Greek Statistical Institute
Proceedings of the 33rd Panhellenic Statistics Conference (2021), pp. 26−38

STATISTICAL ANALYSIS AND APPLICATIONS OF THE MULTI-SIGMOIDAL DETERMINISTIC AND STOCHASTIC LOGISTIC GROWTH

***Antonio Di Crescenzo*[1], *Paola Paraggio*[1], *Patricia Román-Román*[2], *Francisco Torres-Ruiz*[2]**
[1]University of Salerno
{adicrescenzo, pparaggio}@unisa.it
[2]University of Granada
{proman, fdeasis}@ugr.es

ABSTRACT

We consider a multi-sigmoidal generalization of the logistic growth model. The deterministic model is provided together with its stochastic counterpart. More in detail, we analyse two different birth-death processes with linear and quadratic rates, respectively. From the latter we derive a more manageable diffusive approximation by means of a suitable scaling. Furthermore, we study two possible strategies to obtain the maximum likelihood estimates of the parameters. To validate the described procedures, we conclude with a simulation study. The first-passage-time problem is also addressed.

Keywords: Logistic model, multi-sigmoidal growth models, birth-death process, diffusion process, maximum likelihood estimation, first-passage-time problem.

1. INTRODUCTION

The exponential curve is the most common basic model to describe growth of populations in ideal conditions. However, such kind of growth does not occur in nature apart from short time periods. For most living species, indeed, there exists a critical density beyond which the relative population does not find sufficient environmental factors to grow and reproduce. Mathematical models which take into account environmental factors that limit the growth rate of population are characterized by a S-shape and for this reason are called sigmoidal. The logistic model, more in detail, is a sigmoidal growth model with an initial slow growth followed by an explosion of exponential-type which finally flattens up to an equilibrium status, known as carrying capacity. The application of sigmoidal curves are various and they involve several contexts of interest, from biology to medicine, from ecology to software reliability (see, for instance, Erto *et al.* (2020)). For example, in the recent works of Rajasekar *et al.* (2020) the authors analyse a stochastic version of SIR model for the diffusion of the COVID-19 pandemic, by considering a logistic-kind growth for the susceptible individuals.

Anyway, it is possible that a population reaches its limit value after various successive steps. This is the reason why recent investigations address their interest to a generalization of the sigmoidal models by introducing multiple inflections. Such generalizations are the so-called multi-sigmoidal models (cf. Román-Román *et al.* (2019)). The multi-sigmoidal logistic model, in particular, is appropriate to describe maturation of some fruit species as peaches or coffee berries (see Figure 1) which show a growth trend with multiple fluctuations.

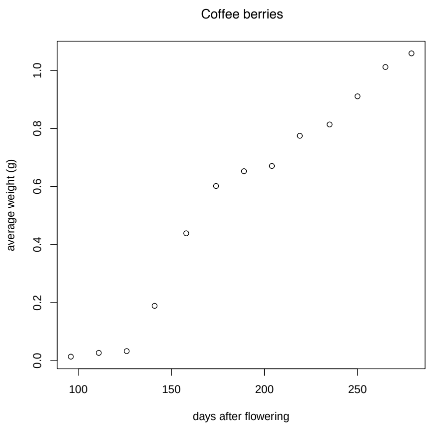

Figure 1: *The multi-sigmoidal growth of coffee berries. The data are taken from Cuhna and Volpe (2011).*

This work is a brief summary of a larger study concerning multi-sigmoidal logistic growth model. The deterministic model together with the birth-death processes and the diffusive approximation have been analysed widely in Di Crescenzo *et al.* (2021a). Whereas, the statistical analysis of the above-mentioned model is the subject of a paper submitted for publication (cf. Di Crescenzo *et al.* (2021b)).

2. THE DETERMINISTIC MODEL

The multi-sigmoidal logistic curve $l_m(t)$ satisfies a generalized version of the Cauchy problem related to the classical logistic model, i.e.

$$\frac{d}{dt} l_m(t) = h_\theta(t) l_m(t), \qquad t \geq t_0, \qquad l_m(t_0) = l_0, \tag{1}$$

where

$$h_\theta(t) = \frac{P_\beta(t) e^{-Q_\beta(t)}}{\eta + e^{-Q_\beta(t)}}, \tag{2}$$

with

$$Q_\beta(t) = \sum_{i=1}^{p} \beta_i t^i, \qquad \beta_p > 0, \qquad P_\beta(t) = \frac{d}{dt} Q_\beta(t), \tag{3}$$

for $\eta > 0$, $\beta_1, \ldots, \beta_{p-1} \in \mathbb{R}$, $\beta_p > 0$, $\theta = (\eta, \beta^T)^T$ and $\beta^T = (\beta_1, \ldots, \beta_p)$, $p \in \mathbb{N}$. Note that when $p = 1$, $Q_\beta(t)$ is linear and from Eq. (1) we come to the classical logistic equation. The solution of the initial value problem (1) is given by

$$l_m(t) = l_0 \frac{\eta + e^{-Q_\beta(t_0)}}{\eta + e^{-Q_\beta(t)}}, \qquad t \geq t_0. \tag{4}$$

In Eqs. (1) and (4) the subscript m means 'multi-sigmoidal'. Various choices of the parameters $\eta, \beta_1, \ldots, \beta_p$ lead to different kinds of shape characterized by multiple inflection points. See, for instance, Figure 2. It is possible to show that the carrying

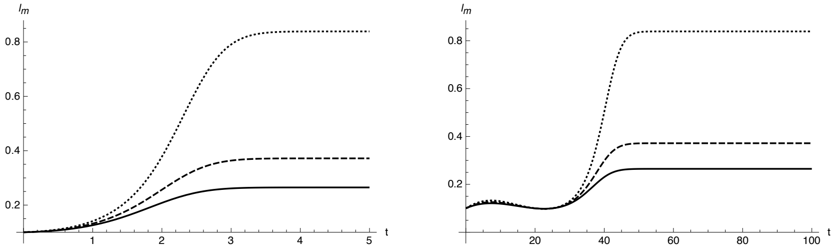

Figure 2: *The multi-sigmoidal logistic function for* $p = 3$, $t_0 = 0$, $l_0 = 0.1$, $\beta_1 = 0.1$ *(a)* $\beta_2 = 0.2$, $\beta_3 = 0.1$, *(b)* $\beta_2 = -0.009$, $\beta_3 = 0.0002$. *In both cases* $\eta = e^{-0.5}, e^{-1}, e^{-2}$ *(from bottom to top).*

capacity of the model depends on the relevant parameters θ and on the initial condition $l_m(t_0) = l_0$. More in detail, it is given by C/η with $C = C(l_0, \theta, t_0) = l_0(\eta + e^{-Q_\beta(t_0)})$. Indeed, from the assumption $\beta_p > 0$, one has the following limit

$$\lim_{t \to \infty} l_m(t) = l_0 \frac{\eta + e^{-Q_\beta(t_0)}}{\eta} = \frac{C}{\eta}.$$

A key-role in the analysis of the multi-sigmoidal logistic function is played by the inflection points which give to the curve the characteristic shape. By performing the second derivative of Eq. (4), it is possible to show that the inflection points are the solutions of the following equation (in the unknown $t \geq t_0$)

$$\frac{d^2}{dt^2} Q_\beta(t) = \left(\frac{d}{dt} Q_\beta(t) \right)^2 \frac{\eta - e^{-Q_\beta(t)}}{\eta + e^{-Q_\beta(t)}}, \tag{5}$$

with $Q_\beta(t)$ defined in Eq. (3). Due to the transcendental nature of the above-mentioned equation, one is forced to use numerical methods to solve it.

Example 1. *With reference to the real data given in Section 1, we now consider an application of the multi-sigmoidal model. To avoid numerical problems, we perform a time shifting so that* $t_0 = 0$ *(this does not affect the generality of the analysis, as pointed*

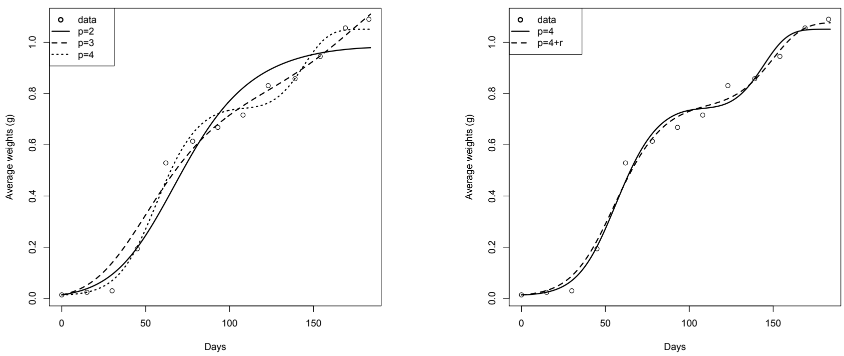

Figure 3: *Fitted multi-sigmoidal logistic curve for the coffee berries with (a) integer and (b) non-integer degrees, with $r = -4.498494 \cdot 10^{-1}$ in case (b).*

out in Remark 2.1 of Di Crescenzo et al. (2021a)). Specifically, we determine the values of the parameters θ minimizing the square error S_p defined as follows

$$S_p(\theta) = \sum_{i=1}^{n} (y_i - l_m(t_i))^2, \qquad \theta = (\eta, \beta^T)^T$$

where y_i are the data, t_i are the shifted times for $i = 1, \ldots, n$ and p is the degree of Q_β. As shown in Figure 3-(a), the best fit is attained when $p = 4$. Note that the minimization of the function S_p has been performed by means of the Nelder-Mead optimization method. Since it is an iterative method, the needed initial solution can be determined as described in Section 3 of Di Crescenzo et al. (2021a). To improve the goodness-of-fit of the proposed model, the last term of the polynomial Q_β can be modified to have a real exponent:

$$\tilde{l}_m(t) = l_0 \frac{\eta + e^{-\tilde{Q}_\beta(t_0)}}{\eta + e^{-\tilde{Q}_\beta(t)}}, \qquad t \geq t_0,$$

where $\tilde{Q}_\beta(t) = \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + \beta_4 t^{4+r}$ and $r \in \mathbb{R}$. Now the aim is to find the best set of parameters $\tilde{\theta} = (\theta^T, r)^T$, i.e. the set which minimizes the cumulative square error defined below

$$S_{4+r}(\tilde{\theta}) = \sum_{i=1}^{n} (y_i - \tilde{l}_m(t_i))^2,$$

where, as above, y_i are the data, t_i are the time instants for $i = 1, \ldots, n$. As shown in Figure 3-(b), the goodness-of-fit increases since $S_4(\theta) > S_{4+r}(\tilde{\theta})$. In particular, we have $S_4(\theta) = S_{4+r}(\tilde{\theta}) + 22,52\%$.

3. BIRTH-DEATH PROCESSES

Birth-death (BD) processes are often adopted to describe stochastic dynamics in various fields of biomathematics, being appropriate to model the random evolution of the number of particles or individuals in a system. Let us consider a time-inhomogeneous BD process $\{N(t); t \geq 0\}$ with state space $\mathbb{N}_0$ and linear birth and death rates given by

$$\begin{aligned} b_n(t) &= n\lambda(t), \qquad n \in \mathbb{N}_0 \\ d_n(t) &= n\mu(t), \qquad n \in \mathbb{N}, \qquad d_0(t) = 0, \end{aligned} \tag{6}$$

where the individual birth and death rates λ and μ are integrable and positive functions in any set $(0, t)$ with $t \geq 0$. In the following proposition, a sufficient and necessary condition to have a conditional mean of multi-sigmoidal logistic type is provided (as done elsewhere, for example in Di Crescenzo and Paraggio (2019) and Di Crescenzo and Spina (2016)).

Proposition 1. *The BD process $N(t)$ with rates given in Eq.* (6) *has conditional mean $\mathbb{E}[X(t)|X(0) = n_0]$ of multi-sigmoidal logistic type if, and only if, the net growth rate $\xi(t) = \lambda(t) - \mu(t)$ is given by $\xi(t) = h_\theta(t)$, for $t \geq 0$.*

See Figure 4 for some plots of the conditional mean considering different choices of the parameters.

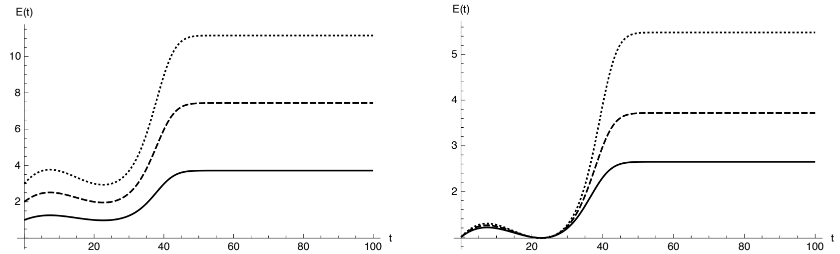

Figure 4: *The conditional mean $\mathbb{E}(t) = \mathbb{E}[X(t)|X(0) = n_0]$ for $p = 3$, $t_0 = 0$, $\beta_1 = 0.1$, $\beta_2 = -0.009$, $\beta_3 = 0.0002$ (a) $\eta = e^{-1}$ and $n_0 = 1, 2, 3$ (from bottom to top) (b) $n_0 = 1$, $\eta = e^{-0.5}, e^{-1}, e^{-1.5}$ (from bottom to top).*

The first-passage-time (FPT) problem is relevant in several applications in population dynamics since the first crossing of a critical high (low) threshold can be viewed as the rising of an overpopulation (extinction). For a fixed threshold $n \in \mathbb{N}$, the FPT of the process $N(t)$ through the state n starting from $N(0) = n_0$ is defined as follows

$$T_{n_0,n} = \inf\left\{t \geq 0 : N(t) = n\right\}, \qquad N(0) = n_0.$$

Let us denote by $g_{n_0,n}$ the corresponding probability density function, i.e.

$$g_{n_0,n}(t) = \frac{d}{dt}\mathbb{P}\left(T_{n_0,n} \leq t\right), \qquad t \geq 0.$$

Considering the same matrix-based approach adopted by Tan in Section 3 of Tan (1986), the FPT density vector $g_n := \left[g_{1,n}, \ldots, g_{n-1,n}\right]^T$ can be expressed as

$$g_n(t) = \lambda(t)\left(P_1 e^D P_1^{-1}\right)^{-\Lambda(t)}\left(P_2 e^D P_2^{-1}\right)^{-M(t)} P_1 D P_1^{-1} \mathbb{I}_{n-1,1},$$

where $A_i = P_i D P_i^{-1}$ for $i = 1, 2$, $A_1 = \left(a_{i,j}^{(1)}\right)$ and $A_2 = \left(a_{i,j}^{(2)}\right)$ defined in such a way

$$a_{i,j}^{(1)} = \begin{cases} -i, & j = i+1 \\ i, & j = i \\ 0, & \text{otherwise} \end{cases}$$

for $i = 1, \ldots, n-2$ and

$$a_{i,j}^{(2)} = \begin{cases} -i, & j = i-1 \\ i, & j = i \\ 0, & \text{otherwise} \end{cases}$$

for $i = 2, \ldots, n-1$. Moreover, $\Lambda(t) = \int_0^t \lambda(s)ds$, $M(t) = \int_0^t \mu(s)ds$ and $\mathbb{I}_{n-1,1}$ is a column of all 1 of dimension $n-1$. In Figure 5 we provide some plots of the FPT density.

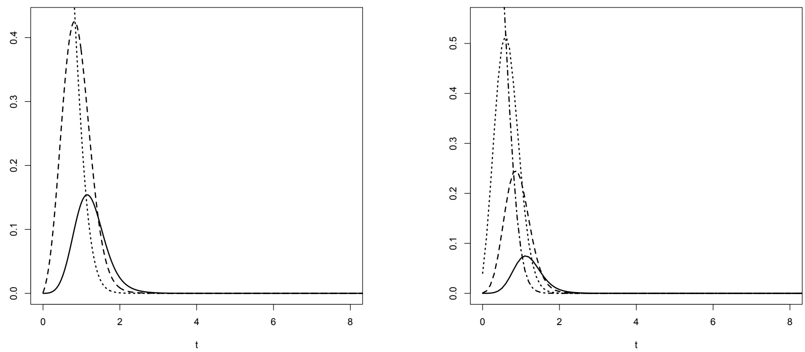

Figure 5: *The FPT density for $\lambda(t) = 2h_\theta(t)$, $\mu(t) = h_\theta(t)$, $Q_\beta(t) = 0.1t + 0.2t^2 + 0.1t^3$, $n_0 = 1$ (solid), 2 (dashed), 3 (dotted), 4 (dot-dashed) and (a) $n = 4$ and (b) $n = 5$.*

In various applications in biomathematics the systems under investigation are subject to dynamics regulated by transitions where rates are allowed to be nonlinear. Let $\left\{\bar{X}(t); t \geq 0\right\}$ be an inhomogeneous nonlinear BD process having $\mathbb{N}_0$ as state space and birth and death rates given by

$$\begin{aligned} \bar{b}_n(t) &= \lambda_1(t) + \lambda_2(t)n + \lambda_3(t)n^2, \qquad n \in \mathbb{N}_0, \\ \bar{d}_n(t) &= \mu_1(t) + \mu_2(t)n + \mu_3(t)n^2, \qquad n \in \mathbb{N}, \qquad d_0(t) = 0, \end{aligned} \tag{7}$$

where λ_1 and μ_1 are non-negative and integrable functions and λ_i and μ_i for $i = 2, 3$ are positive and integrable functions on any set $(0, t)$. It is easy to note that when $\lambda_1(t) = \mu_1(t) = 0$ and $\lambda_3(t) = \mu_3(t)$, then the mean $m_1(t) = \mathbb{E}[\bar{N}(t)|\bar{N}(0) = n_0]$ satisfies the differential equation

$$\frac{d}{dt} m_1(t) = (\lambda_2(t) - \mu_2(t)) m_1(t),$$

which is a differential equation of the same type of Eq. (1). Hence, assuming that $\lambda_2(t) - \mu_2(t) = h_\theta(t)$, the mean can be expressed as

$$m_1(t) = n_0 \frac{\eta + e^{-Q_\beta(t_0)}}{\eta + e^{-Q_\beta(t)}}, \qquad m_1(0) = n_0.$$

4. THE CORRESPONDING DIFFUSION PROCESS

In order to obtain a more manageable description of the growth phenomenon, we perform a diffusive approximation of the BD process with nonlinear rates given in Eq. (7). The diffusive approximation is based on the scaled BD process $N_\varepsilon(t) = \varepsilon \bar{N}(t)$ whose probability $p_n^\varepsilon(t)$, for $\varepsilon \simeq 0$, gives $p_n^\varepsilon(t) \simeq f(x,t)\varepsilon$ with $x = n\varepsilon$. Under some suitable assumptions, the limits below hold for $\varepsilon \to 0$

$$\begin{aligned} &(\mu_i(t) - \lambda_i(t))\varepsilon \to 0, \qquad i = 1,3, \qquad (\mu_2(t) - \lambda_2(t))\varepsilon \to -r(t), \\ &(\mu_i(t) + \lambda_i(t))\varepsilon^2 \to 0, \qquad i = 1,2, \qquad (\mu_3(t) + \lambda_3(t))\varepsilon^2 \to \sigma^2. \end{aligned}$$

Hence, performing the derivative of f with respect to t and expanding f as a Taylor series around x, the density function f of the approximating process satisfies the following Fokker-Plank equation

$$\frac{\partial}{\partial t} f(x,t) = -\frac{\partial}{\partial x}[r(t)xf(x,t)] + \frac{1}{2}\frac{\partial^2}{\partial x^2}[\sigma^2 x^2 f(x,t)].$$

In other terms, for $r(t) = h_\theta(t)$ with h_θ defined in Eq. (2), the BD process $\bar{N}(t)$ leads to the lognormal diffusion process $X(t)$ having infinitesimal moments

$$A_1(x,t) = h_\theta(t)x, \qquad A_2(x) = \sigma^2 x^2.$$

The initial condition $p_{n_0}(0) = 1$ becomes $\lim_{t\to 0} f(x,t) = \delta(x - x_0)$, where δ is the Dirac delta function. The resulting diffusion process $\{X(t); t \geq t_0\}$ has state space $(0, +\infty)$ and is governed by the following SDE:

$$dX(t) = h_\theta(t)X(t)dt + \sigma X(t)dW(t), \qquad X(t_0) = X_0, \tag{8}$$

where $W(t)$ is a Wiener process independent on the initial condition X_0, for any $t \geq t_0$, $\theta = (\eta, \beta^T)^T$ and $\sigma > 0$. By applying Itô's formula to Eq. (8), we obtain

$$X(t) = X_0 \exp\left[H_\xi(t_0,t) + \sigma\left(W(t) - W(t_0)\right)\right], \qquad t \geq t_0, \tag{9}$$

where $\xi = (\theta^T, \sigma^2)^T$ is the vector containing the parameters of the model and

$$H_\xi(s,t) = \log \frac{\eta + e^{-Q_\beta(s)}}{\eta + e^{-Q_\beta(t)}} - \frac{\sigma^2}{2}(t-s), \qquad t_0 \leq s < t. \tag{10}$$

It is worth to notice that if the initial state X_0 is lognormally distributed with parameters μ_0 and σ_0^2 or is degenerate, then the finite dimensional distributions of $X(t)$ are lognormal. Under the above-mentioned assumptions on X_0, the mean of the process is

$$m_1(t) = \mathbb{E}[X(t)] = \mathbb{E}[X_0]\frac{\eta + e^{-Q_\beta(t_0)}}{\eta + e^{-Q_\beta(t)}}, \qquad t \geq t_0,$$

the mode is

$$Mode[X(t)] = Mode[X_0]\frac{\eta + e^{-Q_\beta(t_0)}}{\eta + e^{-Q_\beta(t)}} \exp\left(-\frac{3}{2}\sigma^2(t - t_0)\right), \qquad t \geq t_0,$$

and, finally, the median is

$$med[X(t)] = med[X_0]\frac{\eta + e^{-Q_\beta(t_0)}}{\eta + e^{-Q_\beta(t)}} \exp\left(-\frac{\sigma^2}{2}(t - t_0)\right), \qquad t \geq t_0.$$

5. MAXIMUM LIKELIHOOD ESTIMATES

In this section we describe two different procedures to find the maximum likelihood estimates (MLEs) of the parameters. We consider a discrete sampling of $X(t)$ based on d independent sample paths, with n_i different observation instants for the i-th sample path, i.e. t_{ij} for $j = 1, \ldots, n_i$, $i = 1, \ldots, d$. For simplicity, we assume that the first time instant is the same for all the sample paths, i.e. $t_{i1} = t_0$, $i = 1, \ldots, d$. The vector $\mathbb{X} = (\mathbb{X}_1^T | \ldots | \mathbb{X}_d^T)^T$, where $\mathbb{X}_i = (X(t_{i1}), \ldots, X(t_{in_i}))^T$ for $i = 1, \ldots, d$ and $X(t_0)$ is lognormally distributed with parameters μ_1 and σ_1^2, has density

$$f_{\mathbb{X}}(x) = \prod_{i=1}^{d} \exp\left(-\frac{(\log x_{i,1} - \mu_1)^2}{x_{i,1}\sigma_1\sqrt{2\pi}}\right) \prod_{j=1}^{n_i-1} \frac{\exp\left(-\frac{\left[\log\left(\frac{x_{i,j+1}}{x_{i,j}}\right) - m_\xi^{i,j+1,j}\right]^2}{2\sigma^2\Delta_i^{j+1,j}}\right)}{x_{i,j}\sigma\sqrt{2\pi\Delta_i^{j+1,j}}}$$

where $\Delta_i^{m,n} = t_{i,m} - t_{i,n}$, $m, n = 1, \ldots, n_i - 1$, $m > n$, $\xi = (\theta^T, \sigma^2)^T$ and $m_\xi^{i,m,n} = H_\xi(t_{i,n}, t_{i,m})$ with H_ξ defined in Eq. (10).

If (μ_1, σ_1^2) and ξ are functionally independent, the MLEs of (μ_1, σ_1^2) leads to $\hat{\mu}_1 = \frac{1}{d}\log x_{i,1}$, and $\hat{\sigma}_1^2 = \frac{1}{d}\sum_{i=1}^{d}(\log x_{i,1} - \hat{\mu}_1)^2$. The estimation of ξ is obtained from the following system

$$\begin{cases} \sigma^2\left(n + \frac{\sigma^2}{4}Z_3\right) - Z_1 - A_\theta + 2B_\theta = 0 \\ Y_l^\theta + \frac{\sigma^2}{2}W_l^\theta + X_l^\theta = 0, \quad l = 0, 1, \ldots, p, \end{cases} \tag{11}$$

where for $v_{0i} = x_{i,1}$ and $v_{i,j} = \left(\Delta_i^{j+1,j}\right)^{-1/2} \log\left(\frac{x_{i,j+1}}{x_{i,j}}\right)$, $j = 1, \ldots, n_i - 1$, $i = 1, \ldots, d$, we have set

$$W_l^\theta = \sum_{i=1}^{d} {}_lD_\theta^{i,n_i,1}, \quad Y_l^\theta = \sum_{i=1}^{d}\sum_{j=1}^{n_i-1} \frac{1}{\Delta_i^{j+1,j}} \log\left[\frac{\eta + e^{-Q_\beta(t_{i,j+1})}}{\eta + e^{-Q_\beta(t_{i,j})}}\right] {}_lD_\theta^{i,j+1,j}$$

$$X_l^\theta = \sum_{i=1}^{d}\sum_{j=1}^{n_i-1} \frac{v_{i,j}}{\left(\Delta_i^{j+1,j}\right)^{1/2}} {}_lD_\theta^{i,j+1,j}, \qquad l = 0, 1, \ldots, p,$$

$$\lambda_\theta^{i,m,n} = \log \frac{\eta + e^{Q_\beta(t_{i,n})}}{\eta + e^{Q_\beta(t_{i,m})}}, \quad m > n,\ i = 1, \ldots, d, \quad Z_1 = \sum_{i=1}^{d}\sum_{j=1}^{n_i-1} v_{ij}^2$$

$$Z_3 = \sum_{i=1}^{d} \Delta_i^{n_i,1}, \quad A_\theta = \sum_{i=1}^{d}\sum_{j=1}^{n_i-1} \frac{\left(\lambda_\theta^{i,j+1,j}\right)^2}{\Delta_i^{j+1,j}}, \quad B_\theta = \sum_{i=1}^{d}\sum_{j=1}^{n_i-1} \frac{v_{i,j}\lambda_\theta^{i,j+1,j}}{\left(\Delta_i^{j+1,j}\right)^{1/2}}.$$

From now on, we suppose, without loss of generality, that $t_0 = 0$ and that $n_i = N$ for $i = 1, \ldots, d$. The system (11) cannot be solved explicitly and it is therefore necessary to use a numerical method, such as Newton-Raphson. Hence, an initial approximation is required. An initial solution of σ^2 is calculated by performing a simple linear regression of $\sigma_i^2 = 2\log(m_i/m_i^g)$ where m_i denotes the sample mean and m_i^g the geometric sample mean. Whereas, an initial solution for the coefficients β and η is obtained by a linear regression taking as data the pairs $\left(t_i, -\log\left(\frac{m_N}{m_i} - 1\right)\right)$ where m_N is the last value of the sample mean.

Alternatively, one can obtain the estimates of ξ by maximizing the likelihood function

$$\tilde{L}(\xi) = -\frac{n}{2}\log\sigma^2 - \frac{Z_1 + \Phi_\xi - 2\Gamma_\xi}{2\sigma^2},$$

where

$$n = \sum_{i=1}^{d}(n_i - 1), \quad Z_1 = \sum_{i=1}^{d}\sum_{j=1}^{n_i-1} \frac{1}{\Delta_i^{j+1,j}} \log^2 \frac{x_{i,j+1}}{x_{i,j}}$$

$$\phi_\xi = \sum_{i=1}^{d}\sum_{j=1}^{n_i-1} \frac{(m_\xi^{i,j+1,j})^2}{\Delta_i^{j+1,j}}, \quad \Gamma_\xi = \sum_{i=1}^{d}\sum_{j=1}^{n_i-1} \frac{1}{(\Delta_i^{j+1,j})^{1/2}} \log\frac{x_{i,j+1}}{x_{i,j}}\, m_\xi^{i,j+1,j}.$$

To maximize the function $\tilde{L}$, we use a meta-heuristic optimization method, namely Simulated Annealing (S.A.). This algorithm (see as a reference Kirkpatrick *et al.* (1983)) is used for problems like finding $\arg\min_{\theta\in\Theta} f(\theta)$ and in recent years also in the context of parameters estimation (cf. da Luz Sant'Ana *et al.* (2018) and Román-Román and Torres-Ruiz (2015)). At any step, S.A. generates a new solution in a neighborohood of the previous one and (i) if the new solution improves the objective function, then it replaces the previous, otherwise (ii) if the new solution does not improve the objective function, then it can replace the previous with a probability rate which depends on the increase of the objective function and on a scale factor, called temperature, in agreement with the metallurgical annealing that inspires the method. S.A. avoids in this way local minima but it needs a restriction of the parametric space Θ. In our context of interest, the set Θ contains the parameters ξ. Until now, it is continuous and unbounded, since $\Theta = \{(\eta, \beta^T, \sigma^2) : \eta > 0,\, \beta_1, \ldots, \beta_{p-1} \in \mathbb{R},\, \beta_p > 0,\, \sigma^2 > 0\}$. To bound Θ, we consider $0 < \sigma < 0.1$ so that the simulated sample paths are less variable around the sample mean and thus the multi-sigmoidal logistic profile is advisable. For the parameters $\beta = (\beta_1, \ldots, \beta_p)^T$, we consider the confidence intervals, found by using the data of the polynomial regression performed previously for the initial solutions. Specifically, for β we consider the confidence intervals of the coefficients of the polynomial regression of $-\log\left[\left(\frac{m_N}{m_j} - 1\right)\hat{\eta}\right]$ against t_j, for $j = 1, \ldots, N$, where $\hat{\eta} = \left(\frac{m_N}{m_1} - 1\right)^{-1}$. Finally, for η, we consider the interval (a, b) where

$$a = \min_{1 \le i \le d} \left(\frac{x_{i,n_i}}{x_{i,1}} - 1\right)^{-1}, \qquad b = \max_{1 \le i \le d} \left(\frac{x_{i,n_i}}{x_{i,1}} - 1\right)^{-1}.$$

Regarding the distributions of the MLEs, it is worth to notice that the exact distribution of $\hat{\mu}_1$ is Gaussian $\mathcal{N}(\mu_1, \sigma_1^2/d)$ and the one of $d\hat{\sigma}_1^2/\sigma_1^2$ is chi-square χ^2_{d-1}. Furthermore, the asymptotic distribution of $\hat{\xi}$ is a $(p+2)$-dimensional normal distribution with mean ξ and covariance matrix $I(\xi)^{-1}$, where $I(\xi) \in \mathbb{R}^{(p+2)\times(p+2)}$ is the Fisher information matrix and can be expressed as

$$I(\xi) = \frac{1}{\sigma^2} \begin{pmatrix} \Xi_\xi & -\frac{1}{2}\left(\frac{\partial}{\partial \theta}\gamma_\xi\right) \\ -\frac{1}{2}\left(\frac{\partial}{\partial \theta}\gamma_\xi\right)^T & \frac{n}{2\sigma^2} - \frac{Z_3}{4} \end{pmatrix},$$

where $\Xi_\xi \in \mathbb{R}^{(p+1)\times(p+1)}$ and $\dfrac{\partial}{\partial \theta}\gamma_\xi \in \mathbb{R}^{(p+1)\times 1}$ are defined as

$$\Xi_\xi = \sum_{i=1}^{d} \sum_{j=1}^{n_i - 1} \left(\Delta_i^{j+1,j}\right)^{-1} \left(\frac{\partial}{\partial \theta} m_\xi^{i,j+1,j}\right) \left(\frac{\partial}{\partial \theta} m_\xi^{i,j+1,j}\right)^T$$

and

$$\frac{\partial}{\partial \theta}\gamma_\xi = \sum_{i=1}^{d} \sum_{j=1}^{n_i - 1} \frac{\partial}{\partial \theta} m_\xi^{i,j+1,j}.$$

6. SIMULATIONS

A simulation study is developed to verify the validity of the two aforementioned procedures. As a case study, we use a pattern of 100 independent sample paths simulated by using the expression of the process obtained as the solution of the stochastic differential equation (9). All the sample paths contain the same sumber of data (that is 501), being $(i-1)\cdot 0.1$ for $i=1,\dots,501$ the observation times. The parameters used for the simulation are $\eta = e^{-1}$, $\beta_1 = 0.1$, $\beta_2 = -0.009$, $\beta_3 = 0.0002$, $\sigma = 0.01$. See Figure 6 for the plot of the paths. For simplicity, we have chosen a degenerate initial

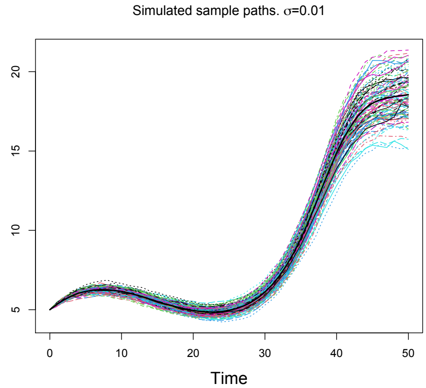

Figure 6: *100 simulated sample paths of the process $X(t)$ for $p=3$, $t_0=0$, $x_0=5$, $Q_\beta(t) = 0.1t - 0.009t^2 + 0.0002t^3$, $\eta = e^{-1}$ and $\sigma = 0.01$.*

distribution centered in $x_0 = 5$, i.e. $\mathbb{P}(X_0 = 5) = 1$. After obtaining each trajectory, we chose 51 values from the first one and using a step equal to 1. The MLEs obtained by solving the system (11) are summarized in Table 1.

Table 1: *The MLEs obtained by solving the system* (11)*.*

	Real	Initial	Estimations	Rel. Err.
η	$e^{-1} = 0.3678794$	0.3695601	0.3695532	$1.673743e\text{-}03$
β_1	0.1	-0.04606928	0.10021729	$2.172891e\text{-}04$
β_2	-0.009	-0.003311808	-0.009030270	$2.030270e\text{-}03$
β_3	0.0002	0.0001356439	0.0002006625	$6.625399e\text{-}07$
σ	0.01	$9.912749e\text{-}03$	$9.982475e\text{-}03$	$3.502022e\text{-}07$

As a further case study, we use the same pattern as before by applying S.A. Moreover, since S.A. is a meta-heuristic algorithm, we apply the procedure 10 times and then we consider the mean of the resulting values. Clearly, if the number of replications increases then the goodness of the results improves but also the computational cost. The MLEs obtained in this way are summarized in Table 2.

Table 2: *The MLEs obtained via S.A.*

	Real	Range	Estimations	Abs. Err.
η	$e^{-1} = 0.3678794$	$[0.287803710, 0.405281061]$	0.3862679	0.04998513
β_1	0.1	$[0.094192479, 0.155306328]$	0.1099219	0.09921900
β_2	-0.009	$[-0.012033562, -0.009118551]$	-0.009722108	0.08023422
β_3	0.0002	$[0.000206216, 0.000245301]$	0.0002131091	0.06554550
σ	0.01	$[0.000000000, 0.010000000]$	0.0001283712	0.13301015

7. CONCLUSIONS

In this paper, we considered the deterministic multi-sigmoidal logistic function and we used the presented model to describe the double-sigmoidal growth of coffee berries. In order to make the model more realistic, we analysed its stochastic counterpart. More in detail, we studied two different birth-death processes, the former with linear rates and the latter with quadratic rates. From the last one, we derive a diffusive approximation by means of a suitable scaling. Then, we found the MLEs of the parameters of the diffusion process by using two different strategies: by solving a non-linear system and by maximizing the log-likelihood function via S.A. We also studied the asymptotic distribution of the resulting MLEs. Finally, to validate the described procedures, we performed a simulation study. Future investigations will be devoted to determine the degree p of the polynomial Q_β, since it is unknown a priori and to use different meta-heuristic strategies to find nice MLEs in a shorter computational time. It will be interesting also to consider a real application based on the diffusion process.

ΠΕΡΙΛΗΨΗ

Θεωρούμε μια πολύ-σιγμοειδή γενίκευση του μοντέλου λογιστικής αύξησης. Το ντετερμινιστικό μοντέλο παρουσιάζεται μαζί με το αντίστοιχό του στοχαστικό μοντέλο. Ειδικότερα, αναλύουμε δύο διαφορετικές διαδικασίες γέννησης-θανάτου με γραμμικούς και τετραγωνικούς ρυθμούς μεταβάσεων, αντίστοιχα. Από το τελευταίο δίνουμε μια πιο εύχρηστη προσέγγιση διάχυσης μέσω μιας κατάλληλης κλιμάκωσης. Επιπλέον, μελετάμε δύο πιθανές στρατηγικές για να υπολογίσουμε τους εκτιμητές μέγιστης πιθανοφάνειας των παραμέτρων του μοντέλου. Για την επαλήθευση των περιγραφόμενων διαδικασιών, παρουσιάζουμε μία μελέτη προσομοίωσης. Διερευνάται επίσης το πρόβλημα του πρώτου χρόνου διέλευσης.

Acknowledgements: Antonio Di Crescenzo and Paola Paraggio are members of the research group GNCS of INdAM (Istituto Nazionale di Alta Matematica). This work was supported in part by the Ministerio de Economía Industria y Competitividad, Spain, under Grant MTM2017-85568-P, FEDER/Junta de Andalucía-Consejería de Economía y Conocimiento, under Grant A-FQM-456-UGR18 and by Italian MIUR-PRIN 2017, project ”Stochastic Models for Complex Systems”, No. 2017HJFFHSH. Paola Paraggio thanks the Department of Statistics and Operations Research, Faculty of Sciences of the University of Granada and the Institute of Mathematics of the University of Granada (IMAG) for the hospitality during the one-month visit carried out in 2019.

REFERENCES

da Cunha, A. R.; Volpe, C.A. (2011) Growth curves of coffee fruits Obatã IAC 1669-20 in different alignments planting. *Semina: Ciências Agrárias, Londrina*, **32**, 1, 49–62.

Di Crescenzo, A.; Paraggio, P. (2019) Logistic growth described by birth-death and diffusion processes. *Mathematics* **7 (6)**, 1–28, 489.

Di Crescenzo, A.; Paraggio, P.; Román-Román, P.; Torres-Ruiz, F. (2021a) Applications of the multi-sigmoidal deterministic and stochastic logistic models for plant dynamics. *Appl Math Model*, **92**, 884–904.

Di Crescenzo, A.; Paraggio, P.; Román-Román, P.; Torres-Ruiz, F. (2021b) Statistical analysis and first-passage-time applications of a lognormal diffusion process with multi-sigmoidal logistic mean (submitted for publication).

Di Crescenzo, A.; Spina, S. (2016) Analysis of a growth model inspired by gompertz and korf laws, and an analogous birth-death process. *Math Biosci*, **282**, 121–134.

Erto, P.; Giorgio, M.; Lepore, A. (2020) The Generalized Inflection S-Shaped Software Reliability Growth Model. *IEEE Trans Reliab*, **69**, 1, 228–244.

Kirkpatrick, S.; Gelatt, C.D.; Vecchi, M.P. (1983) Optimization by simulated annealing. *Science, New Series*, **220**, 4598, 671–680.

da Luz Sant'Ana, I.; Román-Román, P.; Torres-Ruiz, F. (2018) The Hubbert diffusion process: Estimation via simulated annealing and variable neighborhood search procedures - application to forecasting peak oil production. *Appl Stochastic Models Bus Ind* **34**, 376–394.

Rajasekar, S.P.; Pitchaimani, M.; Quanxin Zhu (2020) Progressive dynamics of a stochastic epidemic model with logistic growth and saturated treatment. *Physica A* **538**, 122649, 1–20.

Román-Román, P.; Torres-Ruiz, F. (2015) The nonhomogeneous lognormal diffusion process as a general process to model particular types of growth patterns. *Lecture Notes of Seminario Interdisciplinare di Matematica* **12**, 201–219.

Román-Román, P.; Serrano-Pérez, J.J.; Torres-Ruiz, F. (2019) A note on estimation of multi-sigmoidal Gompertz functions with random noise. *Mathematics* **7**, 541, 1–18.

Tan, W.Y. (1986) A stochastic Gompertz birth-death process. *Stat Prob Lett* **4**, 25–28.