Conecting Markov process and the random variable transformation method to study a particular disease

J.-C. Cortés[1,2], A. Navarro-Quiles[3], J.-V. Romero[1], M.-D. Roselló[1,2], R.-J. Villanueva[2]

1: Colaboradores, 2: Directores. 3: Autor, Programa de Doctorado en Matemáticas.

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

Introduction

We propose a general methodology to study the stroke disease using a Markov model, where some parameters on the transition matrix will be considered random variables (r.v.'s). We will consider three states, Susceptible (S), Reliant (R) and Deceased (D). In Figure 1, we represent the influence diagram associated to the Markov model where transitions among states are included.

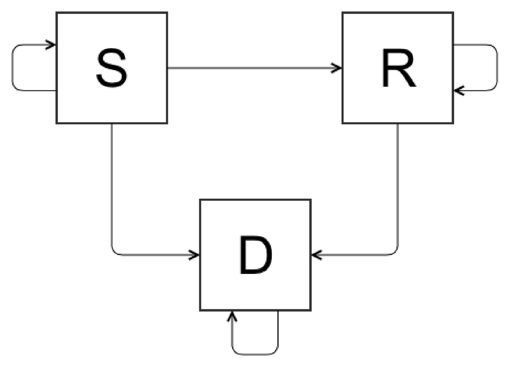

Fig. 1: Influence diagram for the Markov model.

The Markov model is formulated as follows

$$\left.\begin{array}{l}\begin{pmatrix} S_{n+1} \\ R_{n+1} \\ D_{n+1} \end{pmatrix} = \begin{pmatrix} \mathrm{e}^{-t_1 RR/1000} + \mathrm{e}^{-(T_2+t_3(RR-1))/1000} - 1 & 0 & 0 \\ 1-\mathrm{e}^{-t_1 RR/1000} & 1-P & 0 \\ 1-\mathrm{e}^{-(T_2+t_3(RR-1))/1000} & P & 1 \end{pmatrix} \begin{pmatrix} S_n \\ R_n \\ D_n \end{pmatrix}, \\ (S_0, R_0, D_0)^\top = (s_0, r_0, 0)^\top, \quad n = 0, 1, 2, \ldots, \end{array}\right\} \qquad (1)$$

where

- The relative risk, RR, the death rate due to any cause, T_2, and the probability of transition $R \to D$, P, are r.v.'s.
- S_n, R_n and D_n are the proportion of susceptibles, reliants and deceaseds in cycle n, respectively.
- $(s_0, r_0, 0)^\top$ is the initial cohort.
- t_1 and t_3 are the non–moral stroke and the stroke death rates, respectively.

We will assume that $S_n + R_n + D_n = 1$ for each n.

Theoretical Results

To obtain a full probabilistic description of the solution stochastic process to (1), we will apply RVT technique.

Theorem: RVT technique. [1, pp. 24–25]

Let $\mathbf{U} = (U_1, \ldots, U_n)^\top$ and $\mathbf{V} = (V_1, \ldots, V_n)^\top$ 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 $\mathbf{h} = \mathbf{g}^{-1} = (h_1(v_1, \ldots, v_n), \ldots, h_n(v_1, \ldots, v_n))^\top$ represents the inverse mapping of $\mathbf{g} = (g_1(u_1, \ldots, u_n), \ldots, g_n(u_1, \ldots, u_n))^\top$, the joint probability density function of vector $\mathbf{V}$ is given by

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

where $|J|$ is the absolute value of the Jacobian.

Applying this method, we obtain:

- The first probability density function (1-p.d.f.) of susceptible, reliant and deceased subpopulations. For example, for susceptible subpopulation the 1-p.d.f. is given by

$$\begin{aligned} f_1(s,n) = & \int_{\mathcal{D}_{R_n,P}} f_{RR,T_2,P}\left(-\frac{1000\ln(1-m_1)}{t_1}, t_3 + \frac{1000 t_3 \ln(1-m_1)}{t_1} - 1000\ln(1-m_2), p\right) \\ & \times \left|\frac{1000000}{t_1(1-m_1)(1-m_2)}\right| \left|\frac{\left(\frac{s}{s_0}\right)^{1/n}\left(-1+\left(\frac{s}{s_0}\right)^{1/n}+p\right)}{ns\left(s-s_0(1-p)^n\right)}\right| \mathrm{d}p\,\mathrm{d}r, \end{aligned}$$

  where

$$m_1 = \frac{\left(-1+p+\left(\frac{s}{s_0}\right)^{1/n}\right)\left(r_0(1-p)^n - r\right)}{(1-p)^n s_0 - s}, \quad m_2 = \frac{s - s\left(\frac{s}{s_0}\right)^{1/n} + \left(-s_0 + (r_0+s_0)\left(\frac{s}{s_0}\right)^{1/n} + r_0(-1+p)\right)(1-p)^n}{s - s_0(1-p)^n}.$$

- Time until a given proportion of the population remains susceptible, reliant or deceased.

With these 1-p.d.f.'s we can compute, for example, for susceptible subpopulation:

- The mean and the variance for each cycle n.
- Confidence intervals.
- The proportion of susceptibles that lies between a and b at a specific time period, say $\hat{n}$,

$$\mathbb{P}[a \le S_{\hat{n}} \le b] = \int_a^b f_1(s, \hat{n})\mathrm{d}s.$$

Graphical Examples

Based on [2], we assume that

- The initial condition is $(s_0, r_0, 0)^\top = (1, 0, 0)^\top$.
- RR, is a lognormal r.v. with parameters $(1.793, 0.143)$, i.e., $\ln(RR) \sim \mathrm{N}(1.793, 0.143)$.
- P is a beta r.v. with parameters $(80, 120)$, i.e., $P \sim \mathrm{Be}(80; 120)$.
- T_2 is a uniform r.v. on the interval $]21.27, 22.27[$, $T_2 \sim \mathrm{U}(]21.27, 22.27[)$.
- $t_1 = 1.11$ and $t_3 = 1.76$.
- RR, P and T_2 are pairwise independent r.v.'s.

In Figure 2, the 1-p.d.f.'s of susceptibles, reliants and deceaseds have been plotted. We can observe that when time increases the percentage of susceptibles decreases. Besides, the percentage of reliants increases at the beginning, specifically from $n = 1$ to $n = 6$, and afterwards this percentage decreases towards zero. With regard to deceased population, as is an absorbent state, all the population tends to this state. This is in agreement with results shown in Figure 2, where we can see that the percentage of deceaseds increases over the time. Besides, the variability in both susceptible and deceased subpopulations increases when times goes on. It is also interesting to observe that the 1-p.d.f. becomes sharper as standard deviation decreases.

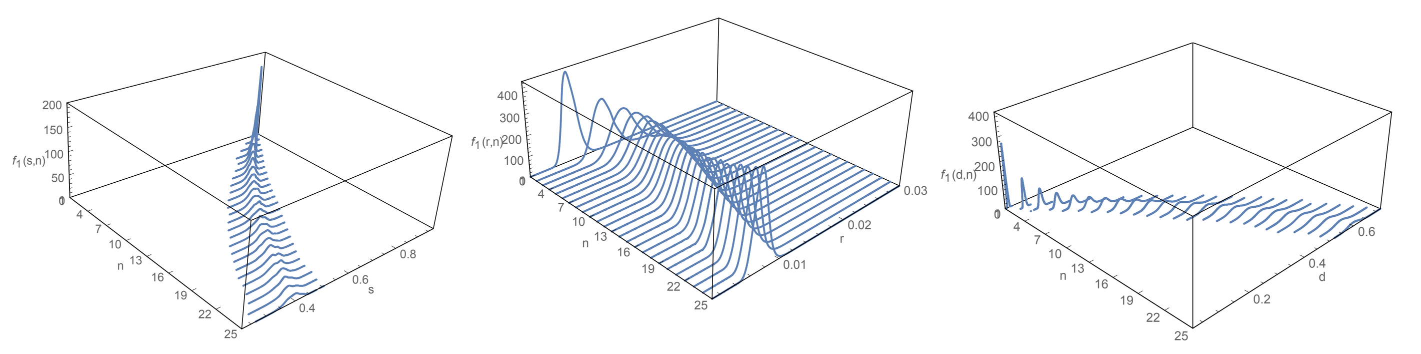

Fig. 2: Plot of the 1-p.d.f.'s of susceptibles (left), reliants (center) and deceaseds (right) at the following values of $n \in \{1, 2, \ldots, 25\}$.

Moreover, in Figure 3 it is shown the mean plus/minus the standard deviation functions of the three subpopulations. Notice that graphical representations shown in Figure 2 and Figure 3 are in agreement.

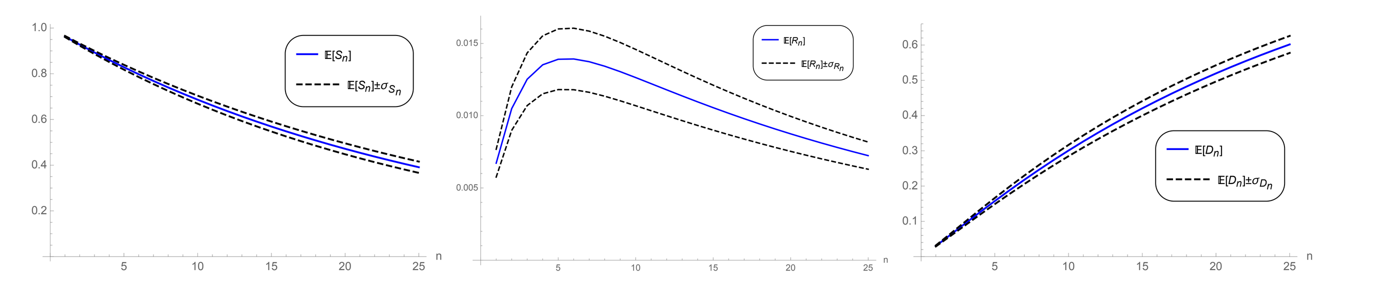

Fig. 3: Expectation plus/minus standard deviation functions for susceptibles (left), reliants (center) and deceaseds (right).

In addition, we can obtain the proportion of reliant subpopulation which lies between $a = 0.010$ and $b = 0.015$ in the time period $\hat{n} = 5$:

$$\mathbb{P}[0.010 \le R_5 \le 0.015] = \int_{0.010}^{0.015} f_1(r, 5)\mathrm{d}r = 0.700602.$$

In Figure 4 it is show the p.d.f. of the time, N_S, until a given proportion, ρ_S, of the population remains susceptible for different values of $\rho_S \in \{0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$.

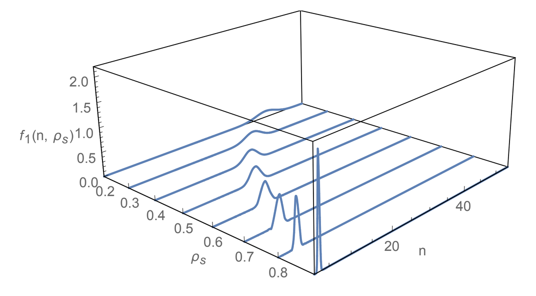

Fig. 4: Plot of the p.d.f. of the time N_S until a proportion $\rho_S \in \{0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$ of the population remains susceptible.

According to the p.d.f. of N_S, we can compute its expectation,

$$\mathbb{E}[N_S] = \int_0^\infty n f_1(n, 0.70)\mathrm{d}n = 9.51904.$$

This means that the middle of the cycle 9 represents, approximately, the average time until 70% of the population will be susceptible. This can be also seen in Figure 4.

References

1. T.T. Soong. Random differential Equations in Science and Engineering, Academic Press, New York, 1973.
2. J. Mar, F. Antoñanzas, R. Pradas, A. Arrospide. Los modelos de Markov probabilísticos en la evaluación económica de tecnologías sanitarias: una guía práctica, Gaceta Sanitaria 24 (3) (2010) 209–214. doi: 10.1016/j.gaceta.2010.02.006.

This work has been partially supported by the Ministerio de Economía y Competitividad grant MTM2013-41765-P. Ana Navarro Quiles acknowledges the doctorate scholarship granted by Programa de Ayudas de Investigación y Desarrollo (PAID), Universitat Politècnica de València.