

A MARKOV CHAIN APPROACH TO MODEL RECONSTRUCTION

VALERIA SCAPINI¹ & EDUARDO ZUÑIGA²

¹ Centro de Investigación en Innovación, Desarrollo Económico y Políticas Sociales (CIDEP), Universidad de Valparaíso, Chile.

² Complex Engineering Systems Institute (ISCI), Universidad de Chile, Chile.

ABSTRACT

Motivated by the fact that Chile is one of the most seismically active countries in the world (located over the ‘Pacific Ring of Fire’), we define a methodology for estimating the cost of housing reconstruction by modelling the occurrence of natural disasters as a Markov chain. Specifically, the states of the chain correspond to the different possible conditions of the housing infrastructure and the transition probabilities represent the possibility of change from one condition to another once the disaster has occurred. We prove that for the case of the 2010 Chilean earthquake, the matrix representing the process admits a stationary state vector. Using this vector, which we interpreted as the portion of time that the chain spends in each state in the long term, we define a cost function associated with total reconstruction. If this cost function is continuous, then this methodology allows policymakers to make decisions when facing the trade-off between current partial reconstruction and future total reconstruction.

Keywords: *earthquake, Markov chain, natural disasters, reconstruction cost.*

1 INTRODUCTION

Earthquakes generate significant losses or damage to housing infrastructure and significant costs associated with the process of rebuilding them. Chile is among the most seismically active places in the world, as the continental territory is located next to the subduction zone of the Nazca plate and next to the Antarctic plate [1] in the southernmost extreme of the country. Thus, it is important to understand the destructive power of this type of disaster, in order to contribute to decision makers and generate appropriate public policies related to reconstruction. These problems are discussed in this document.

We start with studying the 2010 Chilean earthquake motivated by several reasons. Firstly, Chile is a country where this happens regularly. Secondly, the Chilean earthquake that occurred on February 27, 2010 measured 8.8 on the magnitude scale of the moment, and is considered one of the strongest earthquakes in the history of humanity. It strongly hit six of the country’s 13 regions. The last reason is because we have data at the household level that allows us to calculate the level of damage generated by the earthquake. With an earthquake of these characteristics, constructions suffer damages that go from the category of slight to severe, with even cases of total destruction existing. This type of damage has been seen in both new and old constructions and in the entire area affected by the earthquake.

Considering all the above-mentioned facts, what we aim to do is to model the housing conditions as a Markov chain. More specifically, we model a chain whose states are the possible different conditions in which a house can be and consider that the transitions of the process occur because of natural disasters such as earthquakes. In the particular case covered in this article, we define and classify the possible states of the households based on a national government criterion, but the method is flexible in the sense that these states could be defined in different ways. In order to compute the transition probabilities, we use data collected in a national survey prior to and after the earthquake. In this setting, it is possible to argue that the states are positive recurrent and aperiodic, and also that the representative matrix is irreducible. Therefore, it is possible to

compute the stationary state of the process. We use these stationary probabilities, which can be interpreted as the proportion of time that the chain spends on each state on the long run, to analyse the housing reconstruction cost, and consequently, the trade-off faced by the policymaker between partial reconstruction (before the occurrence) versus full reconstruction (after the event).

This work constitutes an extension of the work found in Scapini and Zúñiga [2], which is a particular case of the application of the methodology described here. The present document is structured as follows. Section 2 provides a brief description of the Chilean case. Section 3 describes the data used in the paper. Section 4 explains the mathematical modelling used, calculations made and results obtained. Section 5 shows some extensions of the document and, finally, the conclusions of the study are highlighted in Section 6.

2 THE 2010 CHILE EARTHQUAKE

Earthquakes occur in areas of the earth's surface in which some tectonic plates limit with others. These plates move and collide with each other causing deformation of the rocks on both sides, releasing energy that on reaching the surface makes up the earthquake [3, 4]. Chile is considered one of the countries with the highest seismic activity in the world because it is located along the southwestern shore of South America. The area is known as the 'Pacific Ring of Fire', which corresponds to a zone where the tectonic plates are in permanent friction and concentrates some of the most important subduction zones in the world, where about 90% of the world's earthquakes and 80% of the largest ones occur [5–7]. For this reason, the Chilean experience is important to consider when we talk about disaster management.

The earthquake of February 27, 2010 is particularly interesting to study due to its magnitude, extension and number of affected households. It had a magnitude of 8.8 on the Richter scale, making it the second strongest earthquake in the country's history and the sixth strongest in the world [8]. It began at 3:34 a.m. and lasted nearly 3 min. It affected the central zone of the country over approximately 700 km, where about 80% of the country's population lives. It caused damage to about 500,000 houses. After the event, local seismic regulations were updated in order to improve construction safety [9].

Figures 1 and 2 show some emblematic images of the earthquake that occurred in Chile on February 27, 2010.

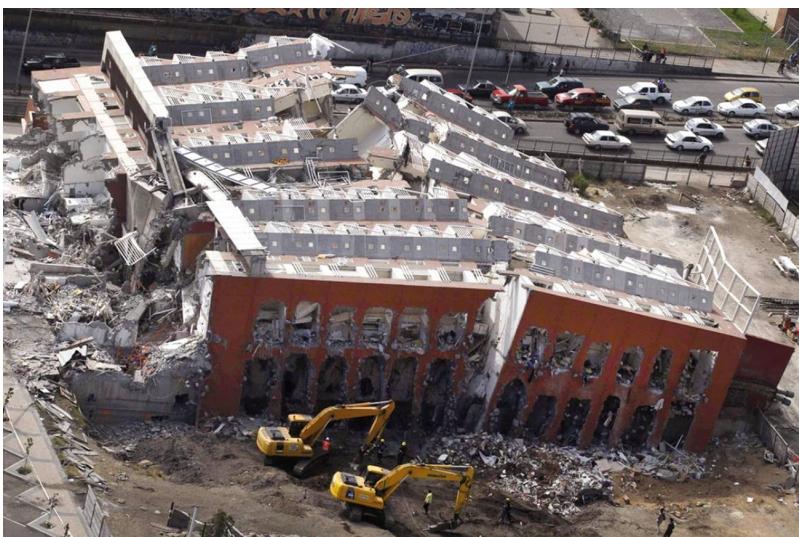


Figure 1: The most emblematic image was the fall of a building in the city of Concepción [10].



Figure 2: The earthquake prompted a major change in building regulations [11].

3 DATA: THE POST-EARTHQUAKE SURVEY

To carry out the study, we used the panel data from the 2010 National Post-Earthquake Socio-Economic Characterization Survey (EPT), which was conducted between May and June 2010 with the purpose of evaluating the main changes in the living standards of the people who were affected by the earthquake and tsunami. Data was collected from a subset of 22,456 households from the 2009 CASEN survey, making it a longitudinal survey that is both nationally and regionally representative. This survey was conducted by applying a questionnaire for each household, and the information was provided by the same person who responded to the 2009 CASEN survey or by another adult in the household if the former was not available. The questionnaire was completed through a direct interview which favours trust between the interviewer and the respondent and allows the interviewer to guide or explain to the respondent in case the questions are not understood.

The data panel contains information about the conditions of the households at two different points in time, approximately 3 months before and 3 months after the earthquake. Preliminary analysis of the data shows that before the earthquake, 78.8% of the homes were in good condition. This means that the floor, walls and roof were in good condition. On the other hand, we can see that 21.2% of the houses were in regular condition, which means that at least one of the structures referred to (floor, roof or walls) was in bad condition. Table 1 shows the number and percentage of houses according to the level of damage in the period prior to the occurrence of the disaster. It should be noted that in this period, the state of housing was good or regular and there were no destroyed houses.

The state of conservation of the housing in the period prior to the earthquake was determined according to the state of conservation of the walls, floors or roof of the housing obtained from the survey in 2009. If at least one of the three structures was found to be in poor condition, the house was considered to be in a regular state. Figure 3 shows the survey questions.

The state of conservation of the house in the subsequent period was determined based on a direct question, in which the state in which the house was left as a result of the earthquake was answered: no damage or minor damage, major damage or destroyed. Figure 4 shows the survey questions.

Table 1: Number and percentage with respect to the level of damage before the earthquake.

	No. of households	Percentage (%)
Good condition	17,441	78.8
Regular condition	4,693	21.2
Bad condition	0	0
Total	22,134	100

v12b. Estado de conservación de los muros

1. Bueno
2. Aceptable
3. Malo

Nº Alternativa 

v13b. Estado de conservación del piso

1. Bueno
2. Aceptable
3. Malo

Nº Alternativa 

v14b. Estado de conservación del techo

1. Bueno
2. Aceptable
3. Malo

Nº Alternativa 

Figure 3: 2009 CASEN survey housing module questions.

v4. Producto del terremoto/tsunami, la vivienda que habitaba junto a su grupo familiar:

1. Fue destruida por terremoto o quedó para ser demolida
2. Fue destruida por tsunami o quedó para ser demolida
3. Tiene daños mayores
4. Tiene daños menores
5. Sin daños o no fue afectada → **Pasa a v10**

Nº Alternativa 

Figure 4: 2010 Post-earthquake survey housing module question.

The data shows that, after the earthquake, 66.8% of the houses were in good condition, while 30% were in regular condition and 3.2% were in bad condition. It should be noted that after an earthquake occurs, the state of housing can be good, regular or bad, with the latter category being the destruction of housing as a result of the earthquake. Table 2 shows the number and percentage of dwellings according to the level of damage in the period following the occurrence of the disaster.

The percentage of houses by the level of damage shown in Tables 1 and 2 can be summarized in Fig. 5, in which the level of damage to homes is measured on the ordinate axis as a percentage and the columns show the period of time referred to, with the one on the left being the period before the earthquake and the one on the right the period after it. The graph shows that, after the disaster, the proportion of homes in regular condition and destroyed to the detriment of homes in good condition increases.

Table 2: Number and percentage with respect to the level of damage after the earthquake.

	No. of houses	Percentage (%)
Good condition	14,625	66.8
Regular condition	6,565	30.0
Bad condition	709	3.2
Total	21,899	100

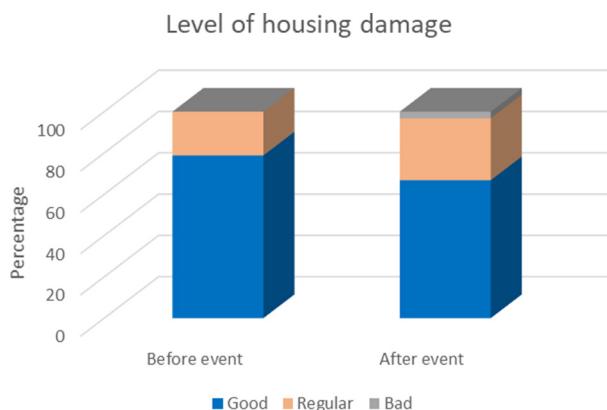


Figure 5: Percentage of homes by the level of damage at both points in time.

4 MATHEMATICAL MODELLING: MARKOV CHAIN APPROACH

As we have said before, we model the transition of one household condition to another as a Markov chain. Therefore, it is essential to properly characterize the chain, in order to define the states space and the transition probabilities. The set of states consists of the possible conditions in which the house can be found after the occurrence of the event. It is given by:

$$E = \{\text{good, regular, bad}\}.$$

The specific definitions of each state are the following:

- Good: not affected or minor damages
- Regular: major damages
- Bad: destroyed or in need to be demolished

This classification comes from the information obtained by the CASEN survey. Specifically, in one of the questions, people were asked to declare in which condition their house was.

We model the realizations of the chain as occurrences of the events, particularly an earthquake in the case covered in this article. Each realization of the chain corresponds to the occurrence of an event, in this case an earthquake. As it is well known, to fully characterize the process, we need to specify the transition probabilities between states. Given the realization definition that we are considering, we will define those probabilities using the data from the CASEN survey. This will be explained in the next subsection. Once we have these probabilities, we will be ready to compute the stationary distribution of the chain, which represents

the percentage of time that the chain spends in each state in the long term and will be used to estimate the reconstruction cost in different settings.

4.1 Transition probabilities

We compute the transition probabilities between states using the Bayes' rule. In particular, we consider the state of the household before and after the earthquake and then apply a favourable over total cases rule. Formally, let $N_{i,j}$ the number of households that went from state i to state j , N_i the number of households in state i prior to the event and N the total number of households in the sample. This is,

$$\sum_{i \in E} = N_i. \quad (1)$$

$$\sum_{i \in E} \sum_{j \in E} N_{i,j} = N \quad (2)$$

Therefore, the probability of going from state i to state j in any given realization is:

$$p_{i,j} = \frac{N_{i,j}}{N_i} \forall i, j \in E. \quad (3)$$

We aggregate these probabilities in a matrix $A = (p_{i,j})_{i,j \in E}$, which is the representative matrix of this Markov chain. Note that this is a stochastic matrix since, for every row i , we have that:

$$\sum_{j \in E} p_{i,j} = \sum_{j \in E} \frac{N_{i,j}}{N_i} = \frac{N_i}{N_i} = 1. \quad (4)$$

After some computations, it is possible to show that the transition probabilities are as indicated in Table 3.

Summarizing Table 4, the transition probabilities are as follows:

$$p_{0,0} = 0.70, \quad p_{0,1} = 0.27, \quad p_{0,2} = 0.03$$

$$p_{1,0} = 0.54, \quad p_{1,1} = 0.40, \quad p_{1,2} = 0.06$$

$$p_{2,0} = 0.01, \quad p_{2,1} = 0.00, \quad p_{2,2} = 0.99$$

Table 3: Number of houses with respect to the level of damage after the earthquake.

		After earthquake			
		Good	Regular	Bad	Total
Before earthquake	Good	12,125	4,697	435	17,257
	Regular	2,500	1,868	274	4,642
	Total	14,625	6,565	709	21,899

Table 4: Percentage of houses with respect to the level of damage after the earthquake.

		After earthquake			
		Good	Regular	Bad	Total
Before earthquake	Good	70.26	27.22	2.52	100
	Regular	53.86	40.24	5.90	100
	Total	66.78	29.98	3.24	100

where state 0 is Good, state 1 is Regular and state 2 is Bad. States and transitions between states can be represented by a directed network, in which nodes represent states and arcs represent transitions between states. The graph associated with the states of conservation of the houses and the probabilities of transition between each of the states is shown in Fig. 6,

Whose transition matrix is:

$$A = \begin{pmatrix} 0.70 & 0.27 & 0.03 \\ 0.54 & 0.40 & 0.06 \\ 0.01 & 0.00 & 0.99 \end{pmatrix}$$

In general terms, it can be seen that the highest probability of occurrence observed in transitions is from a state to itself (self-lapse), which indicates that the state of conservation of a house is more likely to remain than to move to another state. The only exception occurs in the regular state, where there is a greater probability of moving from the regular state to a good state than of remaining the same ($p_{1,0} > p_{1,1}$).

Note that the earthquake only admits transitions between the same state or to a lower condition (the earthquake cannot improve the state of conservation, it can only leave it the same or make it worse), so transitions to better states are due only to the effort of people.

It should be noted that we set $p_{2,1} > 0$ because we have the underlying assumption that in case of personal effort of reconstruction, this is fully achieved. We base this assumption on the fact that this case corresponds mostly to families with higher monetary resources, and it is, therefore, possible for them to fully reconstruct their houses. Also note that $p_{2,0} > 0$, which can appear counterintuitive at the first sight. The reason behind this fact is that we are considering that a small portion of the houses are repaired by their owners after the event. We will go deeper into this assumption in Section 5. In the next section, we tackle the problem of computing the stationary state of the chain, which will be the major tool for cost analysis.

4.2 Stationary state

In this subsection, we compute the stationary distribution of the process. In particular, we aim to find a vector $\pi \in R^{|E|}$, such that [11]:

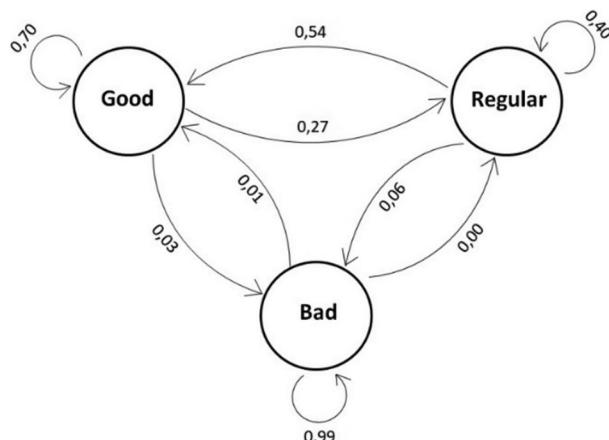


Figure 6: Representative graph of the states of conservation of homes and the probabilities of transition between each of them.

$$A\pi = \pi. \quad (5)$$

where A is the transition matrix. One could simply try to solve this linear system in order to find this vector π , but we prefer to go deeper into the theory since it has interesting interpretations in the context we are considering. In particular, to ensure the existence of this vector π , we consider the following well-known result from probability theory.

Theorem. If the Markov chain is irreducible and aperiodic, then there is a unique stationary distribution π .

Therefore, in order to properly look for this vector in our particular case, we need to argue why these conditions in the theorem are met:

- Matrix being irreducible means that all the states are communicated, and we assume this since a house can deteriorate or improve with the passage of time.
- Every state in E is positive recurrent and aperiodic, since it is always possible to come back to every state with no regularity in the number of steps (aperiodic).

In summary, we can assume that our chain satisfies these two conditions, and therefore, a unique vector π of stationary probabilities exists. Using the transition probabilities found previously, we can define the following linear system to find the stationary state vector:

$$\begin{aligned} 0.27\pi_2 + 0.03\pi_3 &= 0.3\pi_1 \\ 0.54\pi_1 + 0.06\pi_3 &= 0.6\pi_2 \\ 0.01\pi_1 &= 0.1\pi_3. \end{aligned}$$

Solving for π , we find the stationary state probabilities:

As we have said previously in the article, we subscribe the interpretation of these probabilities as the proportion of time that the chain spends in each state on the long run.

5 EXTENSIONS

5.1 Cost of reconstruction

In this section, we aim to exhibit how the previous development, related to the stationary probabilities of the process, might be useful for policymakers. In particular, we believe that this methodology could be useful to define and subsequently perform cost–benefit analysis based on the reconstruction cost as a function of these stationary probabilities. As we have said before, it is possible to interpret these probabilities as the proportion of time that the chain spends in each state on the long run, and therefore, they are useful to express a unit long-term cost of reconstruction. Let us suppose that the process is represented by a matrix $A \in \mathbb{R}^{N \times N}$ and admits the existence of a stationary vector $A \in \mathbb{R}^N$. We define the long-term cost of reconstruction as:

$$C(\cdot) : \mathbb{R}^N \rightarrow \mathbb{R}_+, \quad (6)$$

$$\pi \rightarrow C(\pi). \quad (7)$$

where $C(\pi)$ is defined as:

$$C(\pi) \doteq \sum_{i=1}^N \pi_i * v_i. \quad (8)$$

v_i is the reconstruction cost associated with each state, understood as the monetary value needed to return the household to its optimal condition, given that it is in state i after the event. This v_i might differ when considering different applied problems. For instance, in our previous article [2], we defined v_i as the average reconstruction cost (in dollars), obtained from the 2010 EPT CASEN survey.

Since natural disasters differ in their frequency (e.g. destructive earthquakes occur very distant in time, while hurricanes might happen every year), a policymaker could use this cost definition to perform cost–benefit analysis to face the trade-off between maintenance investments versus the reconstruction cost after the disaster. That is, sometimes it could be cheaper for the government to fully reconstruct the house after the disaster has occurred, while in other occasions, the best option could be to periodically maintain the households to prevent total destruction.

To be more precise about this latter case, let us suppose state N is that of the total destruction, and in a given time, \bar{C} is the maintenance cost needed to avoid total destruction in the occurrence of a disaster. If the cost of reconstruction after total destruction is considerably greater than that of the other states, this is given as:

$$v_N \gg v_i \quad \forall i \neq N.$$

Then, $C(\pi)$ could be big enough, so that $C(\pi) \gg \bar{C}$, leading the policymaker to avoid this situation in any circumstance, of course provided that the probability π_N is positive.

5.2 Extensions and sensitivity

As can be seen from the previous developments, the value of $C(\cdot)$ is extremely sensitive with respect to the stationary probabilities and the value v_i . In fact, in the previous subsection, we defined the cost function thinking of v_i as exogenously given, but, in general, $C(\cdot)$ could depend on both variables and not necessarily be linear. That is,

$$C(\cdot) : \mathbb{R}^{N \times N} \rightarrow \mathbb{R}_+, \quad (8)$$

$$(\pi, v) \rightarrow C(\pi, v) \quad (9)$$

When using this methodology, one could face the fact that π and v could be, at the same time, functions of some other relevant variable x . Therefore, in order to make a sensitivity analysis of how the cost changes, it would be necessary to consider both the direct and indirect effects:

$$\frac{dC(\pi, v)}{dx} = \frac{\partial C}{\partial x} + \sum_{i=1}^N \frac{\partial C}{\partial \pi_i} \cdot \frac{\partial \pi_i}{\partial x} + \frac{\partial C}{\partial v_i} \cdot \frac{\partial v_i}{\partial x}. \quad (10)$$

A future work path that we would like to explore is related to understanding which type of variables x are important when studying natural disasters and also how π and v depend on these variables. For instance, x could be the level of private investment made by the families to maintain their houses, and in that case, would be natural to consider π and v as decreasing functions of x . On the other hand, x could be the level of a ‘nature’-related variable that makes the total destruction easier, in which case π and v would be increasing functions.

6 CONCLUSIONS

In this work, a methodology to model the effect of a natural disaster on the conservation status of a house is presented. In particular, we model the occurrence of a disaster as a Markov

process whose states correspond to the different conditions of the housing. In this setting, we interpret the transition probabilities as the probability of worsening, maintaining or improving the condition of the housing after the occurrence of the disaster. We proved that when the chain defined in this way admits a stationary state vector, this can be used to define a cost associated with housing reconstruction. This cost function can be useful for policymakers when deciding whether to perform partial reconstructions before the event or total reconstructions after it, assuming that the cost function is continuous. In particular, it will always be beneficial for the policymaker to intervene before the event, as long as the intervention cost is lower than the expected cost of full reconstruction. This methodology can be extended to different types of disasters with a good understanding of what are the underlying variables that determine the probabilities of transition and the costs of reconstruction. This last point is essentially important when making sensitivity analysis.

REFERENCES

- [1] Cereceda, P., Errázuriz, A.M. y Lagos, M. Terremotos y Tsunamis en Chile. Origo Ediciones, Santiago, 2011.
- [2] Scapini V. & Zuñiga E., “Conditions of housing infrastructure before the occurrence of an earthquake” *WIT Transactions on Engineering Sciences*, Vol. 129, WIT Press, 2020, ISSN 1743-3533.
- [3] Pérez, E. & Thompson, P., Natural hazards: Causes and effects: Lesson 2—earthquakes. *Prehospital and Disaster Medicine*, **9(4)**, pp. 260–272, 1994. <https://doi.org/10.1017/s1049023x00041510>
- [4] Naghii, M., Public health impact and medical consequences of earthquakes. *Revista Panamericana de Salud Pública*, **18**, pp. 216–221, 2005. <https://doi.org/10.1590/s1020-49892005000800013>
- [5] Alvarez, G., Ramírez, J., Paredes, L. & Canales, M., Zonas oscuras en el sistema de alarma de advertencia de tsunami en Chile. *Ingeniare. Revista Chilena de Ingeniería*, **18(3)**, pp. 316–325, 2010. <https://doi.org/10.4067/s0718-33052010000300005>
- [6] Shabnam, N., Natural disasters and economic growth: A review. *International Journal of Disaster Risk Science*, **5(2)**, pp. 157–163, 2014. <https://doi.org/10.1007/s13753-014-0022-5>
- [7] U.S. Geological Survey Earthquakes FAQ.
- [8] 20 largest earthquakes in the world, USGS, <https://www.usgs.gov/natural-hazards/earthquake-hazards/science/20-largest-earthquakes-world>, Accessed on: 15 Mar. 2019.
- [9] OPS, El terremoto y tsunami del 27 de febrero en Chile: Crónicas y lecciones aprendidas en el sector salud, *Organización Panamericana de la Salud, Santiago de Chile*, 2010.
- [10] Mora, S. *10 años del 27-F: El terremoto que remeció y cambió a Chile para siempre*. [Photographer]. Recuperado de <https://www.24horas.cl/nacional/10-anos-del-27-f-el-terremoto-que-remecio-a-chile-y-para-el-que-nadie-estaba-preparado-3948765>
- [11] Rojas, J. *A 10 años del 27F: El terremoto que cambió la cultura chilena* [Photographer]. Recuperado de <https://www.duna.cl/noticias/2020/02/27/a-10-anos-del-27f-el-terremoto-que-cambio-la-cultura-chilena/>
- [12] Ross, S.M., *Stochastic Processes*. John Wiley & Sons: New York, 1996.