Application of a Discrete-time Markov Chain Simulation in Insurance

A. Fernández-Morales
University of Málaga, Málaga, Spain

Abstract—**This paper describes the application of an online interactive simulator of discrete-time Markov chains to an automobile insurance model. Based on the D3.js library, an interactive visual animation depicts the dynamics of individual policyholders in a *bonus-malus* system of automobile insurance. A survey was conducted among MSc students who used the simulator to obtain a preliminary assessment of the perceived usefulness in several dimensions of their learning process. The main findings indicate that flexible access via different devices was the most valued feature of this resource. In addition, the possibility of experimenting and simulating by means of controlling the main parameter of the model was also found to be particularly useful.**

Index Terms—**Markov chains, Simulation, Insurance, Animation.**

I. INTRODUCTION

Stochastic processes have a determinant role in numerous insurance and finance models. However, many students in these social science fields find it difficult to fully understand some abstract concepts that support this modeling approach. In recent years, there have been many attempts to overcome this problem in statistics education generally using interactive learning tools. It has been widely recognized that the effectiveness of the learning process is enhanced by means of experimenting and simulating, usually with digital tools [1], [2], [3], [4].

The rationale behind this approach consists of (i) letting the computer or the electronic device do the hard calculations while the results, which are presented in a visually appealing way, can be easily interpreted, and (ii) allowing experimentation with interactive controls that permit changing the initial conditions and obtaining the results in real time. With this approach, learners can concentrate their efforts on understanding the logic and reasoning driving the methods and leave the study of calculation techniques to a later stage. In addition, it facilitates a contextualized approach that encourages learners to explore and investigate with real-life or simulated data sets [5], [6].

The first developments in this line were desktop-based programs in the pre-Internet era. These programs used graphical capabilities to display, enter and manipulate data with a limited range of interactive features [7]. The next generation of interactive digital resources for statistics learning focused on some technologies that could be used in connection with Internet browsers, especially Java and Flash. Java applets embedded into web documents were one of the most frequent resources for this purpose [8]. In particular, [3] used Java applets for the central limit theorem, [9] for the confidence interval, and [9] and [10] for hypothesis testing.

All of these attempts to enhance the statistics learning process with interactive digital resources are widely considered as helpful instruments to gain a deeper understanding of the most difficult concepts. Furthermore, the possibility of experimenting and simulating without hard calculation efforts improves conceptual comprehension [11].

New web-based JavaScript libraries and frameworks have recently become available to develop powerful interactive data visualization for web browsers (i.e., D3.js, Cytoscape.js, jQuery, sigma.js, and many others), [12]. The D3.js front-end library is especially versatile and useful in this field. D3.js is an open source JavaScript library for manipulating documents based on data with many graphical capabilities. The library was developed by members of the Stanford Visualization Group, which are now at the University of Washington Data Lab. The interesting combination of the data-driven approach and powerful visualization components gives this framework an extraordinary flexibility [13]. It can also take advantage of the capabilities of CSS, HTML5 and SVG standards, thus avoiding proprietary representations. Moreover, because it is very efficient at supporting dynamic behaviors for interaction and animation, expressiveness and accessibility are improved [14].

Many plug-ins are available for creating several types of graphics, as well as complete libraries for charting built using D3.js, such as dimple, NVD3 or Crossfilter [12]. In this paper, we use a library developed by SETOSA [15].

Built on the author's experience in designing interactive simulations for actuarial education [16], [17] and tourism studies [18], an interactive graphical simulator was designed to study the properties of discrete-time Markov chains in the field of insurance. According to these previous experiences, the simulator is fully contextualized for its application in insurance, focusing on the demands of these students.

In addition, as reported by [19], interactive graphical resources should be included in a consistent and rational pedagogical strategy to truly enhance the learning process. In the same line, [20] stated that performing tasks is the best way to gain a deep understanding. Thus, the graphical simulator is proposed as a complement to a self-contained group of activities carried out in the Stochastic Processes course of the MSc in Actuarial Science at the University of Malaga, Spain. This course is currently being taught by means of a blended strategy that makes use of several online resources.

The paper is structured as follows: the graphical simulator is described in section II; the results of a survey on students' perceptions of the simulator are reported in section III; and finally, the main findings are summarized in section IV.

II. THE INTERACTIVE SIMULATOR

The interactive graphical simulator produces a visual representation of a discrete-time Markov chain in a web page. Three versions are currently available that describe the gambler's ruin problem, a simplified fire insurance model and an automobile insurance *bonus-malus* system (BMS). The last one is described in this paper.

A. The BMS model

Bonus-malus systems are a common framework in automobile insurance in Europe. Although there exist many variations of this type of models [21], [22], [23], most of them can be modeled by means of discrete-time Markov chains [24]. The model behind the simulator is a simple version that contains the main features of a BMS. The insureds can be partitioned into five classes or tariff groups, $\{C_1, C_2, \ldots, C_5\}$, which define the state space of the chain. The associated premium scale is increasing ($b_1 < b_2 < \ldots < b_5$), where the entrance class C_5 is the one with the highest premium. The annual premium depends only on the class. In addition, every insured's class is revised yearly taking into account the previous year class and the number of claims reported in that period, thus ensuring the Markovian property [25]. The probability distribution of the yearly number of claims is the Poisson law with mean μ. Thus, the class X_n to which each policyholder belongs every year is represented by $\{X_n\}_{n \geq 0}$, a homogeneous Markov chain with finite state space.

The transition rules between states in this model are: (i) the policyholder moves one state lower after a year without claims; and (2) moves one state higher for every claim reported in a year (but with three or more claims goes directly to C_5). According to the approach of [24], the transformation rules of the model can be expressed by the following function $T_k(i)=j$ that indicates the transfer from state i to j depending on the number of claims, k, in the previous year:

$$T_0(i) = \begin{cases} i-1 & i > 1 \\ 1 & i = 1 \end{cases}$$
$$T_k(i) = \begin{cases} i+k & i \leq 5-k, k < 3 \\ 5 & otherwise \end{cases} \qquad (1)$$

Therefore, the transition probabilities of the model $p_{ij}(\mu)$, corresponding to the (i,j) elements of the transition matrix, can be calculated by means of:

$$p_{ij}(\mu) = \sum_{i=1}^{5} p_k(\mu) t_{ij}^{(k)}, \qquad (2)$$

where $p_k(\mu)$ stands for the probability of k claims in a Poisson model with mean μ, and $t_{ij}^{(k)} = 1$ if $T_k(i)=j$ and $t_{ij}^{(k)} = 0$ otherwise.

Although this model has some limitations (e.g., the fact that transition probabilities do not depend on time and therefore do not allow for improvements in driving behavior), it is a good starting point for studying dynamic modeling in insurance.

B. The graphical simulator

As mentioned in the previous section, the simulator was designed based on the developments of [15], but with some adaptations required for application to our field of study. Since the library provided by [15] is based on the D3.js framework, the graphical simulator runs directly in the Web browser and does not require any other plug-in or additional software (unlike other interactive statistical simulators that require Java or Flash technologies). This feature greatly improves the usability of the simulator [26], allowing it to be used in some mobile devices that do not incorporate Java or Flash.

The aim of the graphical simulator is to facilitate the learning process of our students, in particular those who are studying Markov chains for the first time. This topic requires the understanding of several abstract concepts, including the combination of probabilistic modeling with a dynamic scenario. It is commonly recognized that simulation and experimentation contribute to a deeper understanding of statistical concepts [6], and that interactivity greatly enhances the learning process, especially when using e-learning resources [27]. For these reasons, the simulator has been designed with a particular focus on interactivity, including a slider to control the value of parameter μ in (2), thus allowing the user to modify this structural parameter and obtain a real time response.

The structure of the simulator is shown in Fig. 1. It is embedded in a web page that follows the general design of the course and contains a brief explanation of the case study (the *bonus-malus* system in the case of Fig.1).

The sliders in the upper left corner control the speed of movement and parameter μ. The right panels show the probabilities of the number of claims in a year, $p_k(\mu)$, for k = 0, 1, 2 and 3+ (three and more), and the complete yearly transition matrix, whose elements are $p_{ij}(\mu)$, for $i,j = 1, 2, \ldots, 5$. All the elements in the right-hand panels are updated in real time to the changes in the left-hand panel.
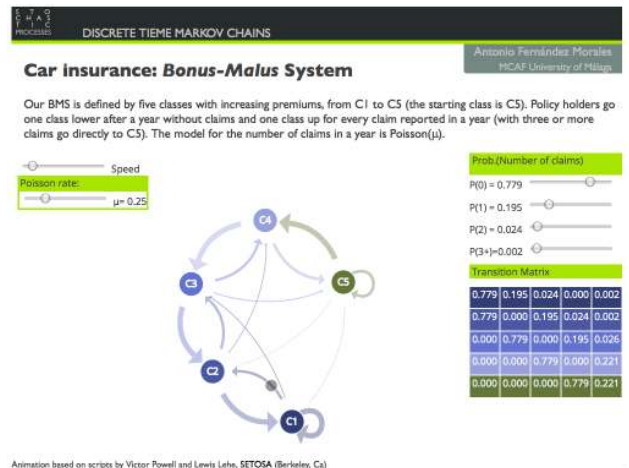

Figure 1. Web page containing the simulator

The dynamic graphical representation of the Markov chain is located in the central part of the screen. Each class is represented by a different colored disc, which matches the rows of the transition matrix. Classes are interconnected by arrows that indicate possible transitions. The width of the arrows is scaled depending on the magnitude of the probability associated to the transitions that are represented.

A gray disc indicates the state of the chain and its movement. It begins in class C_5 and then moves to subsequent classes following a random trajectory of the process that is simulated in real time, according to the model specified in (1), (2) and parameter µ. The slider on the upper left corner is used to control the speed of movement of the gray disc. The continuous movement shown in this graphical visualization of the process is a very interesting alternative to more traditional representations based on a static plot of several random trajectories.

Another remarkable feature of the simulator consists in the possibility of highlighting the arrow corresponding to one specific element of the transition matrix. By placing the cursor over any of the elements of the transition matrix, the corresponding arrow is automatically emphasized and the transparency of the others is increased (Fig. 2). This can be done without interfering in the movement of the chain and provides a very useful instrument to facilitate the interpretation of the model.

To make this simulator a highly useful instrument, several activities were designed that can be solved or explained using the simulator, in the same line as [20], [18]. The activities were contextualized in the field of insurance to enhance students' motivation.

III. RESULTS

As described above, the simulator and the related activities were used in the course on Stochastic Processes taught in the MSc in Actuarial Science at the University of Malaga. At the end of the last academic year, a survey was conducted to obtain a preliminary assessment of the students' perceptions about the simulator and the approach to its use.

The sample comprised 29 MSc students with an average age of 24.2 years and a demographic distribution by sex of 37.4% women and 62.6% men.
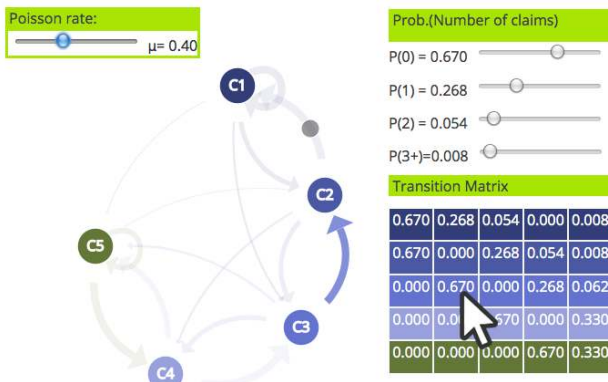
The survey consisted of two parts. The aim of the first part was to investigate students' perception about the usefulness of the simulator by means of three different dimensions: (i) for gaining a better understanding of concepts; (ii) for enhancing engagement in the learning process; and (iii) for autonomous learning.

The survey questions attempted to investigate several aspects of the simulator related to each dimension. The answers to the questions were coded on a 5-point Likert scale. The results of this part of the survey (average scores) are summarized in Fig. 3.

Five general elements were included in this part of the survey: visualizing the process dynamics, online availability, availability in different devices and interactivity. In addition, to gain deeper insight into the students' perceptions about the interactivity of the simulator, they were asked three specific questions related to the possibility of changing the speed of the animation and the parameter of the model in real time, as well as the highlighting effect of the cursor over the elements of the transition matrix.

Regarding the general elements of the simulator, the possibility of online access was the most highly valued of the three dimensions considered, with average scores over 4.75. However, other characteristics of the simulator were ranked differently depending on the dimension. Visualizing the dynamics of the process was the second element in importance (according to the average score) in the dimensions related to facilitating the understanding of concepts and engaging in the learning process, both with an average score of over 4.6. As regards the autonomous learning dimension, however, this element ranked fifth among the general elements and showed a lower average score of 4.15. These results seems to confirm the positive impact that animated visualizations can have in both improving comprehension and achieving a more engaging learning process [28], especially when learning topics with an intrinsically dynamic content.

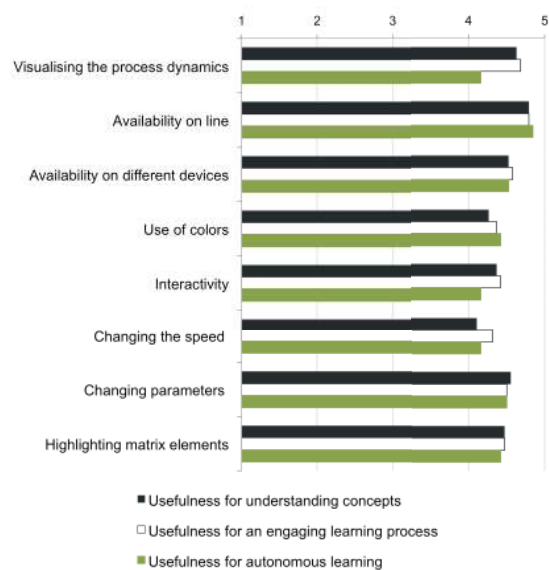Figure 2.   Emphasis on one element of the transition matrix

Figure 3.   Usefulness of the simulator (mean scores)

The availability of the resources in different devices were also scored highly, with averages above 4.5. This aspect was ranked second in the dimension of autonomous learning and third in the other two dimensions. These findings justify the importance of choosing an adequate technology that can satisfy students' demands, which have become increasingly diverse with respect to the devices they use to access Internet. In this case, the flexibility provided by D3.js-based interactive animations was a key factor in the positive responses of our students.

Interactivity, seen as a general feature of the simulator, obtained higher average scores in the dimensions related to the engaging learning process and understanding concepts than to autonomous learning. This indicates that our students' perception of the usefulness of interactivity in resources of this kind is more closely linked to the engaging features of the learning process and to the possibilities for a deeper understanding than to the potential benefits for a self-regulated learning process.

Moreover, the questions about specific features of the simulator's interactivity also yielded some interesting results. With average scores over 4.5 in the three dimensions, the user-driven parameter selection was considered the most useful element. The scores of this element are even higher than those corresponding to the general interactivity question.

The interactive effect of highlighting the transition matrix elements was ranked second, with slightly lower scores (averages above 4.2). The slider that controls the speed of animation was the element with the lowest average scores, although the scores can be considered quite satisfactory, since they are above 4 on a 5-point Likert scale.

Finally, the use of colors is a general feature that was more highly valued in the autonomous learning dimension. This feature ranked third among the general features of the simulator and last in the other two dimensions.

Therefore, the main outcomes of the analysis of the first part of the survey revealed that the usefulness of the interactive simulator and the related activities were highly valued by our students, with some differences in the three dimensions investigated.

The second part of the survey consisted of eight questions about quality, satisfaction in using the simulator and other digital resources, and intention of use. The questions are shown in the lower part of Fig. 4. The answers were also coded on a 5-point Likert scale.

Regarding the intention of use, questions P0, P1 and P2 obtained average scores above 4.4 (4.6, 4.6 and 4.4, respectively). This result suggests that our students were very satisfied with the experience and would therefore not only like to repeat it again and would recommend it to their classmates, but demand more resources of the same type.

In addition, the perceived quality of the simulator was also very high. The average score regarding the quality of the design (P6) was 4.5, while the average score for the quality of contents (P7) was 4.6. Furthermore, the average score was 4.5 for ease of use (P3).

Attitudes towards digital resources were also investigated in the survey, with some remarkable results.

P0: I would like to use more materials like the simulator
P1: I would recommend the simulator to my classmates
P2: I will use the simulator again
P3: The simulator has been easy to use
P4: I feel comfortable using digital learning materials
P5: I prefer traditional to digital learning materials
P6: The quality of the simulator's design is high
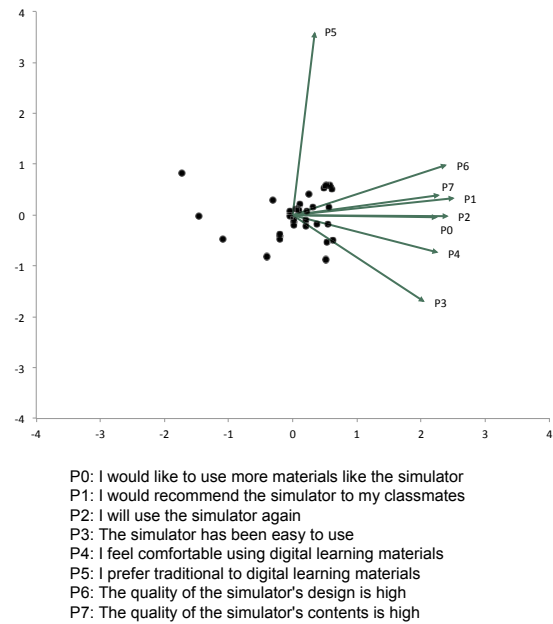P7: The quality of the simulator's contents is high

Figure 4. Biplot of attitudes, satisfaction and quality questions

While our students assigned a high average score to question P4 (4.6), revealing an overall perception of acceptance of digital technologies in learning, the average score of P5 (3.4) shows an intermediate level of preference for traditional resources to digital ones.

In order to get a visual representation of the structure of the relations between questions P0 to P7, Fig. 4 shows a biplot of the sample observations and the variables.

A close relationship was found between questions about the intention of use and recommendation (P0, P1, P2), perception of quality (P6, P7) and attitude towards digital resources (P4). The arrows of these variables are grouped together in the direction of the X-axis. In contrast, the variable representing the preference for traditional resources (P5) stands alone in the direction of the Y axis, showing low correlations with the variables in the group mentioned above. Finally, the question on ease of use (P3) is relatively far from the rest in a location that reflects a moderately negative correlation with the preference for traditional resources (P5).

IV. CONCLUSIONS

The technologies associated with some recently available JavaScript libraries that are contributing to significant developments in online interactive data visualization can undoubtedly be of great interest in the field of interactive resources for statistics learning. These resources are especially important in social science academic fields such as insurance or finance, where an extra degree of motivation is usually needed to engage learners when dealing with the most abstract or complex concepts behind models.

The main benefits of this approach, which is based on combining the capabilities of CSS, HTML5 and SVG standards in conjunction with modern browsers, are the flexibility and the independence from proprietary applications such as those based on Flash or Java. A very desirable outcome of this scheme is that it provides access

to the resources from a wide variety of devices, which is highly demanded by today's students.

The application described in this paper takes advantage of the D3.js library capabilities, particularly the SETOSA developments to build an online simulator of discrete Markov chains. The specific version that visualizes the evolution of a policy in an automobile insurance *bonus-mauls* system introduces another interactive feature that allows controlling the intensity rate of claims. Therefore, in addition to the excellent visual and dynamic characteristics of the SETOSA design, the technical properties of the underlying model, with a Poisson distribution governing the number of claims, are introduced in a fully interactive way.

A survey conducted among MSc students confirmed the good results of the experience presented in this paper. The general perception was that the simulator was highly useful not only for gaining a better understanding of the concepts, but also for a more engaging and autonomous learning process. Moreover, most of the students revealed their intention to recommend the simulator to other classmates and demanded more resources such as this one.

Internet access and availability on different devices were two of the most highly valued items in the survey. This fact reinforces the importance of using technologies like the D3 JavaScript library, which permit very flexible access from different browsers and mobile devices that today's students require.

Regarding the interactive features of the simulator, the students perceived being able to control the main parameter of the statistical model as the most useful. This is the key element that facilitates experimentation with different simulated claims intensities, and corroborates the premise that learners get more involved and gain deeper insight when experimenting with contextualized interactive resources.

Finally, the good results are an incentive to attempt to develop new versions of the simulator with more complex models and additional control parameters in a future line of work. However, before pursuing more complicated implementations, it would be recommendable to analyze the obstacles facing learners who are not fully comfortable with these technologies.

REFERENCES

[1] G. W. Sosa, D. E. Berger, A. T. Saw and J. C. Mary, J. C., "Effectiveness of computer-assisted instruction in statistics: A meta-analysis", *Review of Educational Research*, 81(1), pp. 97–127, 2011.

[2] J. A. González, L. Jover, E. Cobo and P. Muñoz, "A web-based learning tool improves student performance in statistics: A randomized masked trial", *Computers & Education*, 55(2), pp. 704–713, 2010.

[3] I. D. Dinov, N. Christou and J. Sanchez, "Central limit theorem: New SOCR applet and demonstration activity", *Journal of Statistics Education*, 16(2), pp. 1–15, 2008.

[4] M. Lunsford, G. H. Rowell and T. Goodson-Espy, "Classroom research: Assessment of student understanding of sampling distributions of means and the central limit theorem in post-calculus probability and statistics classes", *Journal of Statistics Education,* 14(3), pp. 1–19, 2006.

[5] D. U. Bolliger and S. Supanakorn, "Learning styles and student perceptions of the use of interactive online tutorials", *British Journal of Educational Technology*, 42(3), pp. 470–481, 2011.

[6] B. Chance, D. Ben-Zvi, J. Garfield and E. Medina, "The role of technology in improving student learning of statistics", *Technology Innovations in Statistics Education*, 1(1), pp. 1–25, 2007.

[7] P. Wang, B. K. Vaughn and M. Liu, "The impact of animation interactivity on novices' learning of introductory statistics", *Computers & Education*, 56(1), pp. 300–311, 2011.

[8] J. Byrne, C. Heavey and P.J. Byrne, "A review of web-based simulation and supporting tools", *Simulation Modeling Practice and Theory*, 18(3), pp. 253–276, 2010.

[9] R. W. West and R. T. Ogden, R. T., "Interactive demonstrations for statistics education on the world wide web", *Journal of Statistics Education*, 6(3), pp. 1–9, 1998.

[10] K. Schneiter, K., "Two applets for teaching hypothesis testing", *Journal of Statistics Education*, 16(3), pp. 1–8, 2008.

[11] L. P. Rieber, S. Tzeng and K. Tribble, "Discovery learning, representation, and explanation within a computer-based simulation: Finding the right mix", *Learning and Instruction*, 14(3), pp. 307–323, 2004.

[12] R. Wang, Y. Perez-Riverol, H. Hermjakob and J. A. Vizcaíno, "Open source libraries and frameworks for biological data visualisation: A guide for developers", *Proteomics*, 15, pp. 1356–1374, 2015.

[13] A. David and C. J. M. Tauro, "Web 3D data visualization of spatio temporal data using data driven document (D3js)", International *Journal of Computer Applications*, vol. 111, no. 4, pp. 42-46, 2015.

[14] M. Bostock, V. Ogievetsky and J. Heer, "D3: Data-Driven Documents", *IEEE Transactions on Visualization and Computer Graphics*, vol. 17, no. 12, pp. 2301-2309, 2011.

[15] V. Powell and L. Lehe, "Markov Chains. A visual explanation", http://setosa.io/blog/2014/07/26/markov-chains, accessed 11 June 2015.

[16] A. Fernández-Morales, "Technological resources and non-presential activities for an improvement in the learning process of the subject 'actuarial statistics'", *@Tic Revista d'Innovació Educativa*, 5, pp. 46–56, 2010.

[17] A. Fernández-Morales, "Learning survival models with on-line simulation activities in the Actuarial Science degree". *International Journal of Emerging Technologies in Learning*, 6(1), pp. 15–19, 2011.

[18] A. Fernández-Morales, "Simulating seasonal concentration in tourism series", *Journal of Hospitality, Leisure, Sport & Tourism Education*, 15, pp. 116–123, 2015.

[19] R. C. delMas, J. Garfield and B. Chance, "A model of classroom research in action: Developing simulation activities to improve students' statistical reasoning" *Journal of Statistics Education*, 7(3), pp. 1–19, 1999.

[20] S. Cairncross and M. Mannion, "Interactive multimedia and learning: Realizing the benefits", *Innovations in Education and Teaching International*, 38(2), pp. 156–164, 2001.

[21] R. Mahmoudvand and H. Hassani, "Generalized Bonus-Malus Systems with a frequency and a severity component on an individual basis in automobile insurance", *ASTIN Bulletin*, 39(1), no. 1, pp. 307-315, 2013.

[22] G. Tzougas, S. Vrontos and N. Frangos, "Optimal Bonus-Malus Systems using finite mixture models", *ASTIN Bulletin*, 44(2), pp. 417-444, 2014.

[23] C.I. Tan, J. Li, J. Siu-Hang and U. Balasooriya, "Optimal relativities and transition rules of a Bonus-Malus System", *Insurance: Mathematics and Economics*, 61, pp. 255-263, 2015.

[24] J. Lemaire, *Bonus-Malus Systems in Automobile Insurance*. Boston, MA: Kluwer Academic Publishers 1995.

[25] S. Pitrebois, M. Denuit and J.F. Walhin, ""Bonus-Malus scales in segmented tariffs: Gilde and Sundt's work revisited", *Discussion Paper*, 407, Institut de Statistique, Université Catholique de Louvain, 2003.

[26] A. Zucker, R. Kay and C. Staudt, "Helping students make sense of graphs: An experimental Trial of SmartGraphs software", *Journal of Science Education and Technology*, 23(3), pp. 441–457, 2014.

[27] A. Littlejohn, I. Falconer and L. McGill, "Characterizing effective eLearning resources", *Computers & Education*, 50(3), pp. 757–771, 2008.

[28] Y. Xie, "Animation: An R Package for creating animations and demonstrating Statistical methods", *Journal of Statistical Software*, 53(1), pp. 1-27, 2013.

AUTHOR

A. Fernández-Morales is with the Department of Applied Economics (Statistics and Econometrics), University of Málaga, 29071 Málaga, Spain (e-mail: afdez@uma.es). He is Academic Coordinator of the MSc in Actuarial Science at the University of Málaga, Spain.