed through the approximate likelihood in Eq. 2.4. The results of this chapter outlined some limitations of this framework:

- The model distinguishability procedure showed a bias depending on which model was generating the synthetic data, penalizing the non-linear model in favor of the linear one. This has been attributed to the way the algorithm reconstructs the process path - approximating it to the shortest one - for large values of sampling time, i.e. when the actual jumps are not available.

- The informative window for distinguishability was very narrow, with probability P reaching values of ~ 0.8 only for $M \gtrsim 10000$.

Despite these limitations, it is interesting to note that a window of distinguishability can still be identified: we have robust distinguishability in the range bounded on the right by $\Delta t_s \sim \Delta t_{\text{jumps}}$ and on the left by the total observational time $\frac{T_{\text{obs}}}{\Delta t_{\text{jumps}}} \sim 10$, so that the distinguishability window becomes larger as the sample size M increases.

In conclusion, while the approximate framework introduced in Chapter 2 successfully recovers model distinguishability within specific computational regimes, its observed constraints highlight clear paths for future research.

- First, a natural next step involves testing alternative approximation methods to establish a rigorous comparison with our current approach. This would help decouple actual physical indistinguishability from potential artifacts introduced by the chosen approximate framework.
  For example, a possible improvement of the algorithm concerns the way the hidden path

between two consecutive sampled points is reconstructed. In this thesis, the approximate likelihood is based on the shortest-path reconstruction: as shown in Figure 4.1, among the possible paths connecting two observed states, we only retain the minimum one. A natural improvement could be to replace the shortest-path approximation with a most-likely-path approximation. Instead of selecting the path with the smallest number of jumps, one would select the path with the largest probability among those compatible with the two observed states, taking into account both the reaction rates and the time interval available between observations. This would provide a more physically informed approximation, while still avoiding the computation of the full process propagator.

- A second direction concerns the scalability of the framework. A first simple but interesting step beyond the models considered here would be the study of mediated reaction mechanisms. For instance, one could compare the direct model M1, $A \rightleftharpoons B$, with a model in which the transformation proceeds through an intermediate state C, namely $A \rightleftharpoons C \rightleftharpoons B$. Although still relatively simple, this example would already test whether the approximation remains reliable when additional hidden states are introduced. The intermediate state could represent, for example, an enzyme-substrate complex or the partially unfolded intermediate state of a protein in the transition from a folded state A to an unfolded state B.
  More generally, extending the framework to multivariate systems with several coupled chemical reactions would be essential to assess its robustness in higher-dimensional state spaces.

The lack of closed-form likelihood functions is the main limiting factor for effective inference in stochastic processes. Yet our work suggests that simple analytical approximations to the likelihood can be evaluated at minimal computational cost. While these methods have obvious limitations, which have been discussed above, we envisage that they could serve as tools for generating educated initial estimates that can later be refined using more computationally expensive methods.

Appendix

A Demographic noise

In this appendix, we show how demographic noise manifests as a square-root diffusion function (i.e. $\sim \sqrt{X}$) through the continuous approximation of a discrete birth-death Markov process.

Let $N(t) \in \mathbb{N}$ denote the population abundance at time t. Let us define the transition rates for an individual-based birth-death process in the small time interval Δt as

$$P(N \to N+1) = T^+(N)\,\Delta t + o(\Delta t) \qquad\qquad P(N \to N-1) = T^-(N)\,\Delta t + o(\Delta t),$$

where $T^+(N) = b\,N$ and $T^-(N) = d\,N$, with b, d per-capita birth and death rates. Then, the Master Equation for this process is

$$\frac{\partial P(N,t)}{\partial t} = T^+(N-1)\,P(N-1,t) + T^-(N+1)\,P(N+1,t) - [T^+(N) + T^-(N)]P(N,t).$$

To transition to a continuous variable $X \in \mathbb{R}^+$, we Taylor expand the shifted terms around X:

$$T^{\pm}(X \mp 1)P(X \mp 1,t) \sim T^{\pm}(X)\,P(X,t) \mp \frac{\partial}{\partial X}[T^{\pm}(X)\,P(X,t)] + \frac{1}{2}\frac{\partial^2}{\partial X^2}[T^{\pm}(X)\,P(X,t)].$$

Substituting this expression into the Master Equation yields the Fokker-Plank equation:

$$\frac{\partial P(X,t)}{\partial t} = -\frac{\partial}{\partial X}[\alpha(X)P(X,t)] + \frac{1}{2}\frac{\partial^2}{\partial X^2}[\beta(X)P(X,t)],$$

where $\alpha(X) = T^+(X) - T^-(X)$ and $\beta(X) = T^+(X) + T^-(X)$. The SDE associated with this equation is given by

$$dX_t = \alpha(X)dt + \sqrt{\beta(X)}dW_t = (b-d)\,X_t\,dt + \sqrt{(b+d)\,X_t}\,dW_t,$$

whose diffusion function exhibits the square-root dependence on the state X_t.

B Environmental noise

The deterministic logistic growth is given by (using standard notation)

$$\frac{dX_t}{dt} = rX_t(1 - \frac{X_t}{c}),$$

where r is the growth rate of the population and c is the carrying capacity of the environment, that is, its ability to support the population. Using the notation of [12],

$$\frac{dX_t}{dt} = kX_t(\mu - X_t), \tag{2}$$

the carrying capacity is represented by the parameter μ. Assuming that environmental fluctuations affect carrying capacity as

$$\mu \to \mu + a\,\xi(t),$$

where $\xi(t)$ is Gaussian noise, substituting in 2 and renaming $a\,k \to D$, one obtains

$$\frac{dX_t}{dt} = k\,X_t(\mu - X_t) + DX_t\,\xi(t),$$

which is a Stratonovich SDE. Performing the Stratonovich-Itô conversion [47]

$$A_{\text{Itô}}(x) = A_{\text{Stratonovich}}(x) + \frac{1}{2}B(x)\,B'(x),$$

one gets

$$dX_t = k\,X_t(\mu + \frac{D^2}{2k} - X_t)\,dt + DX_t\,dW_t.$$

Renaming $\mu \to \mu + \frac{D^2}{2k}$, one finally obtains Eq. 1.8.

C Autocorrelation time

The approximate autocorrelation time for model EN is obtained by linearizing around the stable fixed point of the deterministic dynamics, i.e. $X = \mu$. Define a small fluctuation $Y_t = X_t - \mu$ around μ and linearize the equation as

$$-kX_tY_t\,dt\; + X_tD\,dW_t \sim -k\mu Y_t\,dt + \mu D\,dW_t,$$

so that it assumes the form of an OU with $k \to k\mu$.

Since an exact analytical computation for the autocorrelation time of the EN model is not possible, the values used for normalization τ are obtained numerically. From the Wiener-Khinchin theorem [18], the continuous autocorrelation function is given by

$$\mathbb{E}[X_{t+s}X_t] = \mathcal{F}^{-1}\{|\mathcal{F}\{x(t)\}|^2\}.$$

To obtain it computationally, we use the simulated data, before the subsampling procedure, and apply the discrete version of the theorem using the fast Fourier transform FFT and its inverse IFFT. Normalizing we have a time-series for $C(s)$.
The autocorrelation time is then derived by interpolating the point just above and just below the value $\frac{1}{e}$, with indices k_1 and k_2. Thus, we can compute $\tau = (\frac{\frac{1}{e} - C[k_1]}{C[k_2] - C[k_1]} + k_1)\,dt$, where dt is the simulation time step.

D Path likelihood for a jump process

In this section we outline the construction of the path likelihood $L_{[t_0,t_N]}$ for a jump process.

At time t_i the probability for the dynamics $n(t_i) = n_i$ to change state during the infinitesimal time interval dt is given by $W(n_i)\,dt$, by definition of transition rate. Therefore, the probability that this transition does not occur is given by $1 - W(n_i)\,dt$. Considering a concatenation of such infinitesimal probabilities, one can compute the probability that the transition does not occur in a time interval $\Delta t_i = t_{i+1} - t_i$, which reads [48]

$$Q_i(\Delta t_i) = \lim_{dt\to 0}(1 - W(n_i)dt)^{\frac{\Delta t_i}{dt}} = e^{-W(n_i)\Delta t_i}.$$

To construct the likelihood for a particular path, one should multiply the probability to spend a time Δt_i in state n_i, i.e. the probability $Q_i(\Delta t_i)$ of not transitioning during time Δt_i, by the probability for the dynamics to perform exactly the transition $n_i \to n_{i+1}$ at t_{i+1}, i.e. the rate $w(n_i \to n_{i+1})$. Doing this for each state $\{n_i\}_{i=0,\dots,N-1}$ of the trajectory, we obtain the path likelihood in 2.3.

E Waiting-time probability density between jumps

For the Poissonian system in Eq. 2.5, the probability that a single jump occurs in $[t,\, t+dt]$ is given by $P_1(dt) = W(n)\,dt + o(dt)$, where $W(n)$ is the total exit rate from state n, defined in Eq. 2.8. The probability that no jumps occurred in $[0,\, t+dt]$ is given by

$$P_0(t+dt) = P_0(t)P_0(dt) = P_0(t)(1 - W(n)\,dt + o(dt)).$$

Subtracting $P_0(t)$ on both sides, dividing by dt and taking the $dt \to 0$ limit we get

$$\frac{dP_0(t)}{dt} = -W(n)P_0(t) \implies P_0(t) = e^{-W(n)\,t}.$$

Therefore, the waiting-time $f(t)$, which is the probability that no jumps occurred up to t and a single jump occurred in the interval $[t,\, t+dt]$, is given by

$$f(t)dt = P_0(t)P_1(dt) \implies f(t) = W(n)e^{-W(n)\,t}.$$

F The path likelihood approximation becomes exact as $\Delta t_s \to 0$

In this section we show that the approximate path likelihood $\tilde{L}_{[t_0,t_N]}$ in Eq. 2.4 becomes exact as $\Delta t_s \to 0$.

The likelihood approximation is partially due to the temporal approximation $\Delta t_i \simeq n_{s,i}\Delta t_s$. Suppose the system transitions into state n_i at the true time t_i and leaves it at t_{i+1}. The

discretization introduces two distinct sources of error:

- The jump into state n_i occurs at a random time between two consecutive sampling points. In the approximate path, this transition is detected only at the subsequent sampling epoch, delaying the recorded entry time.
- The second jump out of state n_i occurs within another sampling interval. Similarly, this transition is not recorded until the next sampling epoch, shifting the recorded exit time.

Consequently, in the worst-case scenario (where the entry jump occurs immediately after a sampling point and the exit jump occurs immediately before one) the total error introduced in estimating the waiting time Δt_i is bounded by $2\Delta t_s$. Therefore,

$$|\Delta t_i - n_{s,i}\Delta t_s| < 2\Delta t_s.$$

Then, as $\Delta t_s \to 0$, the discrepancy between the approximate interval and the true waiting time vanishes, ensuring that the discretized path exactly recovers the continuous-time dynamics.

G The number of positive jumps J_+ follows a Poisson distribution

Define $J_+(t)$ as the stochastic variable representing the number of positive jumps occurred up to time t, starting with $J_+(0) = 0$. Given a constant transition rate k_+, the probability $p(J_+, t)$[1] of having observed exactly J_+ jumps at time t is governed by the Master Equation

$$\frac{dp(J_+, t)}{dt} = k_+ p(J_+ - 1, t) - k_+ p(J_+, t). \tag{3}$$

To solve this system, we exploit the probability generating function, defined as

$$g(z, t) = \sum_{J_+=0}^{\infty} z^{J_+} p(J_+, t). \tag{4}$$

Let us multiply the ME 3 by z^{J_+} and sum over J_+. We obtain

$$\sum_{J_+=0}^{\infty} z^{J_+} \frac{\partial p(J_+, t)}{\partial t} = z \sum_{J_+=0}^{\infty} z^{J_+-1} k_+ \, p(J_+ - 1, t) \; - \; \sum_{J_+=0}^{\infty} z^{J_+} k_+ p(J_+, t),$$

where we divided and multiplied by z the first term. Redefining $J_+ - 1 \to J_+$, we obtain an equation for $g(z, t)$

$$\frac{\partial g(z, t)}{\partial t} = k_+ g(z, t)(z - 1),$$

whose solution is given by

$$g(z, t) = e^{k_+ t(z-1)}.$$

[1] with $p(J_+, t) = \delta_{J_+,0}$

The probability distribution $p(J_+, t)$ can be recovered by Taylor expanding $g(z, t)$ around $z = 0$:

$$g(z,t) = e^{-k_+ t} e^{k_+ tz} = e^{-k_+ t} \sum_{J_+=0}^{\infty} \frac{(k_+ t)^{J_+}}{J_+!} z^{J_+}. \tag{5}$$

By matching the coefficients of z^{J_+}, we identify the standard Poisson distribution:

$$p(J_+, t) = \frac{(k_+ t)^{J_+}}{J_+!} e^{-k_+ t}, \tag{6}$$

whose expectation value and variance are given by $\langle J_+ \rangle = \mathrm{Var}(J_+) = k_+ t$.

H Continuous limit of the geometric distribution $g(I)$

In this section, we show how the geometric distribution $g(I)$ describing the protein burst sizes converges to a continuous exponential distribution in the limit of $b >> 1$.

Starting from the geometric distribution:

$$g(I) = \frac{1}{b} \left(1 + \frac{1}{b}\right)^{-I}$$

we define a continuous variable $u = \frac{I}{b}$, which represents the burst-size scaled by the mean[2], described by a distribution $p(u)$. The discrete probability $g(I)$ maps onto the continuous probability density function through the relation $p(u)\Delta u = g(bu)$ [18], where $\Delta u = \frac{\Delta I}{b} = \frac{1}{b}$, so that:

$$p(u) = b\, g(bu) = \left(1 + \frac{1}{b}\right)^{-bu}.$$

Taking the limit $b \to \infty$, we obtain:

$$\lim_{b\to\infty} p(u) = \lim_{b\to\infty} \left[\left(1 + \frac{1}{b}\right)^{b}\right]^{-u} = e^{-u},$$

where we used the fundamental limit defining the exponential function $\lim_{b\to\infty}(1 + 1/b)^b = e$. Finally, expressing the distribution in terms of the original burst size I, in the continuous limit, we get:

$$p(u)du = e^{-I/b} \frac{dI}{b} = f(I)dI,$$

where we defined $f(I) = \frac{1}{b} e^{-I/b}$.

I Distribution of the non Gaussian noise

In this section I explore further the non Gaussian noise model for gene expression with SDE

$$dx(t) = (\delta - \gamma_2 x(t))dt + dK(t),$$

[2]For $b >> 1$, $b + 1 \simeq b$.

where $K(t) = \sum_{i=1}^{n(t)} I_i$. So $K(t)$ is the total amount of proteins produced from 0 to $t > 0$, where $\{I_i\}_{i=1,\dots n(t)}$ are independent, identically distributed random variables, with distribution $f(z) = \beta e^{-\beta z}$ with $\beta = \frac{\gamma_1}{k_2}$ s.t.

$$f(z)dz = \mathbb{P}(I_i \in [z,\, z + dz]).$$

The probability of $n(t)$ bursts in $[0, t]$ is given by $q_n(t) = \frac{(k_1 t)^n}{n!} e^{-k_1 t}$.

To better understand how is the noise like, let us explicitly compute the probability distribution of $K(t)$. As a starting point, one can write $p_K(x, t) = \langle \delta(x - K(t)) \rangle$, where $\langle \cdot \rangle$ is the average over all $K(t)$ realizations.
The conditional probability distribution $p_K(x, t|n(t) = 1)$ is the one for $K(t) = I_1$, so

$$p_K(x, t|1) = f(x).$$

Since the distribution of the sum of independent random variables is the convolution of their distribution, one gets

$$p_K(x, t|2) = (f * f)(x) = f(x)^{*2}$$

$$p_K(x, t|n) = f(x)^{*n}$$

So finally one can compute $p(x, t)$ as

$$p_K(x, t) = \sum_{n=0}^{\infty} p_K(x, t|n) q_n(t) = \delta(x) e^{-k_1 t} + \sum_{n=1}^{\infty} f(x)^{*n} \frac{(k_1 t)^n}{n!} e^{-k_1 t}.$$

The convolution can be computed explicitly:

$$f^{*2} = \int_0^x f(y)\ f(x - y)\ dy = \beta^2 e^{-\beta x} \int_0^x dy = \beta^2 e^{-\beta x} x$$

$$f^{*3} = \int_0^x f^{*2}(y)\ f(x - y)\ dy = \beta^3 e^{-\beta x} \int_0^x y\ dy = \beta^3 e^{-\beta x} \frac{x^2}{2}$$

$$f^{*n} = \beta^n \frac{x^{n-1}}{(n-1)!} e^{-\beta x}.$$

Thus the probability distribution is given by

$$\begin{aligned} p_K(x, t) &= \delta(x) e^{-k_1 t} + \frac{e^{-k_1 t - \beta x}}{x} \sum_{n=1}^{\infty} \frac{(k_1 t \beta x)^n}{(n-1)! n!} = \\ &= \delta(x) e^{-k_1 t} + t \frac{e^{-k_1 t - \beta x}}{x} \frac{\partial}{\partial t} I_0(2\sqrt{k_1 t \beta x}) = \\ &= \delta(x) e^{-k_1 t} + \frac{e^{-k_1 t - \beta x}}{x} \sqrt{k_1 t \beta x}\, I_1(2\sqrt{k_1 t \beta x}). \end{aligned}$$

Given that, one can compute the moments for the noise

$$\langle K^n(t)\rangle = \int_0^\infty x^n \; p_K(x,t) \; dx.$$

J Cumulants for the non Gaussian noise

To compute the cumulants for the noise $\Lambda(t)$, the starting point is to compute the characteristic function for $K(t)$

$$\langle e^{izK(t)}\rangle = \sum_{n=0}^{\infty}[g(z)]^n \; q_n(t) = \sum_{n=0}^{\infty}[g(z)]^n \frac{(k_1 t)^n}{n!}e^{-k_1 t} = e^{k_1 t(g(z)-1)}$$

where $g(z) = \langle e^{izI_j}\rangle = \int_0^\infty e^{izx} \; f(x) \; dx = \beta \int_0^\infty e^{izx}e^{-\beta x}dx = \frac{\beta}{\beta - iz}$.
To compute the cumulants, define the logarithm of the characteristic function for $K(t)$

$$\begin{aligned}\log\langle e^{izK(t)}\rangle &= k_1 t(\frac{\beta}{\beta - iz} - 1) = k_1 t(\frac{1}{1-\frac{iz}{\beta}} - 1) = \\ &= k_1 t(\sum_{n=0}^{\infty}(\frac{iz}{\beta})^n - 1) = k_1 t \sum_{n=1}^{\infty}(\frac{iz}{\beta})^n \overset{!}{=} \sum_{n=1}^{\infty} c_n \frac{(iz)^n}{n!}.\end{aligned}$$

Thus

$$c_n = k_1 t \frac{n!}{\beta^n}$$

which are the cumulants for $K(t) = \int_0^t \Lambda(s)ds$. To find the ones for $\Lambda(t)$, one can use the relation

$$c_n = \int_0^t ... \int_0^t \langle\langle \Lambda(t_1)...\Lambda(t_n)\rangle\rangle dt_1...dt_n,$$

and obtain

$$\langle\langle \Lambda(t_1)...\Lambda(t_n)\rangle\rangle = k_1 \frac{n!}{\beta^n}\prod_{i=2}^{n}\delta(t_i - t_1)$$

with $\beta = \frac{\gamma_1}{k_2}$. So the mean value for the noise is $\langle\Lambda(t)\rangle = \frac{k_1 k_2}{\gamma_1}$.

K Master Equation for the burst-like process

The burst-like process can be described through its Master Equation.

For the sake of self-consistency let us derive the general form of the Master Equation for a jump process. A Markov process is fully determined by two functions: the probability density $P_1(x,t) = p(x,t)$ and the conditional probability density or propagator $P_{1|1}(x_1,t_1|x_0,t_0)$, where $x_1 = x(t_1)$. These two functions obey

- the Chapman-Kolmogorov equation:

$$P_{1|1}(y_3,t_3|y_1,t_1) = \int P_{1|1}(y_3,t_3|y_2,t_2)P_{1|1}(y_2,t_2|y_1,t_1)dy_2,$$

- the relation:

$$P_1(y_2,t_2) = \int P_{1|1}(y_2,t_2|y_1,t_1)P_1(y_1,t_1)dy_1$$

For a time-homogeneous Markov process, which is the case of the burst-like gene expression model, the conditional probability density depends on the time interval only

$$P_{1|1}(y_2,t_2|y_1,t_1) = P_\tau(y_2|y_1)$$

where $\tau = t_2 - t_1$. Thus the Chapmann-Kolmogorov become:

$$P_{\tau+\tau'}(y_3|y_1) = \int P_{\tau'}(y_3|y_2)P_\tau(y_2|y_1)dy_2.$$

For small values of τ, one can omit second and higher orders in τ. Then, the expression for $P_\tau(y_2|y_1)$ is

$$P_\tau(y_2|y_1) = \delta(y_2-y_1)[1-\tau\int w(y_2,y_1)\ dy_2] + \tau w(y_2,y_1) + o(\tau)$$

where $w(y_2,y_1)$ is the transition probability per unit time from y_1 to y_2. The first term is the probability of not having a transition in τ, while the second term is the probability that the system will transition from y_1 to y_2 in τ. Substituting this into the Chapmann-Kolmogorov equation, one finds

$$P_{\tau+\tau'}(y_3|y_1) = [1-\tau'\int w(y_2,y_3)\ dy_2]\,P_\tau(y_3|y_1) + \tau'\int w(y_3,y_2)\,P_\tau(y_2|y_1)\,dy_2 + o(\tau').$$

Finally, dividing by τ' and taking $\tau' \to 0$ one get

$$\frac{\partial}{\partial\tau}P_\tau(y_3|y_1) = \int\{w(y_3,y_2)P_\tau(y_2|y_1) - w(y_2,y_3)P_\tau(y_3|y_1)\}dy_2.$$

To simplify the notation, consider $y_1 = 0$ and $t_1 = 0$, so $P_\tau(y_2|y_1) = p(y,t)$ and $P_\tau(y_3|y_1) = p(x,t)$:

$$\frac{\partial}{\partial t}p(x,t) = \int\{w(x,y)p(y,t) - w(y,x)p(x,t)\}dy.$$

In the context of the burst-like gene expression model, the transition rate describes the burst event and is given by

$$w(x,y) = k_1 f(x-y)\Theta(x-y),$$

then the Master Equation part describing the burst events is

$$\frac{\partial}{\partial t}p(y,t) = k_1\int_0^x f(x-y)p(y,t)\,dy - k_1 p(x;t)\int_x^\infty f(y-x)\,dy = k_1\int_0^x f(x-y)p(y,t)\,dy - k_1 p(x;t).$$

Notice that the model we are discussing is composed of jumping dynamics, describing the bursts, and deterministic dynamics, describing the protein degradation. If there are no jumps

the probability density simply satisfy the continuity equation:

$$\frac{\partial}{\partial t}p(x,t) = -\frac{\partial}{\partial x}[(\delta - \gamma_2 x)p(x,t)].$$

Putting everything together one obtains the Master Equation of the full model

$$\frac{\partial}{\partial t}p(x,t) = -\frac{\partial}{\partial x}[(\delta - \gamma_2 x)p(x,t)] + k_1 \int_0^x f(x-y)\, p(y,t)\, dy - k_1\, p(x,t). \tag{7}$$

From Equation 7, one computes the probability flux:

$$j(x,t) = (\delta - \gamma_2 x)p(x,t) - k_1 \int_0^x dz \int_0^z f(z-y)\, p(y,t)\, dy + \int_0^x k_1\, p(y;t) dy.$$

L Chemical reactions: M1 vs M2 distinguishability

For completeness, Figure 4 shows what happens to model discrimination between M1 and M2 for slightly longer values of the sampling time interval. In particular, observe that the bias for P_1 and P_2 is amplified. P_2 decreases below 0.5, confirming that M1 is preferred even if it is not the true model.

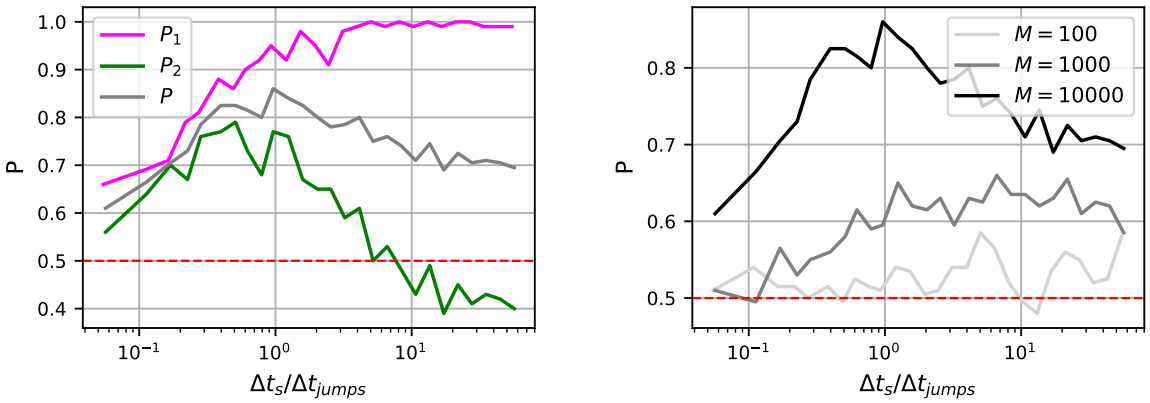

Figure 4: In this Figure are represented the same plots of 4.2, with slightly higher values of Δt_s.

Bibliography

[1] A. R. Ravishankara, David A. Randall, and James W. Hurrell. Complex and yet predictable: The message of the 2021 nobel prize in physics. *Proceedings of the National Academy of Sciences*, 119(4):e2122340119, 2022.

[2] Valerie Isham. Stochastic models for epidemics. *Interdisciplinary Science Reviews*, 29(3):284–298, 2004.

[3] Rama Cont and Adrien de Larrard. Price dynamics in a markovian limit order market. *SIAM Journal on Financial Mathematics*, 4(1):1–25, 2013.

[4] Sandro Azaele, Jayanth R. Banavar, and Amos Maritan. Probing noise in gene expression and protein production. *Physical Review E*, 80(3):031916, 2009.

[5] Lana Descheemaeker and Sophie de Buyl. Stochastic logistic models reproduce experimental time series of microbial communities. *eLife*, 9:e55650, 2020.

[6] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning*. Springer, New York, 2009.

[7] Sheldon M. Ross. *Introduction to probability and statistics for engineers and scientists*. Elsevier Academic Press, 2014.

[8] Bernt Øksendal. *Stochastic Differential Equations*. Springer, Berlin, Heidelberg, 2003.

[9] Simona Cocco, Rémi Monasson, and Francesco Zamponi. *From Statistical Physics to Data-Driven Modelling*. Oxford University Press, 2022.

[10] Peter Craigmile, Radu Herbei, Ge Liu, and Grant Schneider. Statistical inference for stochastic differential equations. *WIREs Computational Statistics*, 14(4):e1564, 2022.

[11] Larry Wasserman. *All of Statistics: A Concise Course in Statistical Inference*. Springer, 2004.

[12] Javier Aguilar, Miguel A. Muñoz, and Sandro Azaele. The limits of inference in complex systems: When stochastic models become indistinguishable. *Physical Review X*, 16, 2026.

[13] Oscar Fajardo-Fontiveros, Jordi Duch, Ignasi Reichardt, Marta Sales-Pardo, Harry R. De Los Ríos, and Roger Guimerà. Fundamental limits to learning closed-form mathematical models from data. *Nature Communications*, 14(1):1043, 2023.

[14] Javier Aguilar, Samir Suweis, Amos Maritan, and Sandro Azaele. Unraveling the temporal dependence of ecological interaction measures. *PRX Life*, 4, 2026.

[15] Jason M. T. Roos. Consumers of experimental observations: Understanding how experimental costs affect sample size and composition. *Management Science*, 64(7):3420–3438, 2018.

[16] Carlos A. Braumann. Environmental versus demographic stochasticity in population growth. *Demographic Research*, 23:631–652, 2010.

[17] Linda J. S. Allen. An introduction to stochastic epidemic models. *Mathematical Epidemiology*, pages 81–130, 2008.

[18] Crispin W. Gardiner. *Handbook of Stochastic Methods*. Springer, 1985.

[19] J. V. Ross, T. Taimre, and P. K. Pollett. On parameter estimation in population models. *Theoretical Population Biology*, 70(4):498–510, 2006.

[20] Tao Feng, Hongjuan Zhou, Zhipeng Qiu, and Yun Kang. Impacts of demographic and environmental stochasticity on population dynamics with cooperative effects. *Journal of Mathematical Analysis and Applications*, 507(1):125746, 2022.

[21] Steinar Engen, Øyvind Bakke, and Aminul Islam. Demographic and environmental stochasticity: Concepts and definitions. *Biometrics*, 54(3):840–846, 1998.

[22] Ivan Dornic, Hugues Chaté, and Miguel A. Muñoz. Integration of langevin equations with multiplicative noise and the viability of field theories for absorbing phase transitions. *Physical Review Letters*, 94(10):100601, 2005.

[23] Linda J. S. Allen. *An Introduction to Stochastic Processes with Applications to Biology*. Pearson, 2010.

[24] Desmond J. Higham. An algorithmic introduction to numerical simulation of stochastic differential equations. *SIAM Review*, 43(3):525–546, 2001.

[25] P. E. Kloeden and T. Shardlow. The milstein scheme for stochastic delay differential equations without using anticipative calculus. *Journal of Difference Equations and Applications*, 20(5-6):832–846, 2014.

[26] John C. Cox, Jonathan E. Ingersoll, and Stephen A. Ross. A theory of the term structure of interest rates. *Econometrica*, 53(2):385–407, 1985.

[27] N. G. van Kampen. *Stochastic processes in physics and chemistry*. Elsevier, 2007.

[28] In Jae Myung. Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology*, 47(1):90–100, 2003.

[29] Neda Mohammadi, Leonardo V. Santoro, and Victor M. Panaretos. Nonparametric estimation for sde with sparsely sampled paths: An fda perspective. *Journal of Multivariate Analysis*, 201:105285, 2024.

[30] Stefano M. Iacus. *Simulation and Inference for Stochastic Differential Equations*. Springer, 2008.

[31] Raul Toral and Pere Colet. *Stochastic Numerical Methods*. Wiley-VCH, 2014.

[32] Yudi Pawitan. *In All Likelihood*. Clarendon Press, 2001.

[33] Cheng Yong Tang and Song Xi Chen. Parameter estimation and bias correction for diffusion processes. *Journal of Econometrics*, 149(1):65–81, 2009.

[34] Kenneth P. Burnham and David R. Anderson. *Model Selection and Multimodel Inference*. Springer, 2002.

[35] Yury A. Kutoyants. *Statistical Inference for Ergodic Diffusion Processes*. Springer, 2004.

[36] Halbert White. Maximum likelihood estimation of misspecified models. *Econometrica*, 50(1):1–25, 1982.

[37] Vincenzo Capasso and David Bakstein. *An Introduction to Continuous-Time Stochastic Processes*. Birkhäuser, 2021.

[38] Darren James Wilkinson. *Stochastic Modelling for Systems Biology*. Chapman & Hall/CRC, 2006.

[39] Arjun Raj and Alexander van Oudenaarden. Nature, nurture, or chance: Stochastic gene expression and its consequences. *Cell*, 135(2):216–226, 2008.

[40] Daniel T. Gillespie. Stochastic simulation of chemical kinetics. *Annual Review of Physical Chemistry*, 58:35–55, 2007.

[41] Roy D. Dar, Nina N. Hosmane, Michelle R. Arkin, Robert F. Siliciano, and Leor S. Weinberger. Screening for noise in gene expression identifies drug synergies. *Science*, 344(6182):419–423, 2014.

[42] Peter Atkins and Julio de Paula. *Physical Chemistry*. Oxford University Press, 2014.

[43] Bree B. Aldridge, John M. Burke, Douglas A. Lauffenburger, and Peter K. Sorger. Physicochemical modelling of cell signalling pathways. *Nature Cell Biology*, 8(11):1195–1203, 2006.

[44] James E. Ferrell Jr. Self-perpetuating states in signal transduction: positive feedback, double-negative feedback and bistability. *Current Opinion in Cell Biology*, 14(2):140–148, 2002.

[45] Irving R. Epstein and John A. Pojman. *An Introduction to Nonlinear Chemical Dynamics*. Oxford University Press, 1998.

[46] Christopher V. Rao, Denise M. Wolf, and Adam P. Arkin. Control, exploitation and tolerance of intracellular noise. *Nature*, 420(6912):231–237, 2002.

[47] N. G. van Kampen. Itô versus Stratanovich. *Journal of Statistical Physics*, 24(1):175–187, 1981.

[48] Javier Aguilar, José J. Ramasco, and Raúl Toral. Biased versus unbiased numerical methods for stochastic simulations. *Communications Physics*, 7(1):23, 2024.