The final publication is available at

https://doi.org/10.1016/j.cnsns.2015.08.009

Additional Information

A comprehensive probabilistic solution of random SIS-type epidemiological models using the Random Variable Transformation technique

M.-C. Casabán[a], J.-C. Cortés[a,*], A. Navarro-Quiles[a], J.-V. Romero[a], M.-D. Roselló[a], R.-J. Villanueva[a]

[a]*Instituto Universitario de Matemática Multidisciplinar,*
Universitat Politècnica de València,
Camino de Vera s/n, 46022, Valencia, Spain

Abstract

This paper provides a complete probabilistic description of SIS-type epidemiological models where all the input parameters (contagion rate, recovery rate and initial conditions) are assumed to be random variables. By applying the Random Variable Transformation technique, the first probability density function, the mean and the variance functions, as well as confidence intervals associated to the solution of SIS-type epidemiological models, are determined. It is done under the general hypothesis that model random inputs have any joint probability density function. The distributions to describe the time until a given proportion of the population remains susceptible and infected are also determined. Lastly, a probabilistic description of the so-called basic reproductive number is included. The theoretical results are applied to an illustrative example showing good fitting.

Keywords: Random SIS-type epidemiological models, first-order nonlinear random differential equations, Random Variable Transformation technique, first probability density function

1. Motivation and Preliminaries

This paper is a continuation of the recent contribution [1] by some of the authors. In [1], a comprehensive probabilistic description of the solution to random SI-type epidemiological models is provided. The study is based on the Random Variable Transformation (RVT) technique. SI-type epidemiological models are appropriate to describe the dynamics from susceptible (S) to Infected (I) for some stages of epidemics. For example, SI model has been used to describe the early stages of the AIDS disease and epidemics with a long incubation period. For these kind of diseases, births and deaths or other transitions have no significant effect on the epidemics [2]. They are also useful to model the spread of new technologies, where potential customers are

---

identified with the *susceptible* subpopulation and, the remainder of the individuals who have already adopted the new technology, corresponds to the *infected* subpopulation [3, 4, 1]. However, SI-models are not appropriate for most epidemics which involve transitions from infected to susceptible. As it shall be justified below, this has motivated the consideration of the so-called SIS-type epidemiological models [5]. The aim of this paper is to extend the results obtained in [1] but now for the SIS-type epidemiological models taking again advantage of the RVT method. This extension entails, among other aspects, the extension to the random framework of the so-called reproductive number which does not appear in dealing with SI-type epidemiological models. Our approach is based on RVT technique, and its consideration is justified because a full probabilistic description of the solution stochastic process for random SIS-type epidemiological models can be obtained under very general assumptions. Indeed, as it shall be seen later in the theoretical development, statistical dependence of all the input parameters involved into the SIS-model via any plausible joint probability density function will be considered. As it will be pointed out later, this is a significant advantage that other available approaches, such as SIS-type epidemiological models based on Itô stochastic differential equations, do not possess.

SIS-type epidemiological models constitute mathematical representations to describe the spread by individual-to-individual contact of infectious diseases [6, 7]. SIS models are useful for modelling diseases whose infection does not confer immunity, that is, any susceptible (S) who has been infected (I) can recover from the disease and then becoming susceptible (S) again. This is usually represented as $S \rightarrow I \rightarrow S$ (Susceptible-Infected-Susceptible). Examples of diseases that have been modelled by SIS models include gonorrhea, meningitis, streptococcal sore throat, etc. [8, 9, 5]. SIS epidemiological models have also been used to model the dynamics of unhealthy lifestyle habits such as the excess of weight [10]. In spite of their mathematical simplicity, SIS models constitute the basis of more refined and sophisticated models where, for instance, mode of transmission, resistance, environmental and cultural factors, and further diseases characteristics can be considered.

SIS models are formulated as initial value problems based on nonlinear systems of differential equations of the form

$$\begin{cases} S'(t) & = & -\beta S(t)I(t) + \gamma I(t), \\ I'(t) & = & \beta S(t)I(t) - \gamma I(t), \end{cases} \qquad t > 0, \tag{1}$$

with initial conditions

$$S(0) = S_0, \quad I(0) = I_0, \tag{2}$$

where, $S(t)$ and $I(t)$ denote the proportion, or equivalently, the percentage of susceptibles and infected at the time instant t, respectively. At the initial time instant, $t = 0$, these values correspond to S_0 and I_0, respectively. It is assumed that

$$S(t) + I(t) = 1, \quad \forall t \geq 0, \tag{3}$$

which means that all the individuals of the population are classified as either susceptible or infected. The parameters $\beta > 0$ and $\gamma > 0$ denote the rate of decline in the percentage of susceptibles and the rate of infected that recover from the disease, respectively. In Figure 1, a flow diagram of the SIS model is depicted.

The consideration of births and deaths, usually referred to as *vital dynamics*, could be included in the model (1)–(3). However, the behaviour of solutions of the new corresponding model is similar when vital dynamics are omitted in deterministic models [5]. This motivates

the analysis of the simpler model (1)–(3) as a first step. The inclusion of vital dynamics using random variables to model births and deaths is a very interesting problem that requires its own subsequent development, although it would probably benefit from the ideas presented here.
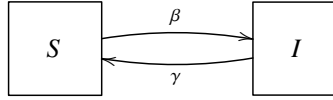

Figure 1: Flow diagram representation of the SIS model (1)–(2).

From a practical point of view, it is more realistic to regard that the input parameters S_0, γ and β are random variables (r.v.'s) rather than deterministic values due to the inherent complexity and uncertainty that involve their determination and interpretation. The initial condition S_0 represents the percentage of susceptibles at the beginning. In practice, this value is usually set after sampling the target population, hence it has got variability. Since S_0 represents a percentage, it lies between 0 and 1, and it could be better modelled by a r.v. like a uniform r.v. on the interval $(0, 1)$, or a beta r.v., which is more flexible and generalizes $(0, 1)$-uniform r.v. The positive parameters γ and β embed complex factors related to the recovery rate from the disease and the contagion rate, respectively. Notice that, in practice, they are both not known with certainty. This motivates that positive probability distributions, such as exponential, gamma, or even Gaussian truncated on positive domain, might be adequate candidates for modelling input parameters γ and β. As it will be seen later, these kind of probability distributions will be assumed in the illustrative example shown in this paper.

The consideration of randomness in SIS-type epidemiological models has been undertaken mainly using Itô-type stochastic differential equations [11, 12]. Under this approach, randomness is introduced by means of a stationary and Gaussian stochastic process (s.p.), usually called white noise, which is the derivative, in a generalized sense, of Brownian motion (or Wiener process). Brownian motion belongs to an important class of s.p.'s termed semimartingales. The resulting stochastic differential equations are handled using a special calculus for semimartingales, usually referred to as Itô stochastic calculus, whose cornerstone result is the Itô Lemma [13, 14]. Alternately to this approach, and based on the previous motivation, uncertainty can be considered directly by assuming that the model input parameters are r.v.'s. Hence, a wide range of probability distributions can be assigned to them such as uniform, beta, exponential, gamma, Gaussian, etc. rather than limiting the introduction of randomness to Gaussian patterns. Therefore, under this approach the transition from the deterministic ordinary and partial differential equations to their random formulation, usually referred to as random differential equations (r.d.e.'s), appears in a natural way. The rigorous treatment of r.d.e.'s requires the application of Mean Square Calculus [15, 16]. Solving a r.d.e. means to compute its solution s.p. including its mean and variance functions, which provide valuable information about its statistical behaviour. The determination of its first probability density function (1-p.d.f.) is more recommendable since from it one gets a full description of the solution in each time instant. In particular, from the 1-p.d.f. one can compute not only the mean and the variance but also any unidimensional statistical moment associated to the solution s.p. In addition, the 1-p.d.f. permits to calculate the probability that the solution lies within a certain set of interest.

As it has been underlined previously, the key result that will be applied in this paper in order to determine the 1-p.d.f. of the solution of the SIS model (1)–(3) is the RVT method.

This technique will also be applied to calculate the p.d.f. of time until a given proportion of the population remains susceptible/infected and, an important epidemiological quantity called the basic reproductive number, usually denoted by R_0. The RVT method constitutes a powerful tool to determine the 1-p.d.f. of a s.p. [17]. In the context of r.d.e.'s, this technique has been successfully applied to compute the 1-p.d.f. of the solution of ordinary and partial r.d.e.'s [18, 19, 20, 21, 22, 23, 24]. Regarding epidemiological models, in [2] authors apply RVT method to investigate the effect of the random initial condition on a deterministic SI-type model, where both subpopulations are given in percentages. In this interesting paper, the study is limited to the case where the initial condition has a beta distribution, hence representing a percentage. Based on the approach initiated in [2], recently a full probabilistic analysis of SI-type models has been done by some of the authors in [1].

As we have previously indicated, Itô-type stochastic differential equations constitute an alternative to random differential equations in order to deal with modelling with uncertainty. In regard to the main objective of this paper, which is the determination of the 1-p.d.f. of the solution s.p. to SIS model (1)–(3) under general hypotheses, it is important to point out that following the Itô approach it is also possible to determine the p.d.f. of the solution by solving the so-called forward Fokker-Planck-Kolmogorov equation. This is a deterministic partial differential equation satisfied by the transition p.d.f. of a diffusion stochastic process [25, 26]. The key fact that permits to take advantage of this result is that solutions of Itô-type stochastic differential equations are Markov diffusion processes [15, Th. 5.2.5]. This is a mathematical requirement which does not need to be satisfied under the approach proposed in this paper.

Below, we state the RVT technique in its general version as well as several particular cases that will be required later. Throughout our exposition all the random variables/vectors are assumed to be absolutely continuous in order to guarantee the existence of their associated p.d.f.'s.

Theorem 1. *(Multidimensional version, [15, pp. 24–25]). Let $\mathbf{U} = (U_1, \ldots, U_n)^\mathsf{T}$ and $\mathbf{V} = (V_1, \ldots, V_n)^\mathsf{T}$ be two n-dimensional absolutely continuous random vectors. Let $\mathbf{g} : \mathbb{R}^n \to \mathbb{R}^n$ be a one-to-one deterministic transformation of $\mathbf{U}$ into $\mathbf{V}$, i.e., $\mathbf{V} = \mathbf{g}(\mathbf{U})$. Assume that $\mathbf{g}$ is continuous in $\mathbf{U}$ and has continuous partial derivatives with respect to $\mathbf{U}$. Then, if $f_\mathbf{U}(\mathbf{u})$ denotes the joint probability density function of vector $\mathbf{U}$, and $\mathbf{h} = \mathbf{g}^{-1} = (h_1(v_1, \ldots, v_n), \ldots, h_n(v_1, \ldots, v_n))^\mathsf{T}$ represents the inverse mapping of $\mathbf{g} = (g_1(u_1, \ldots, u_n), \ldots, g_n(u_1, \ldots, u_n))^\mathsf{T}$, the joint probability density function of vector $\mathbf{V}$ is given by*

$$f_\mathbf{V}(\mathbf{v}) = f_\mathbf{U}(\mathbf{h}(\mathbf{v})) |J|, \tag{4}$$

where $|J|$ is the absolute value of the Jacobian defined by the determinant

$$J = \det\left(\frac{\partial \mathbf{u}^\mathsf{T}}{\partial \mathbf{v}}\right) = \det\begin{pmatrix} \frac{\partial h_1(v_1,\ldots,v_n)}{\partial v_1} & \cdots & \frac{\partial h_n(v_1,\ldots,v_n)}{\partial v_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial h_1(v_1,\ldots,v_n)}{\partial v_n} & \cdots & \frac{\partial h_n(v_1,\ldots,v_n)}{\partial v_n} \end{pmatrix}. \tag{5}$$

In the particular case that $n = 1$ and $\mathbf{g}$ is linear, Theorem 1 reads

Proposition 1. *(Scalar linear transformation, [18, Proposition 2]). Let U be an absolutely continuous random variable defined on the domain $D_U = \{u : u_1 \leq u \leq u_2\}$ and with probability density function $f_U(u)$. Let us consider the random variable $V = aU + b$, $a \neq 0$. Then, the probability density function of V is given by*

$$f_V(v) = \frac{1}{|a|} f_U\left(\frac{v - b}{a}\right), \quad \begin{cases} v_1 = au_1 + b \leq v \leq au_2 + b = v_2 & \text{if} \quad a > 0, \\ v_1 = au_2 + b \leq v \leq au_1 + b = v_2 & \text{if} \quad a < 0. \end{cases}$$

4

If a = 0, then V = b with probability 1 and

$$f_V(v) = \delta(v - b), \quad -\infty < v < +\infty,$$

being $\delta(\cdot)$ the Dirac delta function.

Let $\mathbf{U} = (U_1, U_2)^{\mathsf{T}}$ be an absolutely continuous random vector with joint p.d.f. $f_{\mathbf{U}}(u_1, u_2)$ such that $\mathbb{P}[\{\omega \in \Omega : U_1(\omega) \neq 0\}] = 1$. Let us consider the r.v. $V_1 = U_2/U_1$. Then, introducing the auxiliary r.v. $V_2 = U_1$ and denoting by $\mathbf{V} = (V_1, V_2)^{\mathsf{T}}$, one can define the mapping

$$\mathbf{V} = \mathbf{g}(\mathbf{U}), \quad \mathbf{g} : \mathbb{R}^2 \to \mathbb{R}^2, \quad \mathbf{g}(U_1, U_2) = (g_1(U_1, U_2), g_2(U_1, U_2))^{\mathsf{T}} = (V_1, V_2)^{\mathsf{T}},$$

whose inverse transformation, $\mathbf{h} = \mathbf{g}^{-1}$, is defined as

$$\mathbf{U} = \mathbf{h}(\mathbf{V}), \quad \mathbf{h} : \mathbb{R}^2 \to \mathbb{R}^2, \quad \mathbf{h}(V_1, V_2) = (h_1(V_1, V_2), h_2(V_1, V_2))^{\mathsf{T}} = (V_2, V_1 V_2)^{\mathsf{T}} = (U_1, U_2)^{\mathsf{T}}.$$

Notice that the Jacobian of $\mathbf{h}$ is $J = -V_2 = -U_1 \neq 0$ with probability 1. Then, according to (4) the joint p.d.f. of the random vector $\mathbf{V} = (V_1, V_2)^{\mathsf{T}}$ is given by

$$f_{V_1, V_2}(v_1, v_2) = f_{U_1, U_2}(v_2, v_1 v_2)|v_2|.$$

Hence, the p.d.f. of the quotient $V_1 = U_2/U_1$ is obtained as the V_2-marginal p.d.f. of $f_{V_1, V_2}(v_1, v_2)$, i.e.,

$$f_{V_1}(v_1) = \int_{\mathcal{D}(U_1)} f_{U_1, U_2}(u_1, v_1 u_1)|u_1| \mathrm{d}u_1,$$

where $\mathcal{D}(U_1)$ denotes the domain of U_1 and the relationship $V_2 = U_1$ has been used.

Summarizing, the following result has been established:

Proposition 2. *(RVT technique: quotient of two r.v.'s). Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\mathbf{U} = (U_1, U_2)^{\mathsf{T}}$ an absolutely continuous random vector with joint p.d.f. $f_{U_1, U_2}(u_1, u_2)$ such that $\mathbb{P}[\{\omega \in \Omega : U_1(\omega) \neq 0\}] = 1$. Then, the p.d.f. $f_V(v)$ of the quotient $V = U_2/U_1$ is given by*

$$f_V(v) = \int_{\mathcal{D}(U_1)} f_{U_1, U_2}(\xi, v\xi)|\xi| \, \mathrm{d}\xi. \tag{6}$$

If U_1 and U_2 are independent r.v.'s with p.d.f.'s $f_{U_1}(u_1)$ and $f_{U_2}(u_2)$, respectively, then (6) becomes

$$f_V(v) = \int_{\mathcal{D}(U_1)} f_{U_1}(\xi) f_{U_2}(v\xi)|\xi| \, \mathrm{d}\xi.$$

This paper is organized as follows. Section 2 is devoted to provide a complete probabilistic description of random SIS-type epidemiological model (1)–(3) under general hypotheses regarding the probability distribution of the random inputs. For the sake of clarity, this section is divided into several pieces. Subsection 2.1 deals with the computation of the first probability distributions of the percentages of susceptibles and infected. The determination of mean and variance functions as well as confidence intervals for the percentages of susceptibles is established in Subsection 2.2. Subsection 2.3 is addressed to compute the distributions of time until a given proportion of the population remains susceptible and infected, respectively. Section 2 ends by providing a probabilistic description of the basic reproductive number, which plays a key role in dealing with the behaviour of a disease in the long run. In Section 3, all the theoretical results obtained in Section 2 are applied to study, using available real data, the evolution of Spanish non-smokers men older than 16 years old under the SIS model approach. Conclusions are drawn in Section 4.

2. Probabilistic solution of the random SIS model

This section is devoted to determine the probabilistic solution of the random SIS model (1)–(3). This means the computation of the 1-p.d.f. of each component, $S(t)$ and $I(t)$, of its solution, $(S(t), I(t))^{\mathsf{T}}$, as well as its main statistical properties, such as the mean and variance functions and confidence intervals. To complete useful information in practical situations, the p.d.f.'s of the time until a given proportion of the population remains susceptible/infected and the so-called reproductive number will also be investigated.

To achieve these goals, first, taking into account (3), it is convenient to rewrite (1)–(2) as the following random nonlinear initial value problem (i.v.p.)

$$\begin{cases} S'(t) & = & \beta(S(t))^2 - (\gamma + \beta)S(t) + \gamma, \quad t > 0, \\ S(0) & = & S_0. \end{cases} \tag{7}$$

As it has been motivated in the previous section, throughout the exposition we will assume that the input parameters S_0, γ and β are r.v.'s, and

$$\begin{aligned} \mathcal{D}_{S_0} & = & \{ s_0 = S_0(\omega), \omega \in \Omega : 0 \le s_{0,1} \le s_0 \le s_{0,2} \le 1 \}, \\ \mathcal{D}_{\gamma} & = & \{ \gamma = \gamma(\omega), \omega \in \Omega : 0 \le \gamma_1 \le \gamma \le \gamma_2 \}, \\ \mathcal{D}_{\beta} & = & \{ \beta = \beta(\omega), \omega \in \Omega : 0 \le \beta_1 \le \beta \le \beta_2 \}, \end{aligned} \tag{8}$$

will denote their respective domains. Hereinafter, $f_{S_0,\gamma,\beta}(s_0, \gamma, \beta)$ will indicate the joint p.d.f. of the input random vector (S_0, γ, β). At this point, it is important to underline that in order to provide more generality at the results that will be obtained later, we will handle the joint p.d.f. $f_{S_0,\gamma,\beta}(s_0, \gamma, \beta)$, although in the example the input parameters S_0, γ and β will be assumed pairwise independent r.v.'s. For the sake of clarity in the presentation, we have assumed that the domain of $f_{S_0,\gamma,\beta}(s_0, \gamma, \beta)$ is the product of intervals given by (8) rather than any arbitrary subset of $\mathbb{R}^3$.

From a theoretical point of view is important to guarantee the uniqueness of the solution to the system of r.d.e.'s (1)–(3), or equivalently, the uniqueness of the solution $S(t)$ to initial value problem (7) on a finite interval, $t \in [0, a]$, $a > 0$. To achieve this goal, we will apply Theorem 5.1.2. [15, p.118], which is a natural generalization of the classical Picard theorem based upon convergence of successive approximations. This result is stated in terms of a strong stochastic convergence, namely, mean square convergence. This type of stochastic convergence is defined on the Hilbert space $(L_2, \|\cdot\|)$ of the second-order r.v.'s, where the norm is defined in terms of the expectation operator: $\|X\| = \left(\mathbb{E}\left[X^2 \right] \right)^{1/2}$. Following the standard notation introduced in Theorem 5.1.2. [15, p.118], let us consider the right-hand side of r.d.e. (7), i.e., $f(S, t) = \beta S^2 - (\gamma + \beta) S + \gamma$, and the pair defining the random initial condition $(t_0, X_0) = (0, S_0)$. In accordance with (8), let us assume that β and γ have bounded realizations, hence they are second-order r.v.'s. Moreover, observe that S is also a second-order r.v. because $0 \le S(\omega) \le 1$ for each $\omega \in \Omega$. First, we need to check the following mean square Lipschitz condition

$$\|f(S_1, t) - f(S_2, t)\| \le k(t) \|S_1 - S_2\|, \quad \text{where} \quad \int_0^a k(t)\mathrm{d}t < \infty. \tag{9}$$

For convenience, let us denote by $\delta = \max\{\beta_2, \gamma_2\}$, being β_2 and γ_2 the upper bounds defined in (8). Since S_1 and S_2 are functions of the random inputs S_0, β and γ, and by (3), $0 \le S_1(\omega) +$

$S_2(\omega) \leq 2$ for all $\omega \in \Omega$, one gets

$$
\begin{aligned}
\|f(S_1,t) - f(S_2,t)\|^2 &= \left\| \beta((S_1)^2 - (S_2)^2) - (\gamma + \beta)(S_1 - S_2) \right\|^2 \\
&= \| (\beta(S_1 + S_2) - (\gamma + \beta))(S_1 - S_2) \|^2 \\
&= \mathbb{E}\left[(\beta(S_1 + S_2) - (\gamma + \beta))^2 (S_1 - S_2)^2 \right] \\
&= \int_{\mathcal{D}_{S_0}} \int_{\mathcal{D}_\gamma} \int_{\mathcal{D}_\beta} (\beta(S_1 + S_2) - (\gamma + \beta))^2 (S_1 - S_2)^2 f_{S_0,\gamma,\beta}(s_0, \gamma, \beta) \, d\beta \, d\gamma \, ds_0 \\
&\leq \int_{\mathcal{D}_{S_0}} \int_{\mathcal{D}_\gamma} \int_{\mathcal{D}_\beta} (2\beta - (\gamma + \beta))^2 (S_1 - S_2)^2 f_{S_0,\gamma,\beta}(s_0, \gamma, \beta) \, d\beta \, d\gamma \, ds_0 \qquad (10) \\
&= \int_{\mathcal{D}_{S_0}} \int_{\mathcal{D}_\gamma} \int_{\mathcal{D}_\beta} (\beta - \gamma)^2 (S_1 - S_2)^2 f_{S_0,\gamma,\beta}(s_0, \gamma, \beta) \, d\beta \, d\gamma \, ds_0 \\
&\leq \delta^2 \int_{\mathcal{D}_{S_0}} \int_{\mathcal{D}_\gamma} \int_{\mathcal{D}_\beta} (S_1 - S_2)^2 f_{S_0,\gamma,\beta}(s_0, \gamma, \beta) \, d\beta \, d\gamma \, ds_0 \\
&= \delta^2 \mathbb{E}\left[(S_1 - S_2)^2 \right] \\
&= \delta^2 \|S_1 - S_2\|^2 .
\end{aligned}
$$

Therefore, Lipschitz condition (9) holds true taking $k(t) = \delta$. Notice that

$$
\int_0^a k(t)\,dt = \int_0^a \delta \, dt = \delta a < \infty.
$$

The second condition to be checked is that the function $f(S,t)$ transforms second-order r.v.'s into second-order r.v.'s. This is a direct consequence of the previous development, the triangular inequality and the fact that γ and S are also second-order r.v.'s

$$
\|f(S,t)\| = \left\| \beta S^2 - (\beta + \gamma)S + \gamma \right\| \leq \left\| \beta S^2 - (\beta + \gamma)S \right\| + \|\gamma\| \leq \delta \|S\| + \|\gamma\| < \infty.
$$

2.1. First probability density function (1-p.d.f.)

The aim of this subsection is to obtain the 1-p.d.f., $f_1(s,t)$, of the solution s.p., $S(t)$, of the random i.v.p. (7). This function provides a full probabilistic description of the percentage of susceptibles at every time instant t. Once this goal has been achieved, using (3), the 1-p.d.f. of the percentage of infected $I(t)$ will also be provided.

First, notice that the solution s.p. $S(t)$ of the random i.v.p. (7) is given by

$$
S(t) = \frac{\gamma(1 - S_0) + (S_0\beta - \gamma)e^{(\gamma - \beta)t}}{\beta(1 - S_0) + (S_0\beta - \gamma)e^{(\gamma - \beta)t}}, \quad t \geq 0. \qquad (11)
$$

Now, we fix $t \geq 0$ and we apply the R.V.T. method with the following choice in Theorem 1,

$$
\mathbf{U} = (S_0, \gamma, \beta)^\top, \quad \mathbf{V} = (X, Y, Z)^\top, \quad \mathbf{V} = \mathbf{g}(\mathbf{U})^\top,
$$

7

$$\mathbf{g} : \mathbb{R}^3 \to \mathbb{R}^3, \quad \mathbf{g}(S_0, \gamma, \beta) = (g_1(S_0, \gamma, \beta), g_2(S_0, \gamma, \beta), g_3(S_0, \gamma, \beta))^\mathsf{T} = (X, Y, Z)^\mathsf{T},$$

being

$$X = \frac{\gamma(1 - S_0) + (S_0\beta - \gamma)e^{(\gamma-\beta)t}}{\beta(1 - S_0) + (S_0\beta - \gamma)e^{(\gamma-\beta)t}}, \quad Y = \gamma, \quad Z = \beta.$$

Isolating S_0, β and γ, one gets

$$S_0 = \frac{Y + e^{(Y-Z)t}(-1 + X)Y - XZ}{Y + e^{(Y-Z)t}(-1 + X)Z - XZ}, \quad \gamma = Y, \quad \beta = Z. \tag{12}$$

Thus, in accordance with (4)–(5), the joint p.d.f. of the random vector $(X, Y, Z)^\mathsf{T}$ is given by

$$f_{X,Y,Z}(x, y, z) = f_{S_0, \gamma, \beta}\left(\frac{y + e^{(y-z)t}(-1 + x)y - xz}{y + e^{(y-z)t}(-1 + x)z - xz}, y, z\right) \frac{e^{(y-z)t}(y - z)^2}{(y + e^{(y-z)t}(-1 + x)z - xz)^2}, \tag{13}$$

where $J = \left(e^{(y-z)t}(y - z)^2\right) / \left(y + e^{(y-z)t}(-1 + x)z - xz\right)^2 > 0$ is the Jacobian obtained from the inverse mapping of $\mathbf{g}$,

$$\mathbf{h} : \mathbb{R}^3 \to \mathbb{R}^3, \quad \mathbf{h}(X, Y, Z) = (h_1(X, Y, Z), h_2(X, Y, Z), h_3(X, Y, Z))^\mathsf{T} = (S_0, \gamma, \beta)^\mathsf{T},$$

defined according to (12).

Therefore, the p.d.f of the r.v. $S = S(t)$ with t fixed, is the (γ, β)-marginal p.d.f. of the joint p.d.f. (13), i.e.,

$$f_S(s) = \int_{\mathcal{D}_\gamma} \int_{\mathcal{D}_\beta} f_{S_0, \gamma, \beta}\left(\frac{\xi + e^{(\xi-\eta)t}(-1 + s)\xi - s\eta}{\xi + e^{(\xi-\eta)t}(-1 + s)\eta - s\eta}, \xi, \eta\right) \frac{e^{(\xi-\eta)t}(\xi - \eta)^2}{(\xi + e^{(\xi-\eta)t}(-1 + s)\eta - s\eta)^2} \, d\eta \, d\xi. \tag{14}$$

As the previous development is valid for every $t \geq 0$, the 1-p.d.f. of the solution s.p. $S(t)$ of the i.v.p. (7) is given by

$$f_1(s, t) = \int_{\mathcal{D}_\gamma} \int_{\mathcal{D}_\beta} f_{S_0, \gamma, \beta}\left(\frac{\xi + e^{(\xi-\eta)t}(-1 + s)\xi - s\eta}{\xi + e^{(\xi-\eta)t}(-1 + s)\eta - s\eta}, \xi, \eta\right) \frac{e^{(\xi-\eta)t}(\xi - \eta)^2}{(\xi + e^{(\xi-\eta)t}(-1 + s)\eta - s\eta)^2} \, d\eta \, d\xi. \tag{15}$$

Now, denoting $S = S(t)$ and $I = I(t)$ with t fixed, using the relationship (3) and Proposition 1 with the identification $U = S$, $V = I$, $a = -1 \neq 0$ and $b = 1$, from (14) one obtains the p.d.f. of the r.v. I and, hence straightforwardly the 1-p.d.f. of the percentage of infected

$$I(t) = \frac{(\beta - \gamma)(1 - S_0)e^{(\beta-\gamma)t}}{\beta(1 - S_0)e^{(\beta-\gamma)t} + S_0\beta - \gamma}, \quad t \geq 0, \tag{16}$$

which is given by

$$f_1(i, t) = \int_{\mathcal{D}_\gamma} \int_{\mathcal{D}_\beta} f_{S_0, \gamma, \beta}\left(\frac{\xi - \eta - e^{(\xi-\eta)t}i\xi + i\eta}{\xi - \eta - e^{(\xi-\eta)t}i\eta + i\eta}, \xi, \eta\right) \frac{e^{(\xi-\eta)t}(\xi - \eta)^2}{(\xi - \eta - e^{(\xi-\eta)t}i\eta + i\eta)^2} \, d\eta \, d\xi. \tag{17}$$

8

2.2. Mean and variance functions. Confidence intervals

In the context of the applications, the usefulness of the 1-p.d.f. is manifested in dealing with the computation of the main statistical information associated to the solution $(S(t), I(t))^\mathsf{T}$. Specifically, the mean and the variance functions of $S(t)$ are defined by

$$\mu_S(t) = \mathbb{E}[S(t)] = \int_{-\infty}^{\infty} s f_1(s, t)\, \mathrm{d}s, \quad (\sigma_S(t))^2 = \mathbb{V}[S(t)] = \int_{-\infty}^{\infty} s^2 f_1(s, t)\, \mathrm{d}s - (\mu_S(t))^2, \quad (18)$$

where $f_1(s, t)$ is given by (15). The same can be stated to regard the computation of $\mu_I(t)$ and $\sigma_I(t)$ changing $f_1(s, t)$ by $f_1(i, t)$, which has been determined in (17). Apart from computing the mean statistical behaviour of the solution stochastic process of the SIS model and its variability, for $\hat{t} \geq 0$ fixed, we might be interested in determining the probability that the percentage of susceptibles lies within a specific interval, say, $[s_1(\hat{t}), s_2(\hat{t})] = [\hat{s}_1, \hat{s}_2]$. It can be easily determined, once $f_1(s, t)$ has been computed, by the following integral

$$\mathbb{P}(\{\omega \in \Omega : S(\hat{t}; \omega) \in [\hat{s}_1, \hat{s}_2]\}) = \int_{\hat{s}_1}^{\hat{s}_2} f_1(s, \hat{t})\, \mathrm{d}s. \quad (19)$$

Regarding model validation and predictions, the 1-p.d.f. is also very useful to construct probabilistic intervals for any $(1 - \alpha) \times 100\%$ confidence level. For instance, fixed $\alpha \in (0, 1)$, for each $\hat{t} \geq 0$ fixed, one can determine $s_1(\hat{t})$ and $s_2(\hat{t})$ such that

$$\int_0^{s_1(\hat{t})} f_1(s, \hat{t})\, \mathrm{d}s = \frac{\alpha}{2} = \int_{s_2(\hat{t})}^1 f_1(s, \hat{t})\, \mathrm{d}s, \quad (20)$$

where

$$1 - \alpha = \mathbb{P}(\{\omega \in \Omega : S(\hat{t}; \omega) \in [s_1(\hat{t}), s_2(\hat{t})]\}) = \int_{s_1(\hat{t})}^{s_2(\hat{t})} f_1(s, \hat{t})\, \mathrm{d}s. \quad (21)$$

Commonly $\alpha = 0.05$, which means that 95% confidence intervals are built.

Notice that analogous expressions to (19)–(21) can be given for the percentage of infected simply interchanging $f_1(s, \hat{t})$ by $f_1(i, \hat{t})$, where the latter function is defined by (17).

2.3. Distribution of time until a given proportion of the population remains susceptible

So far we have determined the distribution for the percentage of susceptibles (and infected) at every time instant t. From an applied standpoint, it is very helpful to compute when the percentage of susceptibles in the population will attain a specific level. With this end, let us denote by T_S the time until a given proportion of the population, ρ_S, remains susceptible, i.e., $\rho_S = S(T)$. Then, isolating T_S from the exact solution (11) of the i.v.p. (7) one gets

$$T_S = \frac{1}{\gamma - \beta} \ln\left(\frac{(1 - S_0)(\beta \rho_S - \gamma)}{(1 - \rho_S)(\beta S_0 - \gamma)}\right).$$

Now, we apply Theorem 1 to

$$\mathbf{U} = (S_0, \gamma, \beta)^\mathsf{T}, \quad \mathbf{V} = (X, Y, Z)^\mathsf{T}, \quad \mathbf{V} = \mathbf{g}(\mathbf{U}),$$

$$\mathbf{g} : \mathbb{R}^3 \to \mathbb{R}^3, \quad \mathbf{g}(S_0, \gamma, \beta) = (g_1(S_0, \gamma, \beta), g_2(S_0, \gamma, \beta), g_3(S_0, \gamma, \beta))^\mathsf{T} = (X, Y, Z)^\mathsf{T},$$

being

$$X = \frac{1}{\gamma - \beta} \ln\left(\frac{(1 - S_0)(\beta\rho_S - \gamma)}{(1 - \rho_S)(\beta S_0 - \gamma)}\right), \quad Y = \gamma, \quad Z = \beta.$$

Isolating S_0, γ and β, one gets

$$S_0 = \frac{Y(1 + e^{X(Y-Z)}(-1 + \rho_S)) - Z\rho_S}{Y + Z(e^{X(Y-Z)}(-1 + \rho_S) - \rho_S)}, \quad \gamma = Y, \quad \beta = Z. \tag{22}$$

Then, by applying expressions (4)–(5) and considering that $\rho_S \in (0, 1)$, the joint p.d.f. of the random vector $(X, Y, Z)^\mathsf{T}$ is given by

$$f_{X,Y,Z}(x, y, z) = f_{S_0,\gamma,\beta}\left(\frac{y(1 + e^{x(y-z)}(-1 + \rho_S)) - z\rho_S}{y + z(e^{x(y-z)}(-1 + \rho_S) - \rho_S)}, y, z\right) \frac{e^{x(y-z)}(y - z)^2(1 - \rho_S)}{(y + z(e^{x(y-z)}(-1 + \rho_S) - \rho_S))^2} |y - z\rho_S|, \tag{23}$$

where $J = \left(e^{x(y-z)}(y - z)^2(-1 + \rho_S)(y - z\rho_S)\right) / \left(y + z(e^{x(y-z)}(-1 + \rho_S) - \rho_S)\right)^2$ is the Jacobian obtained from the inverse mapping of $\mathbf{g}$,

$$\mathbf{h} : \mathbb{R}^3 \to \mathbb{R}^3, \quad \mathbf{h}(X, Y, Z) = (h_1(X, Y, Z), h_2(X, Y, Z), h_3(X, Y, Z))^\mathsf{T} = (S_0, \gamma, \beta)^\mathsf{T},$$

defined according to (22).

Therefore, taking into account that $(X, Y, Z)^\mathsf{T} = (T_S, \gamma, \beta)^\mathsf{T}$, the p.d.f of the r.v. T_S is the (γ, β)-marginal p.d.f. of the joint p.d.f. (23)

$$f_1(t, \rho_S) = \int_{\mathcal{D}_\gamma} \int_{\mathcal{D}_\beta} f_{S_0,\gamma,\beta}\left(\frac{\xi(1 + e^{t(\xi-\eta)}(-1 + \rho_S)) - \eta\rho_S}{\xi + \eta(e^{t(\xi-\eta)}(-1 + \rho_S) - \rho_S)}, \xi, \eta\right) \frac{e^{t(\xi-\eta)}(\xi - \eta)^2(1 - \rho_S)|\xi - \eta\rho_S|}{(\xi + \eta(e^{t(\xi-\eta)}(-1 + \rho_S) - \rho_S))^2} \, d\eta \, d\xi. \tag{24}$$

Following an analogous development, one can demonstrate that the distribution of time until a given proportion of the population remains infected, ρ_I, reads

$$f_1(t, \rho_I) = \int_{\mathcal{D}_\gamma} \int_{\mathcal{D}_\beta} f_{S_0,\gamma,\beta}\left(\frac{\xi + \eta(-1 + \rho_I) - e^{t(\xi-\eta)}}{\xi - \eta(1 + (-1 + e^{t(\xi-\eta)})\rho_I)}, \xi, \eta\right) \left|\frac{e^{t(\xi-\eta)}(\xi - \eta)^2(\xi + \eta(-1 + \rho_I))\rho_I}{(\xi - \eta(1 + (-1 + e^{t(\xi-\eta)})\rho_I))^2}\right| \, d\eta \, d\xi.$$

2.4. Basic reproductive number

In epidemiology, the basic reproduction number, R_0, associated to an infection is useful to elucidate whether will spread out or not. In the case of the SIS model (1)–(3), this value and its relationship with the propagation of the epidemic in the long run is given by

$$R_0 = \frac{\beta}{\gamma}, \quad \begin{cases} \text{if} & R_0 < 1 \equiv \beta < \gamma, \quad \text{then the diseases will die out as } t \to +\infty, \\ \text{if} & R_0 > 1 \equiv \beta > \gamma, \quad \text{then the diseases will spread out as } t \to +\infty. \end{cases}$$

This classification is easily derived from expression (16) of $I(t)$, or equivalently of $S(t)$ (see (11)), since

$$\lim_{t\to+\infty} I(t) = \lim_{t\to+\infty} \frac{(\beta - \gamma)(1 - S_0)e^{(\beta-\gamma)t}}{\beta(1 - S_0)e^{(\beta-\gamma)t} + S_0\beta - \gamma} = 0 \quad \text{if} \quad \beta < \gamma,$$

$$\lim_{t\to+\infty} S(t) = \lim_{t\to+\infty} \frac{\gamma(1 - S_0) + (S_0\beta - \gamma)e^{(\gamma-\beta)t}}{\beta(1 - S_0) + (S_0\beta - \gamma)e^{(\gamma-\beta)t}} = 1 \quad \text{if} \quad \beta < \gamma.$$

Notice that this result is in agreement of its intuitive interpretation, the epidemic will die out when the rate of decline in the percentage of susceptibles be less than the rate of infected that recover from the diseases, in other words, when condition $\beta < \gamma$ holds.

In our context, both β and γ are assumed to be r.v.'s, so that the requirement for epidemic extinction in the deterministic framework $\beta < \gamma$ means the computation of the following probability in the stochastic scenario

$$\mathbb{P}[\mathcal{S}], \quad \mathcal{S} = \{\omega \in \Omega : \beta(\omega) < \gamma(\omega)\} = \{\omega \in \Omega : R_0(\omega) < 1\}. \tag{25}$$

This key probability can be computed by taking advantage of Proposition 2. In fact, using the following identification between the notation for the SIS model and the one used in Proposition 2

$$\mathbf{U} = (U_1, U_2)^{\mathsf{T}} = (\gamma, \beta)^{\mathsf{T}}, \quad V = \frac{U_2}{U_1} = \frac{\beta}{\gamma} = R_0,$$

one gets

$$f_{R_0}(r_0) = \int_{\mathcal{D}(\gamma)} f_{\gamma,\beta}(\xi, r_0\xi) \, |\xi| \, \mathrm{d}\xi,$$

where $f_{\gamma,\beta}(\cdot, \cdot)$ denotes the (γ, β)–marginal distribution of the joint p.d.f. of the random inputs S_0, γ and β. As a consequence, taking into account that $R_0(\omega) > 0$ for all $\omega \in \Omega$, the target probability (25) can be computed as follows

$$\mathbb{P}[\mathcal{S}] = \int_0^1 \int_{\mathcal{D}(\gamma)} f_{\gamma,\beta}(\xi, r_0\xi) \, |\xi| \, \mathrm{d}\xi \, \mathrm{d}r_0. \tag{26}$$

3. An illustrative example

In this section we will show how to take advantage of the theoretical probabilistic results previously obtained for the random SIS model (1)–(3) (or equivalently, (7)), in order to study the spread of smoking in Spain.

Example 1. *SIS-type epidemiological models are useful to describe the spread of diseases whose infection does not confer immunity among the individuals of a population which is divided into two groups, susceptible and infected. Many social habits can also be treated by SIS model like smoking. Notice that once a person gives up smoking, he/she could become a smoker again in the future. In this example, we will apply the SIS model (1)–(3) to study the evolution of smoking among the Spanish men aged over 16 years old. In our context this population has been divided into two groups, non-smokers (susceptible) and smokers (infected). Table 1 collects the percentage of non-smokers, denoted by S_j, for the available data during the period 1987– 2006, that correspond to the time values t_j, $j \in \mathcal{J} = \{0, 6, 8, 10, 14, 16, 19\}$, respectively. In order to apply model (7), we have to chose specific probability distributions for the random input parameters S_0, β and γ, which hereinafter we will assume to be independent r.v.'s. As the initial condition S_0 represents a percentage, we have made the decision of assuming that S_0 has a beta distribution of parameters $a > 0$, $b > 0$, i.e., $S_0 \sim Be(a; b)$. This is supported by the domain of a beta r.v., which is the interval $(0, 1)$. In addition, the beta distribution is more flexible than other candidates like the $(0, 1)$-uniform distribution since it is a two-parametric distribution. It permits to adapt it better to model complex situations. For the rate of decline β, we will take an exponential distribution of parameter, $\lambda_\beta > 0$, truncated on the interval $(0, 1000)$, because this*

rate must be modelled by a positive r.v. Finally, for the rate of recovery, γ, we have chosen a truncated Gaussian distribution of parameters $\mu_\gamma > 0$ and $\sigma_\gamma > 0$ on the interval $(0, 1)$. This choice has been made because the positiveness of r.v. γ and the great flexibility of Gaussian distribution. At this point, it is important to emphasize that the generality of the results provided in the previous section permits to choice other probability distributions for the input parameters different to the aforementioned ones whenever they match their epidemiological interpretation (positiveness, boundedness, etc). Moreover, notice that truncation for r.v.'s β and γ allows us to guarantee that hypotheses related to uniqueness are fulfilled.

year	1987	1993	1995	1997	2001	2003	2006
(t_j)	$(j = 0)$	$(j = 6)$	$(j = 8)$	$(j = 10)$	$(j = 14)$	$(j = 16)$	$(j = 19)$
S_j	0.4488	0.5144	0.5278	0.5514	0.5783	0.6244	0.6467

Table 1: Percentage of non-smokers Spanish men aged over 16 years old during the period 1987–2006. Source [27].

In order to determine the positive parameters a, b, λ_β, μ_γ and σ_γ associated to the probability distributions of the model parameters, we will minimize the mean square error, $E(a, b, \lambda_\beta, \mu_\gamma, \sigma_\gamma)$, between the data S_j and the expectation of $S(t) = S(t; a, b, \lambda_\beta, \mu_\gamma, \sigma_\gamma)$ evaluated at the time instants t_j, $j \in \mathcal{J}$. It leads to the following optimization programme

$$\min_{a, b, \lambda_\beta, \mu_\gamma, \sigma_\gamma > 0} E(a, b, \lambda_\beta, \mu_\gamma, \sigma_\gamma) = \sum_{j \in \mathcal{J}} \left(S_j - \mathbb{E}[S(t_j; a, b, \lambda_\beta, \mu_\gamma, \sigma_\gamma)] \right)^2, \tag{27}$$

where, according to (18) and, taking into account that $S(t) = S(t; a, b, \lambda_\beta, \mu_\gamma, \sigma_\gamma) \in (0, 1)$, the above expectation can be computed as

$$\mathbb{E}[S(t_j; a, b, \lambda_\beta, \mu_\gamma, \sigma_\gamma)] = \int_0^1 s f_1(s, t_j)\,\mathrm{d}s, \quad j \in \mathcal{J}. \tag{28}$$

Now, we will explicit the form of the 1-p.d.f. $f_1(s, t)$ that appears in the above integral, and is defined by (15), taking into account all the hypotheses we have assumed up to now. For that, first notice that $S_0 \sim Be(a; b)$, $\beta \sim Exp(\lambda_\beta)$ truncated on the interval $(0, 1000)$ and $\gamma \sim N(\mu_\gamma; \sigma_\gamma)$ truncated on the interval $(0, 1)$, hence according to (8), their domains $\mathcal{D}_{S_0}$, $\mathcal{D}_\beta$ and $\mathcal{D}_\gamma$ are defined by $s_{0,1} = 0$, $s_{0,2} = 1$; $\beta_1 = 0$, $\beta_2 = 1000$, and $\gamma_1 = 0$, $\gamma_2 = 1$, respectively. Therefore,

$$f_1(s, t) = \int_0^1 \int_0^{1000} f_{S_0}\left(\frac{\xi + e^{(\xi - \eta)t}(-1 + s)\xi - s\eta}{\xi + e^{(\xi - \eta)t}(-1 + s)\eta - s\eta}\right) f_\gamma(\xi) f_\beta(\eta) \frac{e^{(\xi - \eta)t}(\xi - \eta)^2}{(\xi + e^{(\xi - \eta)t}(-1 + s)\eta - s\eta)^2}\,\mathrm{d}\eta\,\mathrm{d}\xi, \tag{29}$$

where

$$\begin{aligned} f_{S_0}\left(\frac{\xi + e^{(\xi - \eta)t}(-1 + s)\xi - s\eta}{\xi + e^{(\xi - \eta)t}(-1 + s)\eta - s\eta}\right) &= \frac{\Gamma(a + b)}{\Gamma(a)\Gamma(b)}\left(\frac{\xi + e^{(\xi - \eta)t}(-1 + s)\xi - s\eta}{\xi + e^{(\xi - \eta)t}(-1 + s)\eta - s\eta}\right)^{a-1} \\ &\times \left(\frac{e^{(\xi - \eta)t}(-1 + s)(\eta - \xi)}{\xi + e^{(\xi - \eta)t}(-1 + s)\eta - s\eta}\right)^{b-1}, \end{aligned} \tag{30}$$

$$f_\beta(\eta) = \frac{\lambda_\beta e^{-\lambda_\beta \eta}}{\int_0^{1000} \lambda_\beta e^{-\lambda_\beta \eta}\,\mathrm{d}\eta}, \tag{31}$$

and

$$f_\gamma(\xi) = \begin{cases} e^{-\frac{(\xi-\mu_\gamma)^2}{2(\sigma_\gamma)^2}} \dfrac{1}{\sqrt{2\pi}\sigma_\gamma\left(\frac{1}{2}erfc\left(\frac{\mu_\gamma-1}{\sqrt{2}\sigma_\gamma}\right) - \frac{1}{2}erfc\left(\frac{\mu_\gamma}{\sqrt{2}\sigma_\gamma}\right)\right)}, & if \quad 0 < \xi \le 1, \\ \\ 0, & otherwise, \end{cases} \tag{32}$$

being $erfc(z) = 1 - \frac{2}{\sqrt{\pi}}\int_0^z e^{-t^2}\,dt$ *the complementary error function.*

In order to compute the minimum of the error function, $E(a, b, \lambda_\beta, \mu_\gamma, \sigma_\gamma)$, given by (27)–(32), we have used the Nelder-Mead algorithm [28] carrying out computations by Mathematica®[29]. The obtained results are

$$a^* = 708.755, \quad b^* = 893.394, \quad \lambda_\beta^* = 1362.230, \quad \mu_\gamma^* = 0.0231162, \quad \sigma_\gamma^* = 0.0000526. \tag{33}$$

Figure 2 shows the 1-p.d.f. (29)–(32) for the optimal values a^*, b^*, λ_β^*, μ_γ^* and σ_γ^* given by (33). We observe that the center of the mass of probability of the r.v. percentage of susceptibles at the time t, $S(t)$, rises while its dispersion reduces as time increases. This behaviour is better observed in Figure 3 where the mean, $\mu_S(t)$, and the standard deviation, $\sigma_S(t)$, of $S(t)$ have been depicted. The obtained results agree with the historical data from which one observes that the percentage of Spanish non-smokers men increases as time goes on.
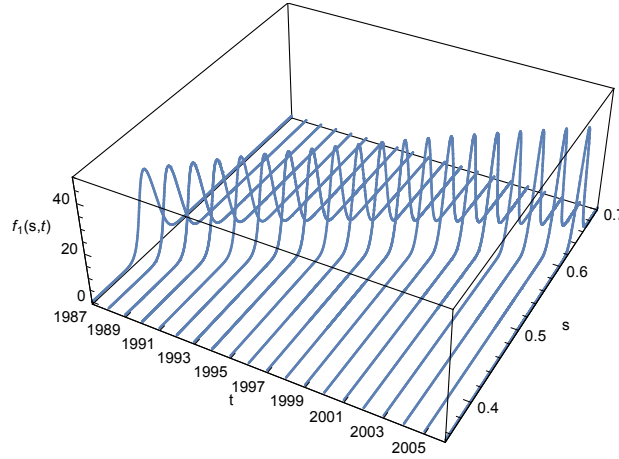

Figure 2: Plot of $f_1(s, t)$ in Example 1 during the period 1987–2006 (corresponding to the solid lines).

In order to validate the model, we have built confidence intervals. As we know exactly the p.d.f. of $S(t)$ for every t, we can determine these intervals using (20)–(21). In this case, we notice that their approximation using Gaussian confidence intervals of the form $\mu_S(t) \pm 2\sigma_S(t)$ provides very similar numerical results for every year of the period 1987–2006. In Figure 4 these intervals together with the real data are plotted.

In Table 2, we present the confidence levels for the approximate intervals $\mu_S(t) \pm 2\sigma_S(t)$ using the exact p.d.f. (15) for the years where data is available, although figures for the rest of the period are similar. Notice that these figures are very close to 95.45%, that corresponds to the case that the p.d.f. is Gaussian. From Figure 4, we observe that all the data lies within the confidence intervals, hence the model is capable of capturing the variability of the real data, and, as a consequence, the model can be validated at 95% confidence level.
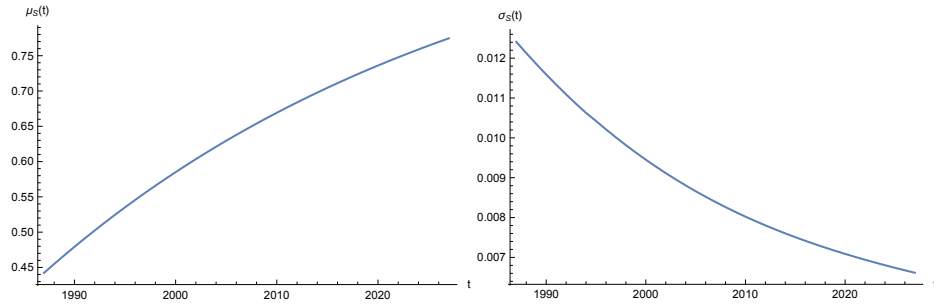
13

Figure 3: Plot of expectation function (left) and standard deviation function (right) in Example 1.
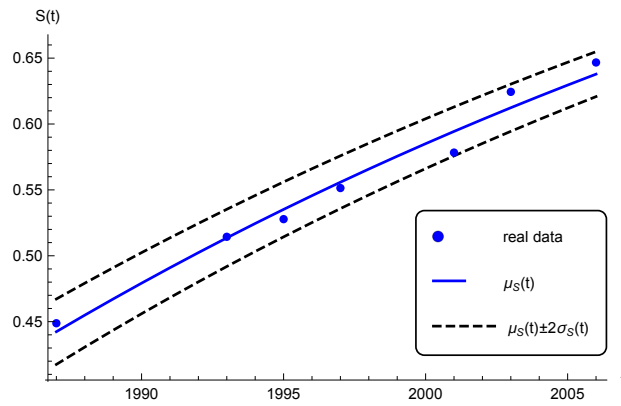

Figure 4: Expectation (solid line) and confidence intervals (dotted lines) in Example 1.

year	1987	1993	1995	1997	2001	2003	2006
(t_j)	$(j = 0)$	$(j = 6)$	$(j = 8)$	$(j = 10)$	$(j = 14)$	$(j = 16)$	$(j = 19)$
Confidence level	0.9550	0.9544	0.9545	0.9546	0.9549	0.9550	0.9552

Table 2: Probabilities associated to the confidence intervals built according to the SIS model.

Figure 5 shows the 1-p.d.f. of the time T_S until a proportion ρ_S of the population remains non-smokers (susceptible) for different values of $\rho_S \in \{0.45, 0.50, 0.55, 0.60, 0.65, 0.70, 0.75\}$. The plot has been performed using the following expression

$$f_1(t, \rho_S) = \int_0^1 \int_0^{1000} f_{S_0}\left(\frac{\xi(1 + e^{t(\xi-\eta)}(-1 + \rho_S)) - \eta\rho_S}{\xi + \eta(e^{t(\xi-\eta)}(-1 + \rho_S) - \rho_S)}\right) f_\gamma(\xi) f_\beta(\eta) \frac{e^{t(\xi-\eta)}(\xi-\eta)^2(1-\rho_S)|\xi - \eta\rho_S|}{(\xi + \eta(e^{t(\xi-\eta)}(-1 + \rho_S) - \rho_S))^2} \, \mathrm{d}\eta \, \mathrm{d}\xi\,,$$

derived from (24), taking into account that S_0, β and γ are independent r.v.'s and, that

$$
\begin{aligned}
f_{S_0}\left(\frac{\xi(1 + e^{t(\xi-\eta)}(-1 + \rho_S)) - \eta\rho_S}{\xi + \eta(e^{t(\xi-\eta)}(-1 + \rho_S) - \rho_S)}\right) &= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\left(\frac{\xi(1 + e^{t(\xi-\eta)}(-1 + \rho_S)) - \eta\rho_S}{\xi + \eta(e^{t(\xi-\eta)}(-1 + \rho_S) - \rho_S)}\right)^{a-1} \\
&\times \left(\frac{(\eta - \xi)e^{t(\xi-\eta)}(-1 + \rho_S)}{\xi + \eta(e^{t(\xi-\eta)}(-1 + \rho_S) - \rho_S)}\right)^{b-1}\,,
\end{aligned}
$$

and $f_\beta(\eta)$ and $f_\gamma(\xi)$ are given by (31) and (32), respectively. For instance, according to the 1-p.d.f. of T_S, one gets

$$\mathbb{E}[T_S] = \int_0^\infty t f_{T_S}(t; 0.75)\, \mathrm{d}t = 35.4013\,.$$

This means that the middle of the year 2023 approximately represents the average time until 75% of the Spanish men aged over 16 years old population will be non-smokers. This can be graphically seen in Figure 5. In Table 3, we have computed $\mathbb{E}[T_S]$ for different values of ρ_S. This information is crucial for health authorities in order to know the evolution of smoking and, hence adopting preventive and treatment campaigns.

ρ_S	0.45	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85
$\mathbb{E}[T_S]$	0.59	4.78	9.42	14.61	20.51	27.32	35.40	45.30	58.10

Table 3: Expectation of time T_S until a proportion, ρ_S, of the population remains non-smoker for different values ρ_S.

Finally, using (26) we compute the probability of the event $\mathcal{S}$ introduced in (25)

$$\mathbb{P}[\mathcal{S}] = \int_0^1 \int_0^1 f_\gamma(\xi) f_\beta(r_0\xi)\,|\xi|\,\mathrm{d}\xi\,\mathrm{d}r_0 = 0.999453\,,$$

where $f_\beta(\eta)$ and $f_\gamma(\xi)$ are defined by (31) and (32), respectively. Notice that this result is in accordance with the interpretation of the basic reproductive number R_0 and, it informs us that very likely the percentage of Spanish smoker men older than 16 years old will disappear as t tends to $+\infty$.

4. Conclusions

In this paper a full probabilistic description of the random SIS model has been provided. The obtained results are very general in the sense that all the involved input parameters have been assumed to be random variables having any probability density function. Therefore, the study includes the possibility that they are statistically either dependent or independent. In this context, significant probabilistic information has been determined. In a first step, the first probability
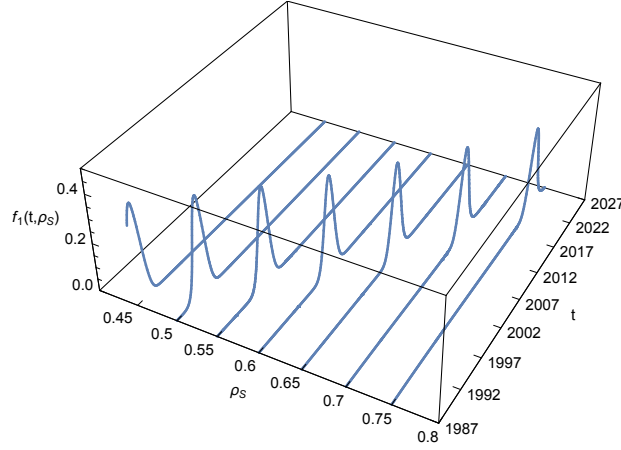
Figure 5: Plot of the 1-p.d.f. of the time T_S until a proportion $\rho_S \in \{0.45, 0.50, 0.55, 0.60, 0.65, 0.70, 0.75\}$ of the population remains susceptible in Example 1.

density function of the solution stochastic process of the governing nonlinear differential equation has been obtained. From this crucial function, the solution is completely characterized in each time instant since every statistical moment of the solution can be computed. In particular, mean and variance functions are easily determined. A key point regarding the explicit determination of the first probability density function is the computation of confidence intervals, whose confidence level is set accurately. This prevents inadvertent use of asymptotic approximations based on the central limit theorem which often are difficult to justify. All the above information permits providing a probabilistic description of the SIS model that generalizes its deterministic counterpart. An important contribution of the paper is the determination of the first probability density function of the time until a given proportion of the population remains susceptible or infected. This is very useful from a practical standpoint since it permits forecasting the earliest time instant at which the susceptible subpopulation will reach a given threshold. The theoretical study has been completed by providing a stochastic interpretation of a very important parameter in epidemiology, namely, the basic reproductive number. To put forward all the theoretical results in practice, an illustrative example using real data has been shown. The obtained results seem to be quite reliable, although it must be noticed that the choice of the probability distributions associated to the random inputs is a delicate issue that constitutes itself a challenge in dealing with real problems. To improve both the theoretical study and its applications, it would be very interesting to consider a demographic model into the SIS model where births and deaths rates should be included. The analysis of this more detailed model may benefit of the results developed in this paper. Analogous comments could be applied to the extensions SIR, SEIR, etc., of the SIS model.

Acknowledgements

Conflict of Interest Statement

The authors declare that there is no conflict of interests regarding the publication of this article.

[1] M. C. Casabán, J. C. Cortés, J. V. Romero, M. D. Roselló, Probabilistic solution of random SI-type epidemiological models using the random variable transformation technique, Communcations Nonlinear Science and Numerical Simulation 24 (1–3) (2015) 86–97. doi:10.1016/j.cnsns.2014.12.016.

[2] B. Kegan, R. W. West, Modeling the simple epidemic with deterministic differential equations and random initial conditions, Mathematical Biosciences 195 (5) (2005) 179–193. doi:10.1016/j.mbs.2005.02.004.

[3] E. Rogers, Diffusion of Innovations, Free Press, New York, 2003.

[4] D. Zhang, A. Ntoko, Computational Methods in Decision-Making, Economics and Finance, Vol. 74 of Applied Optimization, Springer, New York, 2002, Ch. Mathematical Model of Technology Diffusion in Developing Countries, pp. 526–539.

[5] H. Hethcote, Biomathematics, Vol. 18 of Applied Mathematical Ecology, Springer-Verlag, Heidelberg, 1989, Ch. Three Basic Epidemiological Models, pp. 120–144.

[6] H. W. Hethcote, The mathematics of infectious diseases, SIAM Review 42 (4) (2000) 599–653. doi:10.1137/S0036144500371907.

[7] F. Brauer, C. Castillo-Chávez, Mathematical Models in Population Biology and Epidemiology, Springer, New York, 2001.

[8] C. Castillo-Chávez, W. Huang, J. Li, Competitive exclusion in gonorrhea models and other sexually-transmitted diseases, SIAM Journal Applied Mathematics 56 (2) (1996) 494–508. doi:10.1137/S003613999325419X.

[9] H. Hethcote, J. Yorke, Gonorrhea Transmission Dynamics and Control, Vol. 56 of Lecture Notes in Biomathematics, Springer-Verlag, Berlin, 1984.

[10] J. C. Cortés, J. V. Romero, M. D. Roselló, R. J. Villanueva, Applying the Wiener-Hermite random technique to study the evolution of excess weight population in the region of Valencia (Spain), American Journal of Computational Mathematics 2 (4) (2012) 274–281. doi:10.4236/ajcm.2012.24037.

[11] D. Clancy, A stochastic SIS infection model incorporating indirect transmission, Journal of Applied Probability 42 (3) (2007) 726–737. doi:10.1239/jap/1127322023.

[12] C. Chen, K. Yanmei, Dynamics of a stochastic SIS epidemic model with saturated incidence, Abstract and Applied Analysis 2014–ID723825 (3) (2014) 1–13. doi:10.1155/2014/723825.

[13] B. Øksendal, Stochastic Differential Equations: An Introduction with Applications, 6th Edition, Springer, New York, 2010.

[14] E. Allen, Modeling with Itô Stochastic Differential Equations, Springer, New York, 2007.

[15] T. T. Soong, Random Differential Equations in Science and Engineering, Academic Press, New York, 1973.

[16] L. Villafuerte, C. A. Braumann, J. C. Cortés, L. Jódar, Random differential operational calculus: Theory and applications, Computers & Mathematics with Applications 59 (1) (2010) 115–125. doi:10.1016/j.camwa.2009.08.061.

[17] G. Casella, R. Berger, Statistical Inference, Brooks/Cole, New York, 2002.

[18] M. C. Casabán, J. C. Cortés, J. V. Romero, M. D. Roselló, Determining the first probability density function of linear random initial value problems by the random variable transformation (RVT) technique: A comprehensive study, Abstract and Applied Analysis 2014–ID248512 (2014) 1–25. doi:10.1155/2013/248512.

[19] M. C. Casabán, J. C. Cortés, J. V. Romero, M. D. Roselló, Probabilistic solution of random homogeneous linear second-order difference equations, Applied Mathematics Letters 34 (2) (2014) 27–32. doi:10.1016/j.aml.2014.03.0102.

[20] A. Hussein, M. M. Selim, Solution of the stochastic transport equation of neutral particles with anisotropic scattering using RVT technique, Applied Mathematics and Computation 213 (1) (2009) 250–261. doi:10.1016/j.amc.2009.03.016.

[21] A. Hussein, M. M. Selim, A developed solution of the stochastic Milne problem using probabilistic transformations, Applied Mathematics and Computation 216 (10) (2009) 2910–2919. doi:10.1016/j.amc.2010.04.003.

[22] A. Hussein, M. M. Selim, Solution of the stochastic radiative transfer equation with Rayleigh scattering using RVT technique, Applied Mathematics and Computation 218 (13) (2012) 7193–7203. doi:10.1016/j.amc.2011.12.088.

[23] L. T. Santos, F. A. Dorini, M. C. C. Cunha, The probability density function to the random linear transport equation, Applied Mathematics and Computation 216 (5) (2010) 1524–1530. doi:10.16/j.amc.2010.03.001.

[24] M. A. El-Tawil, W. El-Tahan, A. Hussein, Using FEM-RVT technique for solving a randomly excited ordinary differential equation with a random operator, Applied Mathematics and Computation 187 (2) (2007) 856–867. doi:10.16/j.amc.2006.08.164.

[25] A. Friedman, Stochastic Differential Equations and Applications, Vol. I, Academic Press, New York, 1975.

[26] I. I. Gikhman, A. V. Skorokhod, Stochastic Differential Equations, Springer–Verlag, New York, 1975.

[27] National Spanish Health Survey (Encuesta Nacional de Salud de España, ENSE), http://pestadistico.inteligenciadegestion.msssi.es/publicoSNS/comun/ArbolNodos.aspx, accessed: 2015-02-15.

[28] J. A. Nelder, R. Mead, A simplex method for function minimization, Computer Journal 7 (1) (1964) 308–313.

[29] Mathematica, http://www.wolfram.com/mathematica, accessed: 2015-02-15.