



---

*Research article*

## **Probabilistic analysis of linear-quadratic logistic-type models with hybrid uncertainties via probability density functions**

**Clara Burgos, Juan Carlos Cortés\*, Elena López-Navarro and Rafael Jacinto Villanueva**

Instituto Universitario de Matemática Multidisciplinar, Universitat Politècnica de València, València, Spain

\* **Correspondence:** Email: [jccortes@imm.upv.es](mailto:jccortes@imm.upv.es).

**Abstract:** We provide a full stochastic description, via the first probability density function, of the solution of linear-quadratic logistic-type differential equation whose parameters involve both continuous and discrete random variables with arbitrary distributions. For the sake of generality, the initial condition is assumed to be a random variable too. We use the Dirac delta function to unify the treatment of hybrid (discrete-continuous) uncertainty. Under general hypotheses, we also compute the density of time until a certain value (usually representing the population) of the linear-quadratic logistic model is reached. The theoretical results are illustrated by means of several examples, including an application to modelling the number of users of Spotify using real data. We apply the Principle Maximum Entropy to assign plausible distributions to model parameters.

**Keywords:** hybrid uncertainty; random linear-quadratic logistic differential equation; first probability density function; random variable transformation method; uncertainty quantification; Principle Maximum Entropy

**Mathematics Subject Classification:** 34A38, 37H10, 34F05, 60H10

---

### **1. Introduction**

Interaction between discrete and continuous dynamics gives rise to hybrid models. This type of mathematical models mainly appear when describing technological systems, where the continuous dynamics comes from the physical process, while the discrete dynamics appears because of the technological elements of the system [1]. These latter elements, often based upon logical and software tools, are designed to take full control of the technological process [2, 3].

Traditionally, continuous dynamics has been mathematically described via differential equations while automata and Petri nets have been commonly used to model the discrete dynamics. Apart from technological applications, hybrid systems have also been considered to study the dynamics of diseases

in epidemiology. In this setting, hybrid systems have been applied with switching parameters, and to account different control strategies [4–6].

Although most of contributions in the theory of hybrid systems have been developed using deterministic tools, it is natural to consider a stochastic approach, since real systems are affected by uncertainties. The randomness comes from the sampled data used to conduct the corresponding study or from the partial knowledge of the phenomenon under analysis due to inherent complexity that often is present in real-world. This approach leads to stochastic/random hybrid systems where the term *hybrid* is also motivated because of the combination of different types of uncertainties, whose nature can be both continuous and discrete, and that may alter the dynamics of the system.

In dealing with differential equations with uncertainties, one mainly distinguishes two approaches, namely, stochastic differential equations (SDEs) and random differential equations (RDEs) [7, p. 96–98]. SDEs are those where uncertainty is forced (preselected) via stochastic processes whose trajectories or sample behaviour are very irregular, but following specific probabilistic patterns such as the Gaussian Wiener process, whose realizations are nowhere differentiable. The solution of this class of SDEs requires a special stochastic calculus, such as the Itô calculus [8]. RDEs are those in which random effects are directly represented in their input parameters (coefficients, forcing term, initial/boundary conditions). It is usually assumed that these inputs have regular behaviour (e.g., continuous trajectories) with respect to time and/or space [9]. A major advantage of RDEs is their flexibility when assigning probabilistic distributions for each input parameter. This is a key point when RDEs are used to model real-world problems, since we do not prefix the type of uncertainty that appear in the mathematical model but it is determined using inverse probabilistic techniques to assign appropriate distributions to each input parameter [7].

In the setting of hybrid models with uncertainties, SDEs have been widely used. In their formulation these SDEs combine both continuous and discrete stochastic processes. As mentioned before, such processes are previously chosen when formulating the SDEs, so this choice might condition the suitability of the corresponding SDE to describe the specific problem under analysis. This approach has been mainly applied by combining Wiener and Poisson processes [10]. The Wiener process accounts the continuous dynamics as it has continuous trajectories, while the compound Poisson process describes the discrete dynamics as its trajectories have jumps. Lévy-type processes have also been considered as continuous and discrete stochastic terms of SDEs since both the Wiener and the Poisson processes can be regarded as particular cases of the Lévy process [11]. In contrast, to the best of our knowledge, there is a lack of contributions dealing with RDEs whose input parameters combine hybrid uncertainties, i.e., continuous (like Gaussian, Beta, Gamma, etc.) and discrete (like Binomial, Poisson, Geometric, etc.) random variables. In this paper, we tackle this interesting scenario by studying a full randomization of an important dynamic model, namely, the linear-quadratic logistic-type differential equation. As it will be seen later, for the sake of generality, we will assume that model parameters (the coefficients and the initial condition) involve both discrete and continuous random variables. From a stochastic standpoint, this fact confers the problem hybrid nature.

It is important to stress that solving a SDE/RDE consists not only in obtaining, exactly or approximately, the solution stochastic process, say  $X(t)$ , but also its main probabilistic information, usually given via the first moments, such as the mean ( $\mathbb{E}[X(t)]$ ), the variance ( $\mathbb{V}[X(t)]$ ), etc. However, a more ambitious objective is determining the so called *fidis* (finite distributions) of the

solution [9, Ch.3]. The first probability density function (1-PDF),  $f_{X(t)}(x)$ , is the simplest but the most important *fidis* since from it one can calculate, by integration, all the one-dimensional moments,

$$\mathbb{E}[(X(t))^m] = \int_{-\infty}^{\infty} x^m f_{X(t)}(x) dx, \quad m = 1, 2, \dots, \quad (1.1)$$

provided they exist. The 1-PDF also permits computing the probability that the solution lies within a specific interval of interest,

$$\mathbb{P}[x_1 \leq X(t) \leq x_2] = \int_{x_1}^{x_2} f_{X(t)}(x) dx,$$

for every  $t$ . In the context of hybrid systems governed by SDEs, whose uncertainty is modelled by combining both continuous and discrete processes, most contributions have focussed on SDEs driven by a Wiener process (continuous uncertainty) and a Poisson process (discrete uncertainty). In such case, the 1-PDF satisfies the forward generalized Kolmogorov partial differential equation [12]. The exact solution of this partial differential equation can be obtained only under exceptional conditions and, in general, numerical methods are required [13]. In the setting of RDEs, the following result, called Random Variable Transformation (RVT) method, allows us to calculate the 1-PDF of RDEs in the case the solution admits an explicit closed-form or when it can be represented via an analytic expression. The RVT technique avoids solving the corresponding partial differential equation satisfied by the 1-PDF of the solution stochastic process, which might be unaffordable. Moreover, the RVT technique has demonstrated to be very useful to solve both ordinary RDEs [14, 15] and partial RDEs [16, 17]. The RVT method is stated in the following theorem.

**Theorem 1** (Random Variable Transformation method). [9, page 25]. *Let us consider  $\mathbf{U} = (U_1, \dots, U_n)$  and  $\mathbf{V} = (V_1, \dots, V_n)$  two  $n$ -dimensional absolutely continuous random vectors defined on a complete probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ . Let  $\mathbf{r} : \mathbb{R}^n \rightarrow \mathbb{R}^n$  be a one-to-one deterministic transformation of  $\mathbf{U}$  into  $\mathbf{V}$ , i.e.,  $\mathbf{V} = \mathbf{r}(\mathbf{U})$ . Assume that  $\mathbf{r}$  is continuous in  $\mathbf{U}$  and has continuous partial derivatives with respect to each  $U_i$ ,  $1 \leq i \leq n$ . Then, if  $f_{\mathbf{U}}(\mathbf{u})$  denotes the joint probability density function of random vector  $\mathbf{U}$ , and  $\mathbf{s} = \mathbf{r}^{-1} = (s_1(v_1, \dots, v_n), \dots, s_n(v_1, \dots, v_n))$  represents the inverse mapping of  $\mathbf{r} = (r_1(u_1, \dots, u_n), \dots, r_n(u_1, \dots, u_n))$ , the joint probability density function of random vector  $\mathbf{V}$  is given by*

$$f_{\mathbf{V}}(\mathbf{v}) = f_{\mathbf{U}}(\mathbf{s}(\mathbf{v})) |J|,$$

where  $|J|$ , which is assumed to be different from zero, is the absolute value of the Jacobian defined by the determinant

$$J = \det \left( \frac{\partial \mathbf{s}}{\partial \mathbf{v}} \right) = \det \begin{pmatrix} \frac{\partial s_1(v_1, \dots, v_n)}{\partial v_1} & \dots & \frac{\partial s_n(v_1, \dots, v_n)}{\partial v_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial s_1(v_1, \dots, v_n)}{\partial v_n} & \dots & \frac{\partial s_n(v_1, \dots, v_n)}{\partial v_n} \end{pmatrix}.$$

It is important to remark that, in order to apply this results, the random vector  $\mathbf{U} = (U_1, \dots, U_n)$  must be absolutely continuous, so having a PDF. To the best of our knowledge, in the contributions where the RVT technique has been applied to determine the 1-PDF of the solution stochastic process of ordinary/partial RDEs, all the inputs parameters are assumed to be absolutely continuous random variables. However, the interesting case where both discrete and continuous random variables are involved in the RDE, has not been already analysed. As we shall explain right-down, in this paper we

tackle this situation by studying the following random initial value problem (IVP), based on a general linear-quadratic logistic equation,

$$\begin{cases} X'(t) = CX(t)(N - X(t)) + D(N - X(t)), \\ X(0) = X_0. \end{cases} \quad (1.2)$$

Here,  $X(t)$  is the unknown stochastic process, and  $C$ ,  $D$ ,  $X_0$  and  $N$  are random variables defined in a common complete probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ . This IVP can be applied to describe the dynamics of many phenomena. For example, in Biology,  $X(t)$  represents the total population at the time instant  $t$ ,  $X_0 > 0$  is the initial condition,  $N > 0$  is the carrying capacity,  $C > 0$  is the logistic-growth rate and  $D > 0$  is the autonomous-growth rate. The above IVP generalizes the classical Verhulst model, also termed logistic model, by including the linear term,  $D(N - X(t))$ , that represents the autonomous growth of the population,  $X(t)$ , at time instant  $t$ . Apart from Biology, the IVP (1.2) can also be used to model the diffusion of a new technology in a market with  $N$  potential customers taking into account the diffusion can be due both to contacts (encounters) between individuals who already use the new technology and the non-users, and the effect of advertising among those who do not use the new technology [18, 19]. Assuming that at the initial time instant,  $t = 0$ , there are  $X_0$  users of the new technology, the instantaneous change of the number of new users at time  $t$ , given by  $X'(t)$ , is calculated as the sum of  $CX(t)(N - X(t))$  and  $D(N - X(t))$ . The first addend,  $CX(t)(N - X(t))$ , can be interpreted as the part of instantaneous rate due to *successful encounters (contagion)* between those non-users,  $N - X(t)$ , and users,  $X(t)$ , of the new technology, being  $C > 0$ , the *contagion* rate parameter. The second addend,  $D(N - X(t))$ , models the part of instantaneous rate of new technology's users because of their autonomous decision (not being influenced by individuals that already use the new technology). For instance, they might adopt the new technology because advertising campaigns. So, coefficient  $D > 0$  can be interpreted as an autonomous-growth rate. In this illustrative example, is clear that parameters  $C$  and  $D$  embed the factors that account for the transition rates from subpopulation  $N - X(t)$  to  $X(t)$  due to *contagion* and to autonomous behaviour, respectively. These factors are not deterministically known because of the difficulty to determine them. This motivates rates are treated as random variables rather than deterministic quantities. Similarly, in real-world problems the number of users of a new technology is usually estimated by sampling, so containing errors and uncertainties. This suggests to consider the initial condition  $X_0$  as a random variable too. We will illustrate these issues in the last example, where the IVP (1.2) will be applied to modelling the dynamics of Spotify users.

The quadratic logistic equation with uncertainties has been studied in recent contributions using the RVT technique, but only in the case that its inputs are absolutely continuous random variables [15, 20]. In this paper, we go further, and we first perform a full probabilistic study of the aforementioned generalized version of the logistic model, and secondly, we consider hybrid randomness in its formulation. To the best of our knowledge, this latter feature has not been dealt in the extant literature. Henceforth, in order to conduct our probabilistic study, we will assume that  $X_0$  is an absolutely continuous random variable with PDF,  $f_{X_0}(x_0)$ ;  $(C, D)$  is an absolutely continuous random vector with joint PDF,  $f_{C,D}(c, d)$ ; and  $N$  is a discrete random variable such that  $\mathbb{P}[\{\omega \in \Omega : N(\omega) = n_i\}] = p_i$ ,  $i \in I$ , where  $\sum_{i \in I} p_i = 1$  and  $I$  is a discrete set that will be specified later. We will further assume that  $X_0$ ,  $(C, D)$  and  $N$  are independent. This latter hypothesis is intuitive taking into account the interpretation of model parameters and initial condition. In particular, observe that random rates  $C$  and  $D$  are probabilistically related since both determine the transitions from

subpopulation  $N - X(t)$  to  $X(t)$ . Anyway, the case where all random variables  $X_0$ ,  $C$ ,  $D$  and  $N$  are independent, which is a more standard assumption, can be easily inferred from our analysis too. In such scenario the PDFs of  $C$  and  $D$  are denoted by  $f_C(c)$  and  $f_D(d)$ , respectively.

The paper is organized as follows. In Section 2 we revise how to express the probability mass function of a discrete random variable via a PDF using the Dirac delta function. This permits unifying the treatment of discrete and continuous random variables. In Section 3, we determine a closed-form expression, in terms of an expectation, of the 1-PDF of the solution stochastic process of the random IVP (1.2). Section 4 is devoted to determine the PDF of a quantity of great interest in applications, namely, time until a given population size is reached. To illustrate these theoretical findings, two examples are included in Section 3 and 4. In these examples, the distributions of model inputs are chosen just to carry out some illustrative simulations. In Section 5, we apply all the results established in Sections 3 and 4 to describe the dynamics of the number of users of Spotify music streaming service using real data. To this end, we use an inverse probabilistic technique, based on the Principle of Maximum Entropy, to assign appropriate probability distributions to model parameters. Conclusions are drawn in Section 6.

Finally, we point out that in order to facilitate the reading of the paper, a list of all mathematical symbols and abbreviations appearing throughout the paper has been added at the end of the document.

## 2. Obtaining a PDF of a discrete random variable

For the sake of completeness, this section is addressed to explain how a probability mass function associated to a discrete random variable can be represented as a PDF. This is convenient to jointly handle both discrete and continuous random variables when computing the 1-PDF of the solution stochastic process,  $X(t)$ , of the IVP (1.2) via the RVT method. To apply this technique, the random vector  $\mathbf{U}$ , which will be defined as  $\mathbf{U} = (X_0, C, D, N)$ , must have a PDF. As mentioned in Section 1,  $X_0$  and  $(C, D)$  are absolutely continuous random variables, being  $f_{X_0}(x_0)$  and  $f_{C,D}(c, d)$  their respective PDFs. Therefore, a PDF must be associated to the random variable  $N$  to legitimate the application of Theorem 1. To do that, we will take advantage of Dirac delta distribution to express the probability mass function of  $N$  in a generalized form that permits identifying it as a PDF.

The Dirac delta distribution (or generalized function) is defined by [21],

$$\delta_{x_0}(x) = \delta(x - x_0) = \begin{cases} \infty, & x = x_0, \\ 0, & x \neq x_0. \end{cases}$$

This function has the following two properties that will be used in our subsequent development,

$$\int_{-\infty}^{\infty} \delta_{x_0}(x) dx = 1, \quad \int_a^b h(x) \delta_{x_0}(x) dx = \begin{cases} h(x_0), & a < x_0 < b, \\ 0, & \text{otherwise,} \end{cases} \quad -\infty \leq a < b \leq \infty. \quad (2.1)$$

Let  $N$  be a discrete random variable taking a countable number of values, say  $n_i$ , with associated probability  $p_i$ , i.e. having the following probability mass function

$$f_N(n) = \begin{cases} f_N(n_i), & i \in I, \\ 0, & \text{otherwise,} \end{cases} \quad \text{where} \quad f_N(n_i) = \mathbb{P}[\{\omega \in \Omega : N(\omega) = n_i\}] = p_i \in [0, 1],$$

being  $I$  a countable set, and  $p_i$  such that  $\sum_{i \in I} p_i = 1$ .

Then, its probability mass function can be formally represented as

$$f_N(n) = \sum_{i \in I} f_N(n_i) \delta_{n_i}(n) = \sum_{i \in I} p_i \delta_{n_i}(n), \quad -\infty \leq n \leq \infty, \quad (2.2)$$

since, using the first property of the Dirac delta function listed in (2.1),

$$\int_{-\infty}^{\infty} f_N(n) dn = \int_{-\infty}^{\infty} \sum_{i \in I} p_i \delta_{n_i}(n) dn = \sum_{i \in I} p_i \int_{-\infty}^{\infty} \delta_{n_i}(n) dn = \sum_{i \in I} p_i = 1.$$

Moreover, notice that  $f_N(n) \geq 0$ . Thereof, the probability mass function  $f_N(n)$  can be regarded as a PDF. Furthermore, using the second property of the Dirac delta function stated in (2.1) and the representation of  $f_N(n)$  given in (2.2), we can obtain the probability that  $N$  lies in a closed interval, say  $[n_l, n_u]$ ,

$$\begin{aligned} \mathbb{P}[\{\omega \in \Omega : n_l \leq N(\omega) \leq n_u\}] &= \int_{n_l}^{n_u} f_N(n) dn = \int_{n_l}^{n_u} \sum_{i \in I} f_N(n_i) \delta_{n_i}(n) dn \\ &= \sum_{i \in I} \int_{n_l}^{n_u} f_N(n_i) \delta_{n_i}(n) dn = \sum_{i \in I} f_N(n_i) = \sum_{\substack{i \in I \\ n_l \leq n_i \leq n_u}} p_i. \end{aligned}$$

We shall use the above result to associate a PDF to random variable  $N$ , then we will can apply the RVT method, stated in Theorem 1, to determine the PDF of the solution stochastic process of the random IVP (1.2).

### 3. Computing the first probability density function of the solution stochastic process of the random linear-quadratic logistic model

This section is addressed to compute the 1-PDF,  $f_{X(t)}(x)$ , of the solution stochastic process to random IVP (1.2) taking advantage of RVT method.

First, using classical integration methods, it is easy to check that the solution of equation (1.2) is given by

$$X(t) = -\frac{D - e^{(D+CN)t} N \frac{D+CX_0}{N-X_0}}{C + e^{(D+CN)t} \frac{D+CX_0}{N-X_0}}. \quad (3.1)$$

Next, we fix  $t > 0$ . Then, we apply Theorem 1 taking  $\mathbf{U} = (X_0, C, D, N)$  to obtain the PDF of the random vector  $\mathbf{V} = (V_1, V_2, V_3, V_4)$ , defined by the mapping  $\mathbf{r} : \mathbb{R}^4 \rightarrow \mathbb{R}^4$ , whose components  $r_1, r_2, r_3$  and  $r_4$  are defined by

$$\begin{aligned} v_1 = r_1(x_0, c, d, n) &= -\frac{d - ne^{dt+cnt} \left( \frac{d+cx_0}{n-x_0} \right)}{c + e^{dt+cnt} \left( \frac{d+cx_0}{n-x_0} \right)}, \\ v_2 = r_2(x_0, c, d, n) &= c, \\ v_3 = r_3(x_0, c, d, n) &= d, \\ v_4 = r_4(x_0, c, d, n) &= n. \end{aligned}$$

To keep the same notation as in Theorem 1, the inverse mapping of  $\mathbf{r}, \mathbf{s} : \mathbb{R}^4 \rightarrow \mathbb{R}^4, \mathbf{s} = \mathbf{r}^{-1} = (s_1, s_2, s_3, s_4)$ , is given by

$$\begin{aligned}
x_0 &= s_1(v_1, v_2, v_3, v_4) = \frac{v_3 v_4 - v_3 v_4 e^{(v_3+v_2 v_4)t} + v_3 v_1 e^{(v_3+v_2 v_4)t} + v_2 v_4 v_1}{v_3 + v_2 v_4 e^{(v_3+v_2 v_4)t} + v_2 v_1 - v_2 v_1 e^{(v_3+v_2 v_4)t}}, \\
c &= s_2(v_1, v_2, v_3, v_4) = v_2, \\
d &= s_3(v_1, v_2, v_3, v_4) = v_3, \\
n &= s_4(v_1, v_2, v_3, v_4) = v_4.
\end{aligned}$$

To construct the PDF of the random vector  $\mathbf{V}$ , via the RVT method, it is also required to obtain the absolute value of the Jacobian of  $\mathbf{s}$ ,

$$|\mathcal{J}| = \left| \frac{\partial s_1(v_1, v_2, v_3, v_4)}{\partial v_1} \right| = \frac{(v_3 + v_2 v_4)^2 e^{(v_3+v_2 v_4)t}}{(v_3 + v_2 (e^{(v_3+v_2 v_4)t}(v_4 - v_1)) + v_1)^2} \neq 0.$$

Observe that  $|\mathcal{J}| \neq 0$  with probability one (w.p. 1), since  $v_2, v_3$  and  $v_4$  represent arbitrary realizations of random variables,  $C = C(\omega)$ ,  $D = D(\omega)$  and  $N = N(\omega)$ , which are positive for every  $\omega \in \Omega$ . Furthermore the denominator defining the value of  $|\mathcal{J}|$  is different from zero, w.p. 1, since it is defined by absolutely continuous random variables.

Applying Theorem 1, the PDF of  $\mathbf{V}$  is given by

$$\begin{aligned}
&f_{V_1, V_2, V_3, V_4}(v_1, v_2, v_3, v_4) \\
&= f_{X_0, C, D, N} \left( \frac{v_3 v_4 - v_3 v_4 e^{(v_3+v_2 v_4)t} + v_3 v_1 e^{(v_3+v_2 v_4)t} + v_2 v_4 v_1}{v_3 + v_2 v_4 e^{(v_3+v_2 v_4)t} + v_2 v_1 - v_2 v_1 e^{(v_3+v_2 v_4)t}}, v_2, v_3, v_4 \right) \\
&\quad \cdot \frac{(v_3 + v_2 v_4)^2 e^{(v_3+v_2 v_4)t}}{(v_3 + v_2 (e^{(v_3+v_2 v_4)t}(v_4 - v_1)) + v_1)^2}.
\end{aligned}$$

As it has been indicated in Section 1,  $X_0$ ,  $(C, D)$  and  $N$  are assumed to be independent, thereof the PDF of the random vector  $\mathbf{U} = (X_0, C, D, N)$  can be expressed as  $f_{\mathbf{U}}(\mathbf{u}) = f_{X_0, C, D, N}(x_0, c, d, n) = f_{X_0}(x_0) f_{C, D}(c, d) f_N(n)$ . As a consequence,

$$\begin{aligned}
&f_{V_1, V_2, V_3, V_4}(v_1, v_2, v_3, v_4) \\
&= f_{X_0} \left( \frac{v_3 v_4 - v_3 v_4 e^{(v_3+v_2 v_4)t} + v_3 v_1 e^{(v_3+v_2 v_4)t} + v_2 v_4 v_1}{v_3 + v_2 v_4 e^{(v_3+v_2 v_4)t} + v_2 v_1 - v_2 v_1 e^{(v_3+v_2 v_4)t}} \right) f_{C, D}(v_2, v_3) f_N(v_4) \\
&\quad \cdot \frac{(v_3 + v_2 v_4)^2 e^{(v_3+v_2 v_4)t}}{(v_3 + v_2 (e^{(v_3+v_2 v_4)t}(v_4 - v_1)) + v_1)^2}.
\end{aligned}$$

So far, the PDF of the random vector  $\mathbf{V}$  has been computed. To determine the 1-PDF of the solution to random IVP (3.1), which is given by  $V_1$ , we now marginalize with respect to  $V_2 = C$ ,  $V_3 = D$  and

$V_4 = N$ ,

$$\begin{aligned}
 f_{X(t)}(x) &= \int_{\mathcal{D}(C,D)} \int_{\mathcal{D}(N)} f_{X_0} \left( \frac{dn - dne^{(d+cn)t} + dx e^{(d+cn)t} + cnx}{d + cne^{(d+cn)t} + cx - cxe^{(d+cn)t}} \right) f_{C,D}(c, d) f_N(n) \\
 &\quad \cdot \frac{(d + cn)^2 e^{(d+cn)t}}{(d + c(e^{(d+cn)t}(n - x)) + x)^2} dn dc dd \\
 &= \int_{\mathcal{D}(C,D)} \left[ \int_{\mathcal{D}(N)} f_{X_0} \left( \frac{dn - dne^{(d+cn)t} + dx e^{(d+cn)t} + cnx}{d + cne^{(d+cn)t} + cx - cxe^{(d+cn)t}} \right) f_{C,D}(c, d) \left( \sum_{i \in I} p_i \delta_{n_i}(n) \right) \right. \\
 &\quad \cdot \frac{(d + cn)^2 e^{(d+cn)t}}{(d + c(e^{(d+cn)t}(n - x)) + x)^2} dn \left. \right] dc dd \tag{3.2} \\
 &= \int_{\mathcal{D}(C,D)} \left[ \sum_{i \in I} p_i \int_{-\infty}^{\infty} f_{X_0} \left( \frac{dn - dne^{(d+cn)t} + dx e^{(d+cn)t} + cnx}{d + cne^{(d+cn)t} + cx - cxe^{(d+cn)t}} \right) f_{C,D}(c, d) \right. \\
 &\quad \cdot \frac{(d + cn)^2 e^{(d+cn)t}}{(d + c(e^{(d+cn)t}(n - x)) + x)^2} \delta_{n_i}(n) dn \left. \right] dc dd.
 \end{aligned}$$

Applying the second property in (2.1) with  $a = -\infty$  and  $b = \infty$ , the above expression can be simplified as

$$\begin{aligned}
 f_{X(t)}(x) &= \int_{\mathcal{D}(C,D)} \left[ \sum_{i \in I} p_i f_{X_0} \left( \frac{dn_i - dn_i e^{(d+cn_i)t} + dx e^{(d+cn_i)t} + cn_i x}{d + cn_i e^{(d+cn_i)t} + cx - cxe^{(d+cn_i)t}} \right) f_{C,D}(c, d) \right. \\
 &\quad \cdot \frac{(d + cn_i)^2 e^{(d+cn_i)t}}{(d + c(e^{(d+cn_i)t}(n_i - x)) + x)^2} \left. \right] dc dd \tag{3.3} \\
 &= \sum_{i \in I} p_i \int_{\mathcal{D}(C,D)} f_{X_0} \left( \frac{dn_i - dn_i e^{(d+cn_i)t} + dx e^{(d+cn_i)t} + cn_i x}{d + cn_i e^{(d+cn_i)t} + cx - cxe^{(d+cn_i)t}} \right) f_{C,D}(c, d) \\
 &\quad \cdot \frac{(d + cn_i)^2 e^{(d+cn_i)t}}{(d + c(e^{(d+cn_i)t}(n_i - x)) + x)^2} dc dd.
 \end{aligned}$$

It is interesting to observe that the last double integral can be expressed in terms of the expectation,  $\mathbb{E}_{C,D}[\cdot]$ , of random vector  $(C, D)$ ,

$$f_{X(t)}(x) = \sum_{i \in I} p_i \mathbb{E}_{C,D} \left[ f_{X_0} \left( \frac{Dn_i - Dn_i e^{(D+Cn_i)t} + Dxe^{(D+Cn_i)t} + Cn_i x}{D + Cn_i e^{(D+Cn_i)t} + Cx - Cxe^{(D+Cn_i)t}} \right) \frac{(D + Cn_i)^2 e^{(D+Cn_i)t}}{(D + C(e^{(D+Cn_i)t}(n_i - x)) + x)^2} \right]. \tag{3.4}$$

This expression is also useful to compute the 1-PDF of  $X(t)$  via simulations.

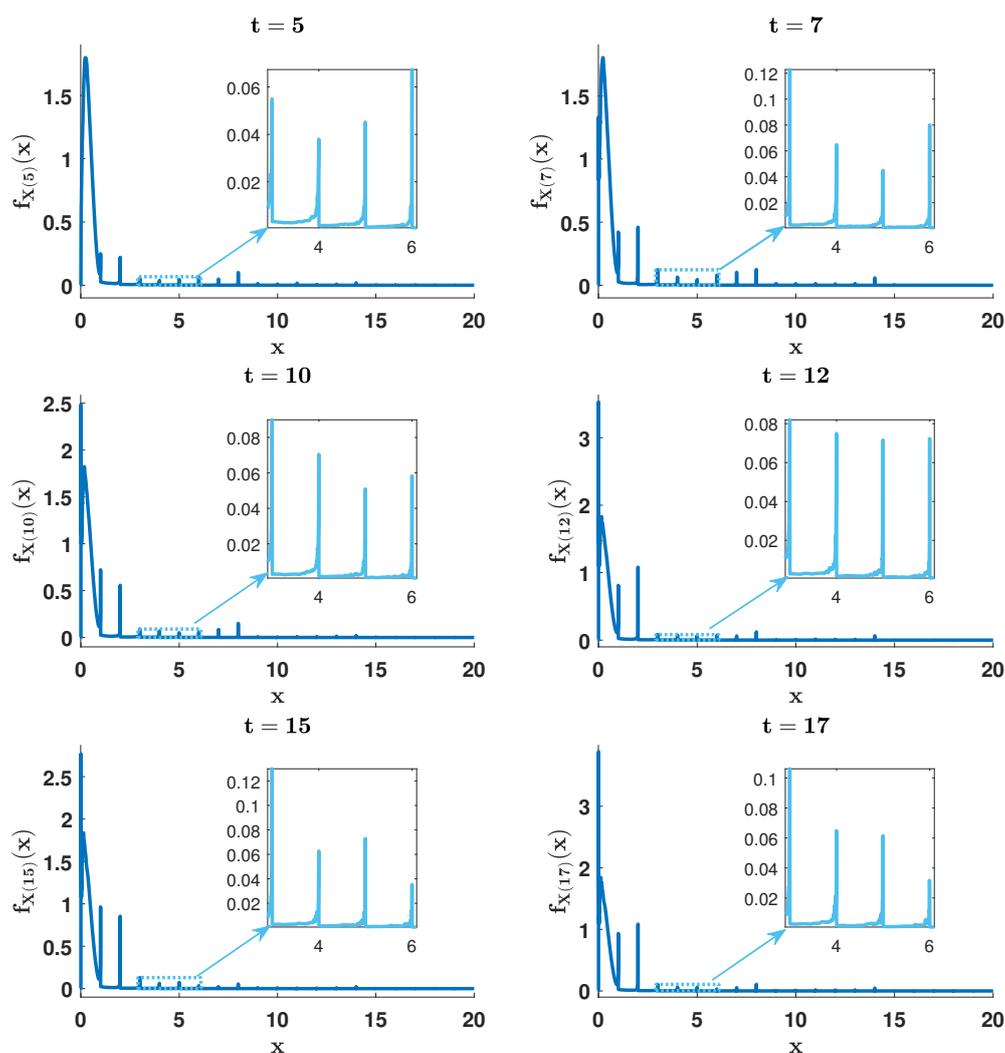
**Remark 1.** In the particular case that  $(C, D)$  are independent random variables, the factor  $f_{C,D}(c, d)$  appearing in formulas (3.2) and (3.3) writes  $f_C(c)f_D(d)$ .

Now, we show a numerical example where the previous theoretical findings are illustrated.

**Example 1.** Let us consider that  $X_0$  is a Beta distribution with parameters 2 and 3, i.e.  $X_0 \sim Be(2, 3)$ . Let  $N$  be a Beta Binomial distribution with parameters 20, 0.1 and 3, i.e.  $N \sim Bi(20, p)$ , being  $p \sim Be(0.1, 3)$ . The random vector  $(C, D)$  has been chosen as a Gaussian copula of dependent Gamma

distributions with parameters 0.2, 0.4 and 0.1, 0.3, respectively, and with correlation factor  $\rho = 0.5$ , i.e.  $(C, D) \sim (Ga(0.2, 0.4), Ga(0.1, 0.3), \rho = 0.5)$  [22].

Figure 1 shows the 1-PDF given in (3.3) at different time instants  $t \in \{5, 7, 10, 12, 15, 17\}$ . As a consequence of considering hybrid uncertainty ( $N$  is a discrete random variable), from these plots we can observe the peaks of the curves when  $x$  takes non-negative entire values.



**Figure 1.** 1-PDF,  $f_{X(t)}(x)$ , of solution stochastic process of random IVP (1.2) where  $X_0 \sim Be(2, 3)$ ,  $(C, D) \sim (Ga(0.2, 0.4), Ga(0.1, 0.3), \rho = 0.5)$  and  $N \sim Bi(20, Be(0.1, 3))$  at different time instants  $t \in \{5, 7, 10, 12, 15, 17\}$ . The 1-PDF has been computed by (3.3). Example 1.

#### 4. Computing the PDF of time until a given population size is reached

Once the 1-PDF of the solution stochastic process of IVP (3.1) has been determined, an important additional information is the PDF of the time,  $T$ , until a given population size, say  $x$ , has been reached.

To this end, we fix  $x$  and we consider the expression of the solution while  $T$  is regarded as a random variable,

$$x = -\frac{D - e^{(D+CN)T} N \frac{D+CX_0}{N-X_0}}{C + e^{(D+CN)T} \frac{D+CX_0}{N-X_0}}. \quad (4.1)$$

Solving for  $T$  yields

$$T = \frac{1}{D + CN} \ln \left( \frac{(xC + D)(N - X_0)}{(D + CX_0)(N - x)} \right).$$

We will determine the PDF of  $T$  applying the RVT method. Let  $\mathbf{r}$  be the transformation of  $\mathbf{U} = (X_0, C, D, N)$  into a random vector  $\mathbf{V}$ , defined as follows

$$\begin{aligned} v_1 &= r_1(x_0, c, d, n) = \frac{1}{d + cn} \ln \left( \frac{(xc + d)(n - x_0)}{(d + cx_0)(n - x)} \right), \\ v_2 &= r_2(x_0, c, d, n) = c, \\ v_3 &= r_3(x_0, c, d, n) = d, \\ v_4 &= r_4(x_0, c, d, n) = n. \end{aligned}$$

The inverse mappings  $\mathbf{s}$  of  $\mathbf{r}$ ,  $\mathbf{s} = \mathbf{r}^{-1}$ ,  $\mathbf{s} : \mathbb{R}^4 \rightarrow \mathbb{R}^4$  has the following components,  $s_i$ ,  $1 \leq i \leq 4$ ,

$$\begin{aligned} x_0 &= s_1(v_1, v_2, v_3, v_4) = \frac{v_3 v_4 - v_3 v_4 e^{(v_3+v_2v_4)v_1} + v_3 x e^{(v_3+v_2v_4)v_1} + v_2 v_4 x}{v_3 + v_2 v_4 e^{(v_3+v_2v_4)v_1} + v_2 x - v_2 x e^{(v_3+v_2v_4)v_1}}, \\ c &= s_2(v_1, v_2, v_3, v_4) = v_2, \\ d &= s_3(v_1, v_2, v_3, v_4) = v_3, \\ n &= s_4(v_1, v_2, v_3, v_4) = v_4. \end{aligned}$$

Now, we calculate the absolute value of the Jacobian of transformation  $\mathbf{s}$ ,

$$|\mathcal{J}| = \left| \frac{\partial s_1(v_1, v_2, v_3, v_4)}{\partial v_1} \right| = \left| -\frac{(v_3 + v_2v_4)^2 (v_4 - x)(v_3 + v_2x)e^{(v_3+v_2v_4)v_1}}{(v_3 + v_2(e^{(v_3+v_2v_4)v_1}(v_4 - x) + x))^2} \right| \neq 0, \quad \text{w.p. 1.}$$

Applying Theorem 1, we obtain the PDF of random vector  $\mathbf{V}$ ,

$$\begin{aligned} & f_{V_1, V_2, V_3, V_4}(v_1, v_2, v_3, v_4) \\ &= f_{X_0, C, D, N} \left( \frac{v_3 v_4 - v_3 v_4 e^{(v_3+v_2v_4)v_1} + v_3 x e^{(v_3+v_2v_4)v_1} + v_2 v_4 x}{v_3 + v_2 v_4 e^{(v_3+v_2v_4)v_1} + v_2 x - v_2 x e^{(v_3+v_2v_4)v_1}}, v_2, v_3, v_4 \right) \\ & \quad \cdot \left| \frac{(v_3 + v_2v_4)^2 (v_4 - x)(v_3 + v_2x)e^{(v_3+v_2v_4)v_1}}{(v_3 + v_2(e^{(v_3+v_2v_4)v_1}(v_4 - x) + x))^2} \right| \\ &= f_{X_0} \left( \frac{v_3 v_4 - v_3 v_4 e^{(v_3+v_2v_4)v_1} + v_3 x e^{(v_3+v_2v_4)v_1} + v_2 v_4 x}{v_3 + v_2 v_4 e^{(v_3+v_2v_4)v_1} + v_2 x - v_2 x e^{(v_3+v_2v_4)v_1}} \right) f_{C, D}(v_2, v_3) f_N(v_4) \\ & \quad \cdot \left| \frac{(v_3 + v_2v_4)^2 (v_4 - x)(v_3 + v_2x)e^{(v_3+v_2v_4)v_1}}{(v_3 + v_2(e^{(v_3+v_2v_4)v_1}(v_4 - x) + x))^2} \right|, \end{aligned}$$

where in the last step we have used that  $X_0$ ,  $(C, D)$  and  $N$  are independent.

Since  $T$  is the first component,  $V_1$ , of random vector  $\mathbf{V}$ , in order to compute the PDF of  $T$ , we marginalize with respect to  $V_2 = C$ ,  $V_3 = D$  and  $V_4 = N$  and we also use the representation (2.2) of the discrete random variable  $N$  via a PDF,

$$\begin{aligned}
 f_T(t, x) &= \int_{\mathcal{D}(C,D)} \int_{\mathcal{D}(N)} f_{X_0} \left( \frac{dn - dne^{(d+cn)t} + dx e^{(d+cn)t} + cnx}{d + cne^{(d+cn)t} + cx - cxe^{(d+cn)t}} \right) f_{C,D}(c, d) f_N(n) \\
 &\quad \cdot \left| \frac{(d+cn)^2 (n-x)(d+cx)e^{(d+cn)t}}{(d+c(e^{(d+cn)t}(n-x)+x))^2} \right| dn dc dd \\
 &= \int_{\mathcal{D}(C,D)} \int_{\mathcal{D}(N)} f_{X_0} \left( \frac{dn - dne^{(d+cn)t} + dx e^{(d+cn)t} + cnx}{d + cne^{(d+cn)t} + cx - cxe^{(d+cn)t}} \right) f_{C,D}(c, d) \left( \sum_{i \in I} p_i \delta_{n_i}(n) \right) \\
 &\quad \cdot \left| \frac{(d+cn)^2 (n-x)(d+cx)e^{(d+cn)t}}{(d+c(e^{(d+cn)t}(n-x)+x))^2} \right| dn dc dd \\
 &= \int_{\mathcal{D}(C,D)} \left[ \sum_{i \in I} p_i \int_{-\infty}^{\infty} f_{X_0} \left( \frac{dn - dne^{(d+cn)t} + dx e^{(d+cn)t} + cnx}{d + cne^{(d+cn)t} + cx - cxe^{(d+cn)t}} \right) f_{C,D}(c, d) \right. \\
 &\quad \cdot \left. \left| \frac{(d+cn)^2 (n-x)(d+cx)e^{(d+cn)t}}{(d+c(e^{(d+cn)t}(n-x)+x))^2} \right| \delta_{n_i}(n) \right] dn dc dd.
 \end{aligned} \tag{4.2}$$

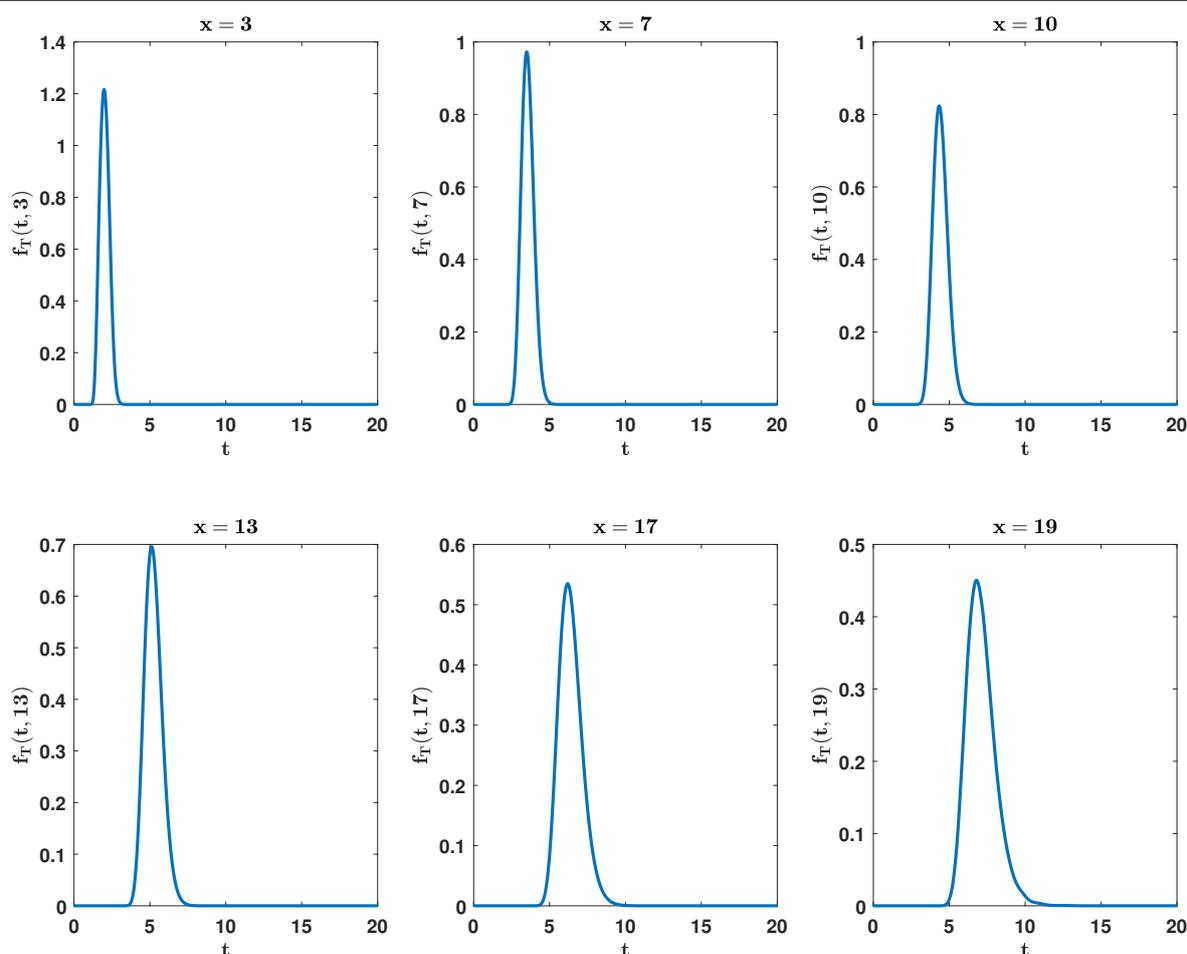
Now, we apply the second property in (2.1) with  $a = -\infty$  and  $b = \infty$ ,

$$\begin{aligned}
 f_T(t, x) &= \int_{\mathcal{D}(C,D)} \left[ \sum_{i \in I} p_i f_{X_0} \left( \frac{dn_i - dn_i e^{(d+cn_i)t} + dx e^{(d+cn_i)t} + cn_i x}{d + cn_i e^{(d+cn_i)t} + cx - cxe^{(d+cn_i)t}} \right) f_{C,D}(c, d) \right. \\
 &\quad \cdot \left. \left| \frac{(d+cn)^2 (n-x)(d+cx)e^{(d+cn)t}}{(d+c(e^{(d+cn)t}(n-x)+x))^2} \right| \right] dc dd \\
 &= \sum_{i \in I} p_i \mathbb{E}_{C,D} \left[ f_{X_0} \left( \frac{Dn_i - Dn_i e^{(D+Cn_i)t} + Dxe^{(D+Cn_i)t} + Cn_i x}{D + Cn_i e^{(D+Cn_i)t} + Cx - Cxe^{(D+Cn_i)t}} \right) \right. \\
 &\quad \cdot \left. \left| \frac{(D+CN)^2 (N-x)(D+Cx)e^{(D+CN)t}}{(D+C(e^{(D+CN)t}(N-x)+x))^2} \right| \right].
 \end{aligned} \tag{4.3}$$

Again, the Remark 1 applies to express (4.2) and (4.3) in terms of the marginal distributions  $f_C(c)$  and  $f_D(d)$  of  $C$  and  $D$ , respectively, when  $(C, D)$  are independent random variables.

Now, we shall illustrate the results by means of an example.

**Example 2.** Let us consider the random IVP (1.2), where the initial condition has a Beta distribution of parameters,  $(2, 3)$ , i.e.  $X_0 \sim Be(2, 3)$ . Let  $C$  and  $D$  be independent random variables with Uniform distributions in the interval  $[0.2, 0.25]$ , i.e.  $C, D \sim U([0.2, 0.25])$ , and  $N$  has a shifted binomial distribution with shift 20 and binomial parameters,  $(10, 0.5)$ , i.e.  $N \sim 20 + Bi(10, 0.5)$ . In Figure 2, we have plotted the PDF of  $T$  for different values of  $x$ . From these graphical representations, we can observe that the mean and the variance of  $T$  increase as  $x$  does.



**Figure 2.** PDF of solution stochastic process described in (4.3) considering  $X_0 \sim Be(2, 3)$ ,  $C, D \sim U([0.2, 0.25])$  and  $N \sim 20 + Bi(10, 0.5)$  at different values of  $x = 3, 7, 10, 13, 17, 19$ . Example 2.

## 5. Modelling Spotify users with uncertainty in the data

So far, we have illustrated the theoretical results by choosing arbitrary distributions to model parameters, however when dealing with real-world problems such distributions must be carefully determined so that the model *properly captures data uncertainty*. This is a critical point which requires inverse estimation techniques [23, Chapter 2]. In this section, we tackle this very interesting problem to studying the dynamics of Spotify’s users from the random IVP (1.2), assuming that  $X_0$ ,  $C$ ,  $D$  and  $N$  are independent random variables. We will use the yearly data about Spotify’s users (measured in million of individuals) during the period 2015–2019.

At this point, it is important to stress the meaning of the above sentence: “properly captures output uncertainty”. Indeed, most of the contributions focus on computing the model parameters by fitting the expectation of solution stochastic process to sampled data. To this end, some specific goodness-of-fit measure is minimized (typically, the mean square error). Then, confidence intervals are built to capture data uncertainty. Therefore, this approach is based on the punctual information contained in the the expectation to perform the fitting, and then one calculates the standard deviation to build confidence

intervals. In our example, we will perform the model fitting using a more complete information than the one based only on the two first moments, since we will determine the distribution of each model parameter. This will be done by minimizing a goodness-of-fit measure, that will be specified later, based on the 1-PDF of the solution stochastic process,  $f_{X(t)}(x)$ , given by (3.3) (or, equivalently, (3.4)). In this manner, as we directly work with the 1-PDF of the solution stochastic process, our model fitting is more complete since it takes into account more information of the solution than only the expectation.

To perform the aforementioned fitting via the 1-PDF,  $f_{X(t)}(x)$ , we will apply the methodology recently developed in [24]. This approach takes extensively advantage of the Principle of Maximum Entropy (PME) to assign reasonable distributions to sample data and model inputs.

Given a real random variable, say  $Y$ , the PME is an efficient method to assign it a probabilistic distribution using all the available statistical information of  $Y$ , such as its support, moments, etc. For instance, if  $Y$  is an absolutely continuous random variable with support,  $\mathcal{D}(Y)$ , and its two first moments,  $m_1 = \mathbb{E}[Y]$  and  $m_2 = \mathbb{E}[Y^2]$ , are known, then the PME looks for an approximation,  $f_Y$ , of the true PDF of  $Y$  that maximizes the Shannon's entropy

$$\mathcal{S}(f_Y) = - \int_{\mathcal{D}(Y)} f_Y(y) \ln(f_Y(y)) dy,$$

subject to the following constraints

$$\int_{\mathcal{D}(Y)} f_Y(y) dy = 1, \quad \int_{\mathcal{D}(Y)} y f_Y(y) dy = m_1, \quad \int_{\mathcal{D}(Y)} y^2 f_Y(y) dy = m_2. \quad (5.1)$$

As  $f_Y$  must represent a PDF, the first restriction indicates the normalization condition that guarantees it integrates the unit over the whole real line. Using the variational formulation of method of Lagrange multipliers, it can be seen that the PDF of  $Y$  has the following form [25]

$$f_Y(y) = \mathbb{1}_{\mathcal{D}(Y)} e^{-1-\lambda_0-\lambda_1 y-\lambda_2 y^2},$$

where  $\mathbb{1}_{\mathcal{D}(Y)}$  denotes the characteristic function on the domain  $\mathcal{D}(Y)$ , and parameters,  $\lambda_i$ ,  $i \in \{0, 1, 2\}$ , are calculated by solving the constraints given in (5.1).

Roughly speaking and using the interpretation of entropy, the PME method seeks for the PDF,  $f_Y$ , that corresponds to the maximal uncertainty (based on the Shannon's entropy measure) and the minimal quantity of information (based on the available statistical moments).

Now, we explain how the PME is applied to assign a PDF to sampled data for the number of Spotify's users. These data are collected in third column of Table 1 and considered as the mean of sampled distribution, [26]. For convenience, hereinafter, we will denote by  $X_t$  the random variable that represents the number of Spotify's users at the time instant  $t$ . In our setting,  $t$  is measured in months. Since the number of Spotify's users is known yearly during the period 2015–2019, the data is determined at  $t \in \mathcal{T} = \{0, 12, 24, 36, 48\}$  (see second column of Table 1). To construct the constraints required to apply the PME, we proceed as follows. On the one hand, we will assume that sampled data represent the mean or expectation,  $m_{1,t}$ , of random variable  $X_t$ . On the other hand, since sampled data are rounded-off to million units, we will assume they contain an intrinsic sampling error depending on the number of users in every time instant. Thereof, we will take as standard deviation 1% of each sampled data, i.e.  $\sigma_t = 0.01m_{1,t}$ . This allows us to compute the second-order moment,  $m_{2,t} = m_{1,t}^2 + \sigma_t^2$

(see last column in Table 1). Notice that, by means of this reasoning, we are implicitly considering that, at every time instant  $t \in \mathcal{T}$ , each sampled data comes from a random variable,  $X_t$ , whose mean ( $m_{1,t}$ ) and second moment ( $m_{2,t}$ ) are listed in Table 1.

Random Variable	Months	Mean ( $m_{1,t}$ )	2nd moment ( $m_{2,t}$ )
$X_0$	$t = 0$ (December 2015)	91	8.281e+03
$X_{12}$	$t = 12$ (December 2016)	123	1.513e+04
$X_{24}$	$t = 24$ (December 2017)	160	2.560e+04
$X_{36}$	$t = 36$ (December 2018)	207	4.285e+04
$X_{48}$	$t = 48$ (December 2019)	271	7.344e+04

**Table 1.** Number of Spotify's users, at different time instants  $t$  (in months), represented by random variable  $X_t$ , with the mean,  $m_{1,t}$ , and second-order momen,  $m_{2,t}$ , [26].

With this information, we apply the PME to assign a PDF,  $f_{X_t}$ , of each random variable  $X_t$  of the form

$$f_{X_t}(x) = \mathbb{1}_{\mathcal{D}(X_t)} e^{-1-\lambda_0^t - \lambda_1^t x - \lambda_2^t x^2}, \quad (5.2)$$

where  $\mathcal{D}(X_t)$  denotes its domain and, the parameters  $\lambda_0^t$ ,  $\lambda_1^t$  and  $\lambda_2^t$ , are calculated by numerically solving the following system of nonlinear equations

$$\int_{\mathcal{D}(X_t)} f_{X_t}(x) dx = 1, \quad \int_{\mathcal{D}(X_t)} x f_{X_t}(x) dx = m_{1,t}, \quad \int_{\mathcal{D}(X_t)} x^2 f_{X_t}(x) dx = m_{2,t}, \quad (5.3)$$

where  $m_{1,t}$  and  $m_{2,t}$  are given in Table 1. The results are shown in Table 2.

Months	$\lambda_0^t$	$\lambda_1^t$	$\lambda_2^t$
$t = 0$	1867.722	-41.045	0.226
$t = 12$	2175.312	-35.370	0.144
$t = 24$	1941.066	-24.269	0.076
$t = 36$	1849.090	-17.872	0.043
$t = 48$	1811.340	-13.376	0.025

**Table 2.** Values of  $\lambda_0^t$ ,  $\lambda_1^t$  and  $\lambda_2^t$  obtained after solving the system of nonlinear equations given in (5.3) for every  $t$ .

Once reliable PDFs have been assigned to sampled data, we proceed to assign suitable distributions for every model parameters so that the 1-PDF of the logistic model, given by (3.2), fits the PDFs of data detailed above at  $t \in \{0, 12, 24, 36, 48\}$ . To do that for,  $X_0$ ,  $C$  and  $D$ , we will apply the PME method again. First observe that, for consistency, the PDF of  $X_0$  is given by (see values of  $\lambda_0^{t=0}$ ,  $\lambda_1^{t=0}$  and  $\lambda_2^{t=0}$  in Table 2)

$$f_{X_0}(x_0) = e^{-1-1867.22+14.045x_0-0.226x_0^2}, \quad x_0 \in [x_{0,1}, x_{0,2}] = [0, 119.6182]. \quad (5.4)$$

To determine the endpoints,  $x_{0,1}$  and  $x_{0,2}$ , we have used the  $\sigma$ -rule [27, page 122]. Specifically, we have taken  $x_{0,1} = \max\{0, m_{1,t=0} - k\sigma_{t=0}\}$  (as  $X_0$  must be positive) and  $x_{0,2} = m_{1,t=0} + k\sigma_{t=0}$  with  $k = 30$ . This guarantees the interval  $[x_{0,1}, x_{0,2}]$  embraces more than 99.9% of the probability regardless the distribution of  $X_0$ . As  $m_{1,t=0} - 30\sigma_{t=0} < 0$  and  $m_{1,t=0} + 30\sigma_{t=0} = 119.6182$ , this results  $[x_{0,1}, x_{0,2}] = [0, 119.6182]$ .

The PDFs,  $f_C(c)$  and  $f_D(d)$ , of random variables  $C$  and  $D$ , respectively, are constructed via PME method taking into account that their corresponding domains are non-negative, since they represent growth-rates. Then, they have the following form

$$f_C(c) = e^{-1-\lambda_0^C-\lambda_1^C c-\lambda_2^C c^2}, \quad c \in [0, c_2], \quad (5.5)$$

and

$$f_D(d) = e^{-1-\lambda_0^D-\lambda_1^D d-\lambda_2^D d^2}, \quad d \in [0, d_2]. \quad (5.6)$$

Finally, the probability mass function of discrete random variable,  $N$ , is modelled by a shifted Binomial distribution, with parameters  $n_1, n_2$  and  $\nu_N$ , i.e.,  $N \sim n_1 + \text{Bi}(n_2, \nu_N)$ , where

$$p_i = \mathbb{P}[\{\omega \in \Omega : N(\omega) = n_1 + i\}] = \frac{n_2!}{i!(n_2 - i)!} (\nu_N)^i (1 - \nu_N)^{n_2 - i}, \quad i \in I = \{0, 1, \dots, n_2\}. \quad (5.7)$$

At this point, the parameters  $\lambda_0^C, \lambda_1^C, \lambda_2^C, c_2, \lambda_0^D, \lambda_1^D, \lambda_2^D, d_2, n_1, n_2$  and  $\nu_N$  need to be calculated. This will be done by fitting the 1-PDF,  $f_{X(t)}(x)$ , calculated in (3.3) (or equivalently (3.4)), to the PDF,  $f_{X_t}(x)$ , given by (5.2), at the time instants  $t \in \mathcal{T} = \{0, 12, 24, 36, 48\}$ . Notice that  $f_{X(t)}(x)$  depends on  $f_{X_0}(x_0)$ ,  $f_C(c)$ ,  $f_D(d)$  and  $p_i$ , where  $f_{X_0}(x_0)$  is completely determined in (5.4);  $f_C(c)$  depends on  $\lambda_0^C, \lambda_1^C, \lambda_2^C$  and  $c_2$  (see (5.5));  $f_D(d)$  depends on  $\lambda_0^D, \lambda_1^D, \lambda_2^D$  and  $d_2$  (see (5.6)), and  $p_i$  depends on  $n_1, n_2$  and  $\nu_N$  (see (5.7)). Similarly as in the algorithm recently described in Steps 7 and 8 in [24, Sec. 5], we take as goodness-of-fit function

$$E = \sum_{t \in \mathcal{T}} E_t, \quad E_t = \sum_{i=0}^N \frac{|f_{X(t)}(x_i) - f_{X_t}(x_i)|}{f_{X_t}(x_i)}, \quad \mathcal{T} = \{0, 12, 24, 36, 48\}, \quad (5.8)$$

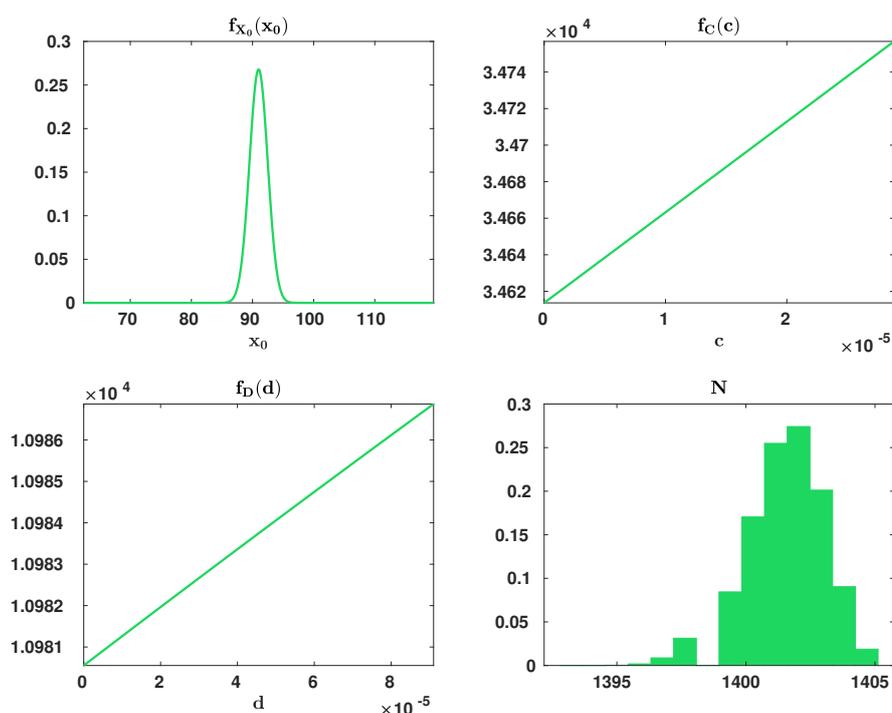
where  $x_i$  is a mesh on a domain determined by the PDF,  $f_{X_t}(x)$ , of sampled data (see further details in Step 3 in [24, Sec. 5]). Notice that  $E_0 = 0$  since by construction  $f_{X(0)} = f_{X_0}$ .

We have used Particle Swarm Optimization [28], to find the values of  $\lambda_0^C, \lambda_1^C, \lambda_2^C, c_2, \lambda_0^D, \lambda_1^D, \lambda_2^D, d_2, n_1, n_2$  and  $\nu_N$  that minimize the function  $E$  functional defined in (5.8). In Table 3, we show the obtained values. To carry out these computations, the 1-PDF  $f_{X(t)}(x)$  has been evaluated via the expectation given in (3.4). To this end, we firstly fix  $t \in \mathcal{T}$ . Secondly, we obtain a simulation of each random variable  $C$  and  $D$  (recall that, according to (5.5) and (5.6)), we know their respective PDF. Thirdly, we evaluate the argument of the expectation. This process is repeated many times and the results are averaged. This permits approximating  $f_{X(t)}(x)$ . More details of this computational procedure can be seen in [24, Sec. 5].

From figures listed in Table 3, and using expressions (5.5), (5.6) and (5.7), we have calculated the distributions  $f_C(c)$ ,  $f_D(d)$  and  $f_N(n)$ , respectively. In Figure 3, we show the graphical representations of these distributions and of  $f_{X_0}$  from expression (5.4).

Parameters	Values
$\lambda_0^C$	-11.4520
$\lambda_1^C$	-136.6524
$\lambda_2^C$	151.444
$c_2$	$7.7305e - 05$
$\lambda_0^D$	-10.3465
$\lambda_1^D$	-105.6717
$\lambda_2^D$	$1.1367e + 03$
$d_2$	$8.6869e - 05$
$n_1$	1353
$n_2$	15
$\nu_N$	0.8818

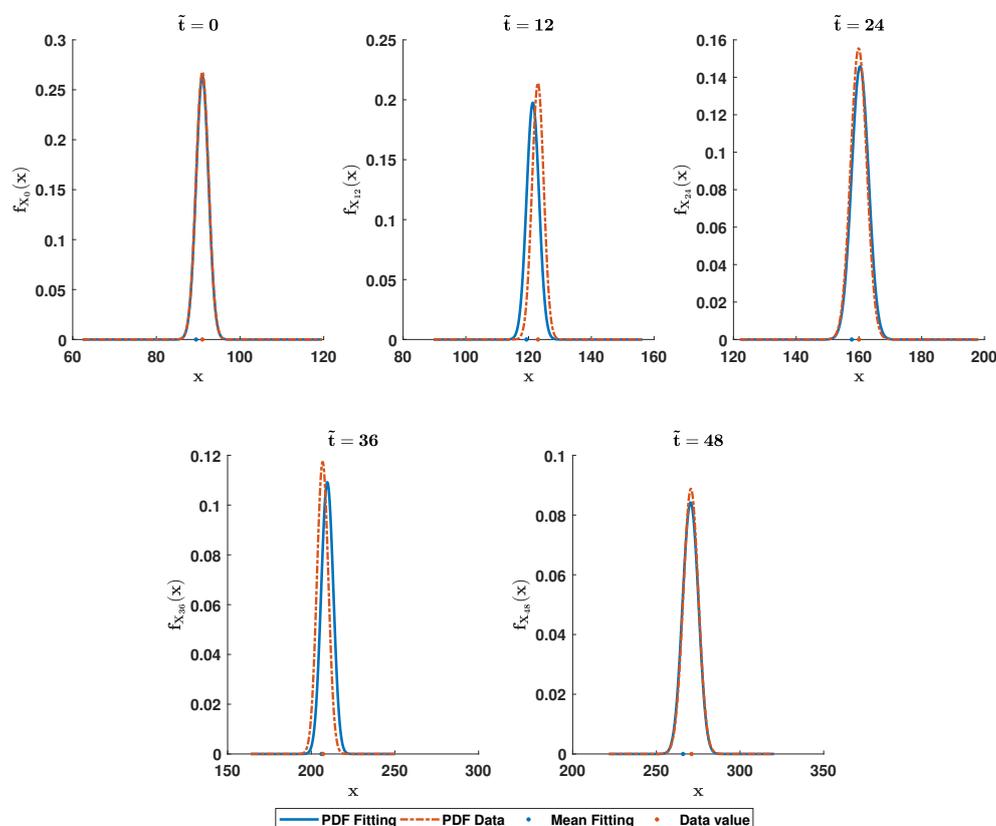
**Table 3.** Values of parameters  $\lambda_0^C$ ,  $\lambda_1^C$ ,  $\lambda_2^C$ ,  $c_2$ ,  $\lambda_0^D$ ,  $\lambda_1^D$ ,  $\lambda_2^D$ ,  $d_2$ ,  $n_1$ ,  $n_2$  and  $\nu_N$  that define the distributions for random variables  $C$ ,  $D$  and  $N$ .



**Figure 3.** PDFs assigned to random model parameters  $X_0$ ,  $C$  and  $D$  and probability mass function for  $N$  according to the obtained values of Table 3.

To better compare the obtained results, in Figure 4 we plot the PDFs associated to sampled data (red lines) and the PDFs of the solution stochastic process, given by (3.3), at the time instants  $t \in \mathcal{T} = \{0, 12, 24, 3, 6, 48\}$ . Red and blue points, plotted on the horizontal axis represent, respectively, the sampled data of Spotify's users listed in the second column ( $m_{1,t}$ ) of Table 1 and the expectation of the solution stochastic process calculated by means of expression (1.1) with  $m = 1$  where  $f_{X(t)}(x)$  is given

by (3.3). From these plots we can see the fitting via the PDFs is very good.



**Figure 4.** Comparison between the PDF assigned to each sampled data (red dashed lines) via the PME according to (5.2) and the PDF (blue lines), given by (3.3) and constructed via the RVT method. This comparison is shown at the time instants  $t \in \mathcal{T} = \{0, 12, 24, 26, 48\}$  (in months). Red and blue points, plotted on the horizontal axis, represent, respectively, the sampled data of Spotify's users and the expectation of the solution stochastic process.

## 6. Conclusions

In this paper we have performed a probabilistic study of a generalized version of the random logistic differential equation which has a linear term in its formulation. This additional term describes the autonomous behaviour that affects the model dynamics. For the sake of generality, in our analysis we have assumed that all model parameters and the initial condition are random variables having arbitrary distributions. The randomness is treated from a hybrid standpoint in the sense we consider a combination of discrete and continuous random variables for model inputs. The key tools to conduct our study are, on the one hand, the Dirac delta function to unify the treatment of discrete and continuous random variables with their respective probability density functions and, on the other hand, the Random Variable Transformation method that permits computing the density of certain differentiable mappings that act onto absolutely continuous random vectors. We have included a number of examples that illustrate different cases where our theoretical findings can be applied

including the application of a model using real-world data. Nevertheless, we must point out that our approach relies on two critical points, first the availability of a closed-form expression of the solution stochastic process of the general linear-quadratic differential equation and, secondly, the exact computation of the inverse of the mapping defined when applying the Random Variable Transformation method. In other models, where some of the above exact expressions are not available, the method can still be applied using numerical approximations, which certainly will increase the computational burden. Finally, we think that approach proposed in this paper may open new avenues in the analysis of problems formulated by means of differential equations with hybrid (discrete-continuous) randomness.

## Acknowledgments

This work has been supported by the Spanish Ministerio de Economía, Industria y Competitividad (MINECO), the Agencia Estatal de Investigación (AEI) and Fondo Europeo de Desarrollo Regional (FEDER UE) grant MTM2017-89664-P. Computations have been carried thanks to the collaboration of Raúl San Julián Garcés and Elena López Navarro granted by European Union through the Operational Program of the European Regional Development Fund (ERDF)/European Social Fund (ESF) of the Valencian Community 2014–2020, grants GJIDI/2018/A/009 and GJIDI/2018/A/010, respectively.

## Conflict of interest

The authors declare that there are not conflict of interest.

## Nomenclature

$(\Omega, \mathcal{F}, \mathbb{P})$	complete probability space.
$\mathbb{E}[X]$	mean of the random variable $X$ .
$\mathbb{V}[X]$	variance of the random variable $X$ .
$f_X(x)$	probability density function of the random variable $X$ .
$\delta_{x_0}(x)$	Dirac delta function at value $x_0$ .
$\mathbb{1}_{[a,b]}$	characteristic function on the interval $[a, b]$ .
fidis	finite distributions.
IVP	initial value problem.
PDF	probability density function.
1-PDF	first probability density function.
PME	Principle of Maximum Entropy.
RDE	random differential equation.
RVT	Random Variable Transformation.
SDE	stochastic differential equation.
w.p. 1	with probability one.

---

**References**

1. R. Goebel, R. G. Sanfelice, A. R. Teel, *Hybrid Dynamical Systems: Modeling, Stability, and Robustness*, Princeton University Press, 2012.
2. J. Lunze, F. Lamnabhi-Lagarrigue, *Handbook of Hybrid Systems Control: Theory, Tools, Applications*, ser. Nonlinear Systems and Complexity, Cambridge University Press, 2009.
3. E. Villani, P. E. Miyagi, R. Valette, *Modelling and Analysis of Hybrid Supervisory Systems: A Petri Net Approach*, ser. Advances in Industrial Control, Springer, 2007.
4. X. Yang, X. Li, J. Lu, Z. Cheng, Synchronization of time-delayed complex networks with switching topology via hybrid actuator fault and impulsive effects control, *IEEE Trans. Cybernetics*, **50** (2019), 4043–4052.
5. H. Zhu, X. Li, J. Lu, Z. Cheng, Input-to-state stability of impulsive systems with hybrid delayed impulse effects, *J. Appl. Anal. Comput.*, **50** (2019), 777–795.
6. X. Liu, P. Stechliniski, *Infectious Disease Modeling: A Hybrid System Approach*, ser. Nonlinear Systems and Complexity, Springer International Publishing, **19** (2017).
7. R. C. Smith, *Uncertainly Quantification: Theory, Implementation, and Applications*, ser. Computational Science & Engineering, SIAM, 2013.
8. E. Allen, *Modeling with Itô Stochastic Differential Equations*, Springer Science & Business Media, **22** (2007).
9. T. T. Soong, *Random Differential Equations in Science and Engineering*, New York: Academic Press, 1973.
10. F. B. Hanson, *Applied Stochastic Processes and Control for Jump-Diffusions*, SIAM, 2007.
11. J. Bertoin, *Lévy processes*, Cambridge University Press, 1996.
12. M. Grigoriu, *Applied Non-Gaussian Processes*, Prentice-Hall, 1995.
13. M. Grigoriu, Numerical solution of stochastic differential equations with Poisson and Lévy white noise, *Physical Rev. E*, **80** (2009), 02670.
14. M. C. Casabán, J. C. Cortés, A. Navarro-Quiles, J. V. Romero, M. D. Roselló, R. J. Villanueva, A comprehensive probabilistic solution of random SIS-type epidemiological models using the random variable transformation technique, *Commun. Nonlinear Sci. Numer. Simul.*, **32** (2016), 199–210. Available from: <https://doi.org/10.1016/j.cnsns.2015.08.009>.
15. F. A. Dorini, M. S. Ceconello, L. B. Dorini, On the logistic equation subject to uncertainties in the environmental carrying capacity and initial population density, *Commun. Nonlinear Sci. Numer. Simul.*, **33** (2016), 160–173. Available from: <https://doi.org/10.1016/j.cnsns.2015.09.009>.
16. F. A. Dorini, M. C. C. Cunha, Statistical moments of the random linear transport equation, *J. Comput. Phys.*, **227** (2008), 8541–8550. Available from: <https://doi.org/10.1016/j.jcp.2008.06.002>.
17. M. Hussein, M. Selim, A complete probabilistic solution for a stochastic milne problem of radiative transfer using KLE-RVT technique, *J. Quantit. Spectrosc. Radiat. Transfer*, **232** (2019), 54–65.

18. J. C. Cortés, I. C. Lombana, R. J. Villanueva, Age-structured mathematical modeling approach to short-term diffusion of electronic commerce in Spain, *Math. Comput. Model.*, **52** (2010), 1045–1051, *Mathematical Models in Medicine, Business and Engineering 2009*. Available from: <http://www.sciencedirect.com/science/article/pii/S0895717710000944>.
19. R. Cervelló-Royo, J. C. Cortés, A. Sánchez-Sánchez, F. J. Santonja, R. Shoucri, R. J. Villanueva, Probabilistic European country risk score forecasting using a diffusion model, In: *Computational models of complex systems*, Springer, (2014), 45–58.
20. J. Calatayud, J. C. Cortés, F. A. Dorini, M. Jornet, Solution of the finite Milne problem in stochastic media with RVT Technique, *Comput. Appl. Math.*, **39** (2020), 288.
21. R. F. Hoskins, *Delta functions: Introduction to generalised functions*, Horwood Publishing, 2009.
22. Matlab, Cupula Matlab, 2020, Available from: <https://es.mathworks.com/help/stats/copularnd.html>.
23. L. Devroye, Nonuniform random variate generation, *Handbooks Oper. Res. Manage. Sci.*, **13** (2006), 83–121.
24. C. Burgos, J. C. Cortés, D. Martínez-Rodríguez, R. J. Villanueva, Modeling breast tumor growth by a randomized logistic model: A computational approach to treat uncertainties via probability densities, *Europ. Phys. J. Plus*, **135** (2020). Available from: <https://doi.org/10.1140/epjp/s13360-020-00853-3>.
25. J. V. Michalowicz, J. M. Nichols, F. Bucholtz, *Handbook Differ. Entropy*, CRC Press, 2013.
26. Statista, Number of Spotify monthly active users (MAU) worldwide from 1st quarter 2015 to 3rd quarter 2020. Available from: <https://www.statista.com/statistics/367739/spotify-global-mau/>.
27. G. Casella, R. Berger, *Statistical Inference*, ser. Duxbury Advanced Series, New York: Brooks Cole, 2002.
28. The MathWorks Inc. (2020), Particle swarm optimization. Available from: <https://es.mathworks.com/help/gads/particleswarm.html>.



AIMS Press

©2021 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)