On a Non-homogeneous Gompertz-Type Diffusion Process: Inference and First Passage Time

Giuseppina Albano[1(✉)], Virginia Giorno[2], Patricia Román-Román[3], and Francisco Torres-Ruiz[3]

[1] Dip. di Scienze Economiche e Statistiche, Università di Salerno, Fisciano, Italy
pialbano@unisa.it
[2] Dip. di Informatica, Università di Salerno, Fisciano, Italy
giorno@unisa.it
[3] Dpto. de Estadística e Investigación Operativa, Universidad de Granada, Granada, Spain
{proman,fdeasis}@ugr.es

Abstract. A stochastic diffusion model based on a generalized Gompertz deterministic growth in which the carrying capacity depends on the initial size of the population is considered. The growth parameter of the process is then modified by introducing a time-dependent exogenous term. The first passage time problem is considered and a two-step procedure to estimate the model is proposed. Simulation study is also provided for suitable choices of the exogenous term.

1 Introduction

The models for the description of growth phenomena, originally associated to the evolution of animals populations, currently play an important role in several fields like that economic, biological, medical, ecological, among others. For this reason many efforts are oriented to the development of always more sophisticated mathematical models for the description of a particular type of behaviour. The most representative curves for modeling growth are of exponential-type as the logistic and Gompertz curves because they are characterized by a finite carrying capacity, that represents, in general terms, the limitation of the natural resources. Specifically, the Gompertz curve is frequently used because in several contexts it seems to fit the experimental data in enough precise way. It is described by the equation:

$$x(t) = \exp\Big\{ \frac{m}{\beta}[1 - e^{-\beta(t-t_0)}] + \ln x_0 \, e^{-\beta(t-t_0)}\Big\}, \tag{1}$$

where m and β are positive constants that represent the growth parameters of the population. Equation (1) is able to describe growth dynamics in a lot of contexts

so, for instance animal, vegetable, tumor growth. Equation (1) is solution of the following ordinary differential equation:

$$\frac{dx}{dt} = m\beta - \beta x \ln x, \qquad x(t_0) = x_0.$$

We note that in (1) the carrying capacity is $K = \lim_{t\to\infty} x(t) = \exp\{m/\beta\}$. However, in several contexts, the carrying capacity can depend from the initial size of the population (cf. [4,6,8], for example). In order to take into account this aspect, in [8] the authors modified Eq. (1) as follows

$$x(t) = x_0 \exp\left\{\frac{m}{\beta} \left[e^{-\beta t_0} - e^{-\beta t}\right]\right\}. \tag{2}$$

Equation (2) is the solution of

$$\frac{dx}{dt} = m\,e^{-\beta t}x, \qquad x(t_0) = x_0. \tag{3}$$

We note explicitly that the limit for long time of (2) depends from the initial size of the population being $K = x_0 \exp\{m/\beta\,e^{-\beta t_0}\}$.

In this paper, we consider the stochastic diffusion process associated to (2), denoted by $X(t)$. Then we derive the process $X_C(t)$ by modifying the growth parameter m to introduce the effect of a *therapy* interpreted as a continuous time dependent function $C(t)$. We study both the processes by focusing on the First Passage Time (FPT) problem. Moreover, in experimental studies the effect of a new therapy has to be tested so the term $C(t)$ is usually unknown. The knowledge of such functional form is fundamental since it allows to introduce an external control to the system and to explain how the therapy acts. Further, the study of some problems related to the process $X_C(t)$, i.e. modeling and forecasting, requires the knowledge of $C(t)$. For these reasons, the functional form of $C(t)$ has to be estimated. In this direction we propose a two-step estimation procedure applicable when data from a control group, modeled by means of $X(t)$, and from one or more treated groups, described by $X_C(t)$, are available. In the first step the parameters of the control group, are estimated by maximum likelihood method (see [8,9]). In the second step the function $C(t)$ is estimated making use of relationships between the processes $X(t)$ and $X_C(t)$. Finally, in order to evaluate the goodness of the proposed procedure a simulation study is presented.

The paper is organized as follows. In Sect. 2 the stochastic model $X_C(t)$ is introduced, the transition distribution and the related moments are provided. In Sect. 3 we study the FPT through suitable boundaries of interest in real applications. In Sect. 4 a two-step procedure is proposed to estimate the parameters of $X_C(t)$. A simulation is also provided to validate the procedure. Some conclusions close the paper.

2   The Stochastic Model

In the following we consider the stochastic version of the Eq. (3). Precisely, let $X(t)$ be a stochastic process defined in $\mathbb{R}^+$ described by the following stochastic differential equation (SDE)

$$dX(t) = m\,e^{-\beta t}X(t)dt + \sigma X(t)\,dW(t), \qquad X(t_0) = x_0 \ a.s. \tag{4}$$

It can be obtained from (3), following a standard procedure (see, for instance [5]). The parameters m and β are positive constants that represent the growth rates of the population $X(t)$, $\sigma > 0$ is the width of random fluctuations and $W(t)$ represents a standard Brownian motion.

In real-life situations, intrinsic growth rates of the population can be modify by means of exogenous terms generally depending on time. Examples of such situations could be suitable food treatments in growth of animals (see [1]) or therapeutic treatments in tumor growth (see for instance [2,3,7]). In order to model such situations, we modify the drift of $X(t)$ by introducing a continuous time dependent function $C(t)$ to model the effect of an exogenous factor, namely therapy, obtaining the stochastic process $X_C(t)$ described by the following SDE

$$dX_C(t) = [m - C(t)]e^{-\beta t}X_C(t)dt + \sigma X_C(t)dW(t), \qquad X_C(t_0) = x_0 \ a.s. \tag{5}$$

The solution of (5) is a diffusion process with sample paths

$$X_C(t) = x_0 \exp\left\{ \int_{t_0}^{t} [m - C(s)]\,e^{-\beta s}\,ds - \frac{\sigma^2}{2}(t - t_0) + \sigma\,[W(t) - W(t_0)] \right\}. \tag{6}$$

Clearly, the solution of (4) can be obtained by (6) choosing $C(t) = 0$. Equation (5) defines a stochastic diffusion process taking values in $\mathbb{R}^+$, characterized by drift and infinitesimal variance given by

$$A_1^C(x,t) = [m - C(t)]e^{-\beta t}x, \qquad A_2^C(x) = \sigma^2 x^2,$$

respectively. Let $f_C(x,t|y,\tau) = \frac{\partial}{\partial x}P[X_C(t) \leq x|X_C(\tau) = y]$ be the transition probability density function (pdf) of $X_C(t)$. The function $f_C(x,t|y,\tau)$ is solution of the Fokker-Planck equation:

$$\frac{\partial f_C(x,t|y,\tau)}{\partial t} = -[m-C(t)]e^{-\beta t}\frac{\partial}{\partial x}[xf_C(x,t|y,\tau)] + \frac{\sigma^2}{2}\frac{\partial}{\partial x^2}[x^2 f_C(x,t|y,\tau)] \tag{7}$$

and of the Kolmogorov equation:

$$\frac{\partial f_C(x,t|y,\tau)}{\partial \tau} + [m - C(\tau)]e^{-\beta \tau}y\frac{\partial f_C(x,t|y,\tau)}{\partial y} + \frac{\sigma^2 y^2}{2}\frac{\partial f_C(x,t|y,\tau)}{\partial y^2} = 0. \tag{8}$$

Furthermore, $f_C(x,t|y,\tau)$ satisfies the initial delta condition: $\lim_{t\to\tau} f_C(x,t|y,\tau) = \lim_{\tau\to t} f_C(x,t|y,\tau) = \delta(x - y)$. Note that the transformation

$$z = \ln x + d(t), \qquad z_0 = \ln y + d(\tau)$$

with

$$d(t) = \frac{\sigma^2}{2}t + \frac{m}{\beta}e^{-\beta t} + \int^{t} C(\theta)e^{-\beta \theta}d\theta,$$

reduces (7) and (8) to the analogous equations for a Wiener process $Z(t)$ defined in $\mathbb{R}$ with drift and infinitesimal variance $B_1 = 0, B_2 = \sigma^2$, respectively. So one obtains

$$f_C(x, t|y, \tau) = \frac{1}{x\sqrt{2\pi\sigma^2(t-\tau)}} \exp\left\{-\frac{[\ln(x/y) + d(t) - d(\tau)]^2}{2\sigma^2(t-\tau)}\right\}.$$

Moreover one has

$$E[X_C^k(t)|X_C(\tau) = y] = y \exp\left\{-k\left[\frac{m}{\beta}\left(e^{-\beta t} - e^{-\beta\tau}\right)\right.\right.$$
$$\left.\left. + \int_\tau^t C(\theta)e^{-\beta\theta}\,d\theta\right] + \frac{k(k-1)}{2}\sigma^2(t-\tau)\right\}.$$

3   First Passage Time Through a Single Boundary

Let

$$T_{x_0, S(t)}^C = \begin{cases} \inf\limits_{t \geq t_0}\{t \ : \ X_C(t) > S(t)|X_C(t_0) = x_0\}, x_0 < S(t_0) \\ \inf\limits_{t \geq t_0}\{t \ : \ X_C(t) < S(t)|X_C(t_0) = x_0\}, x_0 > S(t_0) \end{cases}$$

be the FPT of $X_C(t)$ through the boundary $S(t)$ and let $g_C[S(t), t|x_0, t_0]$ be the FPT pdf. If $S(t) \in C^2[t_0, +\infty)$ the FPT pdf $g_C[S(t), t|x_0, t_0]$ is solution of the following second kind Volterra integral equation:

$$g_C[S(t), t|x_0, t_0] = 2\rho\left[\psi_C[S(t), t|x_0, t_0] - \int_{t_0}^t g_C[S(\tau), \tau|x_0, t_0]\,\psi_C[S(t), t|S(\tau), \tau]d\tau\right]$$

with

$$\psi_C[S(t), t|y, \theta] = \frac{f_C[S(t), t|y, \theta]}{2}\left\{S'(t) - [m - C(t)]e^{-\beta t}S(t)\right.$$
$$\left. -S(t)\frac{\ln[S(t)/S(\theta)] - \int_\theta^t [m - C(s)]e^{-\beta s}\,ds}{t - \theta}\right\}$$

Note that if

$$S(t) = A\exp\left\{Bt + \int_{}^{t} [m - C(s)]\,e^{-\beta s}\,ds\right\}$$

with $A > 0$ and $B \in \mathbb{R}$, then $\psi_C[S(t), t|S(\tau), \tau] = 0, \forall\tau \in [t_0, t]$, so that the FPT pdf can be expressed in the following closed form:

$$g_C[S(t), t|x_0, t_0] = \frac{\left|\ln\frac{S(t_0)}{x_0}\right|}{\sqrt{2\pi\sigma^2(t-t_0)^3}} \exp\left\{-\frac{\left[(\frac{\sigma^2}{2}+B)(t-t_0)+\ln\frac{S(t_0)}{x_0}\right]^2}{2\sigma^2(t-t_0)}\right\}, \quad S(t_0) \neq x_0.$$

Moreover, by choosing in (9) $C(t) = Be^{\beta t}$ and $A = px_0\exp\left\{\frac{m}{\beta}e^{-\beta t_0}\right\}$, one has

$$S(t) = pE[X(t)|X(t_0) = x_0] = px_0\exp\left\{\frac{m}{\beta}(e^{-\beta t_0} - e^{-\beta t})\right\} \tag{9}$$

that, for $0 < p < 1$, represents a percentage of the mean of the process $X(t)$. In other words, for the process $\{X_C(t); t_0 \leq t\}$ characterized by infinitesimal moments

$$A_1^C(x,t) = [m - Be^{\beta t}]e^{-\beta t}x, \quad A_2^C(x) = \sigma^2 x^2, \tag{10}$$

the FPT pdf through boundary (9) is given by

$$g_C[S(t), t | x_0, t_0] = \frac{|\ln p|}{\sqrt{2\pi\sigma^2(t-t_0)^3}} \exp\left\{-\frac{[(\sigma^2/2 + B)(t - t_0) + \ln p]^2}{2\sigma^2(t - t_0)}\right\}. \tag{11}$$

For the process defined in (10) with $m = 0.75, \beta = 0.18, \sigma = 0.07$, in Fig. 1 the FPT pdf (11) is plotted for $p = 0.7$ (on the left) and $p = 0.85$ (on the right) for various choices of B.
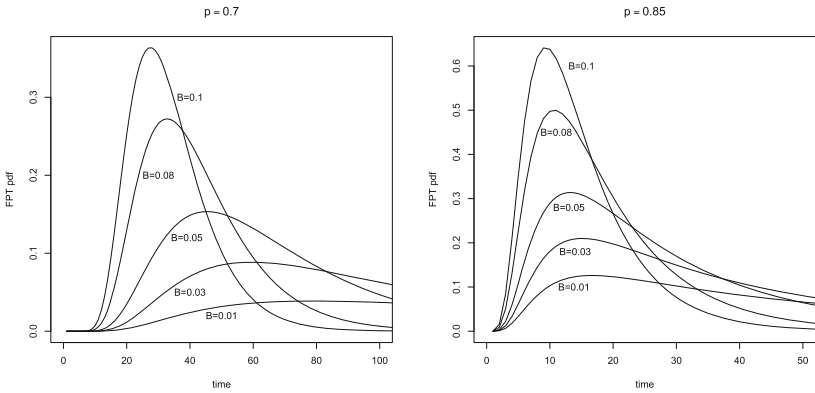
Fig. 1. FPT pdf (11) for $X_C(t)$ in (10) with $m = 0.75, \beta = 0.18, \sigma = 0.07$ through $S(t)$ in (9) for $p = 0.7$ (left) and $p = 0.85$ (right) and some values of B.

4   Inference

In this section we propose a two step estimation procedure that can be used when data from a control group and from one or more treated groups are available. In the first step, from the control group, modeled by means of $X(t)$, the parameters m, β and σ are estimated by maximum likelihood method (see [8,9]). In the second step the function $C(t)$ is estimated making use of some relationships relating the process $X(t)$ describing the control group, i.e. an untreated group, and $X_C(t)$ modeling the treated group. The idea is to take the model $X(t)$ as a starting point and then to use the information provided by the treated group to fit the function $C(t)$. Therefore, after estimating the parameters of $X(t)$, $C(t)$ is estimated by the trajectories of the treated and non treated groups by means of suitable relations between the two models.

In order to relate the trajectories of the processes $X_C(t)$ and $X(t)$, we assume that $X_C(t_0) = X(t_0) = x_0$, i.e. the therapy is applied from time t_0, so that from Eq. (6) one obtains:

$$X_C(t) = \exp\left\{ -\int_{t_0}^t C(s)\, e^{-\beta s} ds \right\} X(t). \tag{12}$$

From (12), looking at the conditional mean functions, we find

$$E[X_C(t)|X_C(t_0) = x_0] = \exp\left\{ -\int_{t_0}^t C(s)\, e^{-\beta s} ds \right\} E[X(t)|X(t_0) = x_0]$$

from which we have

$$C(t) = -e^{\beta t} \frac{d}{dt} \ln \left(\frac{E[X_C(t)|X_C(t_0) = x_0]}{E[X(t)|X(t_0) = x_0]} \right).$$

4.1 Proposed Methodology

The data required for the proposed strategy are the values of d_1 sample paths of a non-treated group ($x_{ij}, i = 1, \ldots, d_1,\ j = 1, \ldots, n$) and d_2 sample paths of a treated group ($x_{ij}^C, i = 1, \ldots, d_2,\ j = 1, \ldots, n$), observed in the same time instants $t_1, \ldots, t_n$.

- From the data of the control group, estimate the parameters of process $X(t)$. From this first step, we obtain ML estimations $\widehat{m}, \widehat{\beta}$ and $\widehat{\sigma}^2$.
- Denoting by x_j and x_j^C the means of x_{ij} and of x_{ij}^C at any instant t_j, respectively, i.e.

$$x_j = \frac{1}{d_1} \sum_{i=1}^{d_1} x_{ij}, \qquad x_j^C = \frac{1}{d_2} \sum_{i=1}^{d_2} x_{ij}^C,$$

we obtain

$$\gamma_j = -\ln \left[\frac{x_j^C}{x_j} \right].$$

- Interpolate the points γ_j for $j = 1, 2, \ldots n$ (for example by using cubic spline interpolation) obtaining the function $\widehat{\gamma}(t)$. Finally, consider the following function as an approximation of $C(t)$.

$$\widehat{C}(t) = -e^{\widehat{\beta} t}\, \widehat{\gamma}\,'(t).$$

4.2 A Simulation Study

In order to evaluate the goodness of the proposed procedure we present a simulation study. We consider some specific functions for therapies: constant, linear, logarithmic and periodic. 50 sample paths of the control group $X(t)$ with $m = 0.1$, $\beta = 0.01$, $\sigma = 0.01$ and $t_0 = 0$ have been simulated assuming a random initial state x_0 chosen according $\Lambda(1, 0.16)$.

The paths include 300 observations of the process starting from $t_1 = t_0 = 0$ with $t_i - t_{i-1} = 2$.

The first step of the procedure gives the estimation of the control group parameters: $\widehat{m} = 0.09578, \widehat{\beta} = 0.01085$ and $\widehat{\sigma} = 0.0157$. Hence, as control group we consider the stochastic process $X(t)$ with infinitesimal moments

$$A_1(x,t) = 0.09578\,e^{-0.01085\,t}\,x, \qquad A_2(x) = 0.0157^2\,x^2. \tag{13}$$

Then, 50 sample paths of the treated group $X_C(t)$ have been simulated with $X_C(t_0) = X(t_0) = x_0$ and considering the following therapies: $C(t) = -0.005, C(t) = \pm 0.001t, C(t) = 0.005\sin(t), C(t) = 0.02\ln(1 + 0.15t)$. The results obtained by applying the proposed procedure are shown in Fig. 2. The dashed curves represent the functions $C(t)$ whereas the full curves represent the corresponding estimation $\widehat{C}(t)$. We note explicitly that in the considered cases the proposed procedure is able to capture the trend of $C(t)$. To evaluate the performance of the proposed procedure we calculate the mean absolute error (MAE), root mean square error (RMSE) and

$$d = 1 - \frac{\sum_{i=1}^{N}(\widehat{C}(t_i) - C(t_i))^2}{\sum_{i=1}^{N}\left(|\widehat{C}(t_i) - \overline{C(t)}| + |C(t_i) - \overline{C(t)}|\right)^2}$$

where N is the number of estimated values for the considered cases and $\overline{C(t)}$ is the mean of function $C(t)$. The results are shown in Table 1 for the aforementioned cases. For all the choices of the function $C(t)$ the procedure provides very satisfactory estimates of the function $C(t)$.
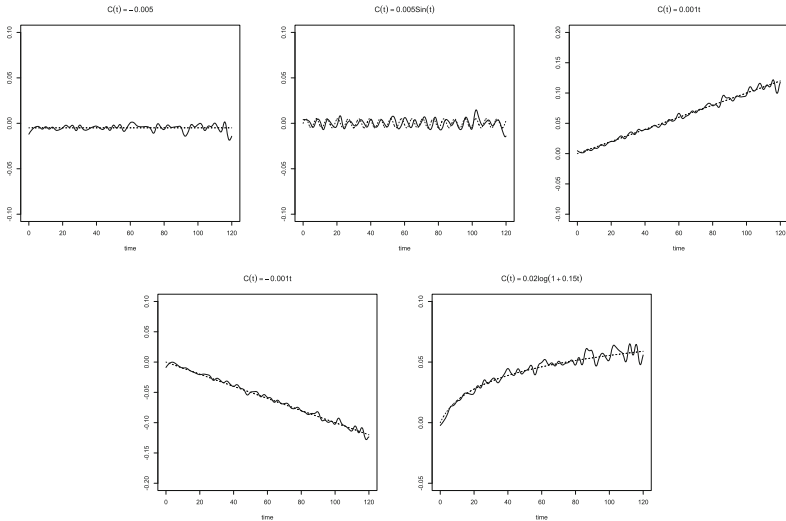
Fig. 2. For the process (13), $C(t)$ and its estimate are shown for different cases.

Table 1. MAE, RMSE and d for the considered therapies.

$C(t)$	MAE	RMSE	d
-0.005	0.095	0.216	0.999
$0.005\sin t$	0.081	0.169	0.995
$0.001t$	0.0918	0.219	0.999
$-0.001t$	0.085	0.175	0.999
$0.02\log(1+0.15t)$	0.086	0.181	0.998

5   Conclusions

We have analyzed a non-homogeneous Gompertz-type stochastic diffusion processes characterized by a carrying capacity depending on the initial state. For such a process we have considered a perturbation of a growth parameter by introducing the effect of a exogenous term $C(t)$ generally depending on time and we have studied the first passage time of the process through a time dependent boundary. Moreover, a two-step procedure has been proposed in order to estimate the model parameters and to fit the function $C(t)$ when data from a control group and one or more treated groups are available. Our simulation study has shown that the proposed procedure is able to capture the trend of $C(t)$ in a very satisfactory way.

References

1. Albano, G., Giorno, V., Román-Román, P., Torres-Ruiz, F.: On the therapy effect for a stochastic Gompertz-type model. Math. Biosci. **235**, 148–160 (2012)
2. Albano, G., Giorno, V., Román-Román, P., Torres-Ruiz, F.: Inference on a stochastic two-compartment model in tumor growth. Comput. Stat. Data Anal. **56**, 1723–1736 (2012)
3. Albano, G., Giorno, V., Román-Román, P., Román-Román, S., Torres-Ruiz, F.: Estimating and determining the effect of a therapy on tumor dynamics by means of a modified Gompertz diffusion process. J. Theor. Biol. **364**, 206–219 (2015)
4. Blasco, A., Piles, M., Varona, L.: A Bayesian analysis of the effect of selection for growth rate on growth curves in rabbits. Genet. Sel. Evol. **35**, 21–41 (2003)
5. Capocelli, R.M., Ricciardi, L.M.: Growth with regulation in random environment. Kibernetik **15**, 147–157 (1974)
6. Di Crescenzo, A., Spina, S.: Analysis of a growth model inspired by Gompertz and Korf laws, and an analogous birth-death process. Math. Biosci. **282**, 121–134 (2016)
7. Giorno, V., Román-Román, P., Spina, S., Torres-Ruiz, F.: Estimating a non-homogeneous Gompertz process with jumps as model of tumor dynamics. Comput. Stat. Data Anal. **107**, 18–31 (2017)
8. Gutiérrez-Jáimez, R., Román-Román, P., Romero, D., Serrano, J.J., Torres-Ruiz, F.: A new Gompertz-type diffusion process with application to random growth. Math. Biosci. **208**, 147–165 (2007)
9. Román-Román, P., Romero, D., Rubio, M.A., Torres-Ruiz, F.: Estimating the parameters of a Gompertz-type diffusion process by means of simulated annealing. Appl. Math. Comput. **218**, 5121–5131 (2012)