MARKOV CHAIN MODELS IN PRACTICE: A REVIEW OF LOW COST SOFTWARE OPTIONS

Jiaru Bai*[1], Cristina del Campo**[2], L. Robin Keller*[3]

*   Paul Merage School of Business, University of California, Irvine, 92697-3125, USA.

** Corresponding author, Facultad de Ciencias Económicas y Empresariales, Campus de Somosaguas, Universidad Complutense de Madrid, 28223 Pozuelo de Alarcón (Madrid), SPAIN

ABSTRACT

Markov processes (or Markov chains) are used for modeling a phenomenon in which changes over time of a random variable comprise a sequence of values in the future, each of which depends only on the immediately preceding state, not on other past states. A Markov process (PM) is completely characterized by specifying the finite set S of possible states and the stationary probabilities (i.e. time-invariant) of transition between these states. The software most used in medical applications is produced by TreeAge, since it offers many advantages to the user. But, the cost of the Treeage software is relatively high. Therefore in this article two software alternatives are presented: Sto Tree and the zero cost add-in package "markovchain" implemented in R. An example of a cost-effectiveness analysis of two possible treatments for advanced cervical cancer, previously conducted with the Treeage software, is re-analyzed with these two low cost software packages.

You can find a Spanish version of this paper in the following link: http://faculty.sites.uci.edu/lrkeller/publications/

KEYWORDS: cervical cancer treatments, cost-effectiveness analysis, Markov decision trees, stationary transition probabilities

MSC: Primary 60J10; Secondary 90B99

RESUMEN

Procesos de Markov (o cadenas de Markov) se utilizan para el modelización de un fenómeno en el que cambia con el tiempo de una variable aleatoria comprenden una secuencia de valores en el futuro, cada uno de los que depende sólo del estado inmediatamente anterior, no en otros estados pasados. Un proceso de Markov (PM) se caracteriza por completo mediante la especificación del conjunto S finito de estados posibles y las probabilidades estacionarias (es decir, invariante en el tiempo) de transición entre estos estados. El software más utilizado en aplicaciones médicas es producido por TreeAge, ya que ofrece muchas ventajas para el usuario. Pero, el costo del software TreeAge es relativamente alta. Por lo tanto, en este artículo se presentan dos alternativas de software: Sto Tree y el coste cero de complemento de "markovchain" paquete implementado en R. Un ejemplo de un análisis de coste-efectividad de dos tratamientos posibles para el cáncer de cuello uterino avanzado, llevaron a cabo previamente con el TreeAge software, se volvió a analizar con estos dos paquetes de software de bajo coste.

Puede encontrar la versión en español de este artículo en el siguiente enlace: http://faculty.sites.uci.edu/lrkeller/publications/

1. INTRODUCTION

Markov processes (or Markov chains) are well known tools for modeling a wide range of phenomena in which changes over time of a random variable comprise a sequence of values in the future, each of which depends only on the immediately preceding state, not on other past states. A Markov process (PM) is completely characterized by specifying the finite set S of possible states and the stationary probabilities (i.e. time-invariant) of transition between these states (define $P_{ss'} = P(S_{t+1} = s'/S_t = s)$). Markov processes have applications in modeling and analysis of a wide range of phenomena in all branches of science, ranging from linguistics and biology to political science, medicine, economics, computer science, and many others.

Now if each change of state is assigned a numerical quantity, we have a Markov model with rewards. During the sixties and seventies, Markov processes with rewards (PMR) were applied in engineering-related (Howard [6]) fields, whereas in other fields, such as the economy, applications occurred a bit later (see, for example Norberg [9] for a PMR application to life insurance). A PMR is simply a Markov process in which in each state a different reward ($R_s = E[R_{t+1}/S_t = s]$) is received. The broader definition also includes a possible temporal discount factor ($\gamma \in [0, 1]$), since thereby an infinite return on Markov processes is avoided. If the reward is financial, the immediate rewards can be set to be larger than delayed future rewards. If the model does not discount, then $\gamma = 1$.

However, Markov processes, with or without rewards, cannot be used to represent problems in which each of the states of the Markov process has a large number of embedded decisions, which often occurs in situations requiring sequential decisions where the decision-maker can choose from several options, each with a potential

[1] JiaruB@uci.edu

[2] campocc@ccee.ucm.es

[3] LRKeller@uci.edu

reward and different transition probabilities. This problem is modeled as a Markov decision process (MDP), defined as a PMR where, in addition to the finite set S of states, there is a finite set of alternatives or actions (A), whose transition probabilities, $P_{ss',a} = P(S_{t+1} = s'/S_t = s, A = a)$, also depend on the proposed action. Obviously, the reward function, $R_{s,a} = E[R_{t+1}/S_t = s, A = a]$, also depends on the considered alternative.

For a more in-depth study of powerful and systematic methods for analyzing stochastic systems such as Markov processes, see, for example, Gouberman and Siegle [3], Rubio-Terrés [10] or Soto Álvarez [11].

The software most used in medical applications is by TreeAge, since it offers many advantages including a user-friendly integrated graphical user interface, Markov process modeling capability, graphs and reports for cost-effectiveness analysis, the ability to perform sensitivity analyses and display results graphically, Monte-Carlo simulation and other features. For a detailed description of the TreeAge Pro Healthcare software, see the documentation and the trial at www.treeage.com.

But a basic problem is the price of the TreeAge software, and the need to renew annually the license (for annual licenses) or the maintenance (for the standard perpetual license), which, in the medium to long term, results in an amount that may seem excessive. Therefore in this article two software alternatives are demonstrated to compare treatments for cervical cancer in a cost-effectiveness analysis: the rather low cost Sto Tree program or using the zero cost "markovchain" package implemented in R.

2. BACKGROUND

Information technology and decision support have evolved rapidly in recent decades. Billions of dollars invested in technology research have stimulated improvements in computer processing power, hardware, software, networks and telecommunications. The rapid evolution of these technologies has advanced so far that most users now have personal computers, either at work, at home or both, and a high proportion of them use packages of integrated applications, such as for example, Microsoft Office for Windows, LibreOffice or OpenOffice suite for Linux or Mac own suite, in day to day life.

Of course, the idea of reviewing the software programs that are used to support decision-making is not new, but based on what the authors found in their review of the literature, Menn and Holle [7] have published so far the only existing software review to help in making health decisions based on Markov models that implement and evaluate the cost-effectiveness of different medical treatments. There are some other assessments, such as Bhargava, Sridhar and Herrick [1], but they refer to decision support systems in general, without focusing solely on Markov models. However the review is mainly based on the comparison of these programs with respect to ease of use and implementation of a model, runtime, the characteristics of the presentation of results, and flexibility, while in this Article the authors are interested in three aspects:
- Ease of learning to use the program - many of the available systems are too complex for users to know and use them successfully in the short time that is usually available.
- Ability to examine the logic of the system - A user who uses a program wants it to be transparent in its operation. It is difficult to understand what is happening unless you can understand the logic of the operations performed by the system.
- Cost - with an always limited budget when investing in decision support software, it should be free or at a reduced cost to fit within the available budget.
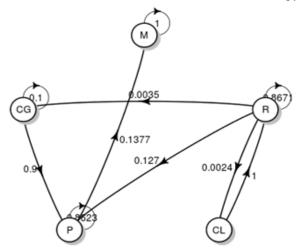
There seems to not be a wide range of support systems for medical and health decisions proposed in the abundant literature since the vast majority of the studies reviewed by the authors of this article use TreeAge. This makes it even more necessary to review the options now available, among which are the two presented in this article, with the added feature of their low or null cost.

3. THE EXAMPLE

Markov chain models were introduced in the medical literature by Beck and Pauker [2] and provide a modeling format that has been used in numerous subsequent analyses to aid choices between risky medical treatments with costs and health outcomes occurring not only in the present but also in the near or distant future. Specifically, Minion et al. [8] performed an analysis of the cost-effectiveness of chemotherapy alone compared with chemotherapy combined with a new drug called Bevacizumab (B), with the brand name Avastin[4], in patients with persistent or metastatic cervical cancer. This analysis is modeled by a Markov process with rewards (monthly health utilities), in which five health states are possible: response (R), limited complications (CL), progress (P), severe complications (CG) and death (M), see Figure 1. This is the example that the authors use to illustrate how to implement a Markov process with rewards in the two low cost software packages.

---

[4] Avastin is given as an infusion with the same schedule as the chemotherapy.

Figure 1. Diagram of Transitions in the Markov Chain for Chemotherapy (Created with R)

In this model, a patient who is in a state for a month can either remain in the same state or transition to a different state, with a certain probability and a certain cost (different for each of the two treatments under study), in the following month (Table 1). (In private communication, the Minion et al. authors pointed out typographical errors in their originally published [8] Table 2: In their transitions probability table, the probability of going from "severe complications" to "progress" in one month is 0.9 (not 0), and the probability of going from "severe complications" to "death" is 0 (not 0.9), for both treatment options. We are using the corrected probabilities here.)

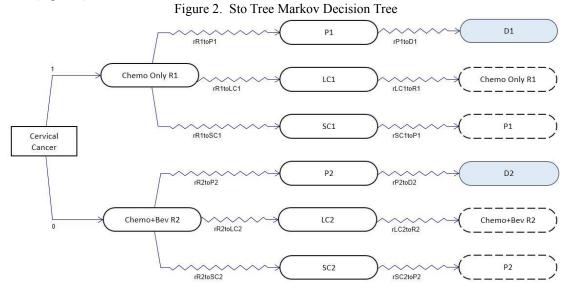Table 1. Transition Probability from State i to State j in One Cycle

	From i to j	Response R	Limited Complications CL	Progress P	Severe Complications CG	Death M
Chemo	R	0.8671	0.0024	0.1270	0.0035	0
	CL	1	0	0	0	0
	P	0	0	0.8623	0	0.1377
	CG	0	0	0.9	0.1	0
	M	0	0	0	0	1
Chemo + B	R	0.8720	0.0273	0.0823	0.0184	0
	CL	1	0	0	0	0
	P	0	0	0.8771	0	0.1229
	CG	0	0	0.9	0.1	0
	M	0	0	0	0	1

In the Markov model, the patient experiences a health "reward" or "utility" in each month, representing the effectiveness of the treatment that depends on the health state during that month. The patient's overall effectiveness is the sum of these utilities over all months. The utilities were assumed in Minion et al. [8] to be 1 for the state Response, indicating that the patient lived one month in the health state of responding to treatment for advanced cervical cancer, 0.75 for the state Limited Complications and 0.5 for the states Progress and Severe complications, as the life quality is adjusted downward for that month, and finally 0 for the state Die.

4. STO TREE SOFTWARE

Another option to make a Markov decision tree model is to use the Sto Tree software by Gordon Hazen ([5], http://users.iems.northwestern.edu/~hazen/StoTreeOv.html), which is a macro that is embedded in Microsoft Excel, so that other features of Excel are also available at the same time. The Sto Tree software only costs $20. Regarding the specific use of the software (user manual in [13]), first note that Sto Tree does not generate a

Markov transition diagram (as in Figure 1), but instead uses a Markov tree similar to that created in the TreeAge software (Figure 2).

Figure 2. Sto Tree Markov Decision Tree

The construction of a Markov model in Sto Tree is in a freer style than in the Treeage software. There is no standard formulation, Sto Tree nodes and arcs can be dragged anywhere on the worksheet, to suit the user. It is easy to build a Markov model in the software by simply adding nodes and arcs, no coding work is needed. The only thing to note is that there are two types of Sto Tree arcs: a wavy arrow represents a transition with a certain speed; while the straight arrow shows an instantaneous transition. The software is best suited to continuous Markov models, and is somewhat less efficient for discrete Markov models. (Although any probability p can be converted at a rate r using the formula $p = 1 - e^{-r\Delta s}$, where Δs is the cycle time.) The software does not consider time duration for those transitions with a straight arrow as they are seen as something that happens instantly.

Whenever there is a transition from a previous pre-existing condition, you have to select a duplicate node to represent this transition. For example, the state of "limited complications" (CL) can return to the "response" (R) state, so a duplicate node R (with a dotted circle in Figure 2) is created after the CL state. Unlike the TreeAge software, wherein the terminal node must be one of the states listed in the Markov tree, the Sto Tree software can select any state that is already created.

The parameters can be defined using the name of the function defined in Excel, making it easy to change the values of variables in the named cells and also allowing sensitivity analysis to be conducted manually in this way. Automatic sensitivity analysis is available in Sto Tree, also allowing probabilistic sensitivity analysis.

The cell in the last column is a logical value that would show 'True' if this row has all time counts in an initial chance or decision node. In our case (Figure 3), the chance/decision node is 'Cervical Cancer ', but at time 0 all counts are at 'Chemo Only R1', since the straight arrow takes 0 time. So the cell shows 'False'.

The Sto Tree software does not allow optimization for Markov processes, which means that a decision node is treated just as random nodes are treated. You can, though, indicate the chosen decision action by assigning a probability of 1 to the selected branch and 0 to the other decision branches. For example, in Figure 2, the patient can choose between two types of treatment: chemotherapy alone and chemotherapy plus Bevacizumab. We assign the first choice the probability 1 and the second 0, meaning that the branch with chemotherapy alone is active now and that allows us to obtain the results for the first treatment option. And the same could be done with the chemotherapy plus Bevacizumab option. However, the software does not compare the results of the two options and therefore does not indicate which is the optimal alternative, that has to be done by the user.

Finally, you can perform a cohort analysis once the stochastic tree is complete, as shown in Figure 3. Just specify the cycle time and the time horizon, and, where applicable, the discount rate, since the software also accepts a discount rate as in any PMR. The quality adjusted life years (QALY) and costs can be calculated once you have entered the parameters of the model in the table. In the figure, each cycle is one month, so the calculation is of "Quality Adjusted Cervical Cancer Life Months". The spreadsheet is simple and its implementation presents no difficulties.

The user can click on "cycle one step ahead" (go forward one cycle) to check the calculations in each step, or you can use "calculate sheet" to get the overall results for the specified time horizon. However, the software does not provide a detailed output of the analysis of the cohort, but the user has to record the result of each step and summarize the incremental and cumulative results accordingly.

And as a last detail, an unlicensed version lets you create stochastic trees, but you cannot do any of the computations. However, if a stochastic tree has been created in a licensed version, this model can be shared and used on an unlicensed version.

Figure 3. Sto Tree Spreadsheet for Calculating Costs and Quality Adjusted Cervical Cancer Life Months (Health Utilities) over 60 Monthly Treatment Cycles

		Cervical Cancer:	Cervical Cancer	Chemo Only R1	P1	LC1	SC1		
Cycle Time =	1	Chance/Decision?	1	0	0	0	0		
Starting Time =	0								
Time Horizon =	60								
Discount Rate =	0	Initial Cohort =	1						
Calculate Sheet									
		Time	Counts						
Reset to Start Time		0	0	1	0	0	0	0	FALSE
Step ahead one cycle		1	0	0.8671	0.127	0.0024	0.0035	0	
		Quality Coefficients:		1	0.5	0.75	0.5		
		Cost Coefficients:		524	262	809	4157		
		Incremental QALY Tolls:	0	0	0	0	0		
		Incremental QALYs:	0	1	0	0	0		
		Incremental Total:	0	1	0	0	0		
		Cum Discounted QALYs:	0	1	0	0	0		
		Overall Cum QALYs =	1						
		Incremental Cost Tolls:	0	0	0	0	0		
		Incremental Costs:	0	524	0	0	0		
		Incremental Total:	0	524	0	0	0		
		Cum Discounted Costs:	0	524	0	0	0		
		Overall Cum Costs =	524						

5. R: THE FREE OPTION TO CONSIDER

The three authors of this article are all teachers in a school of Administration and Management. Specifically, they are teaching students earning degrees in Administration, Management or Economics, and teaching these students a wide range of statistical techniques, both basic and advanced. In the modern era, the statistical package used most often is R. The advantages of this program are that it includes all the techniques needed, is open source and available without cost in all platforms (Windows, Linux or Mac), is easy to install and use, and can even be run from a CD or USB.

For those who have not heard of R, or those that have actually heard of it, but feel frightened at the thought of having to use it by writing code, we recommend that you install it (https://www.r-project.org/) and download some of the manuals that exist (e.g. in Spanish: https://cran.r-project.org/doc/contrib/R-intro-1.1.0-espanol.1.pdf). But above all we recommend that one of the graphical user interfaces specially designed to work with R be used. An interface provides the user access to a selection of analysis techniques using commonly used simple menus (equivalent to those installed, for example, when you have SPSS). The authors recommend R-commander (commonly known as Rcmdr) as, in our opinion, the most used and perhaps most documented. See, for example, the http://www.uv.es/innovamide/l4u/R/R0/R0.wiki web page, in Spanish, on how to install R for Windows and how to install Rcmdr and see, for example, Guisande González, Vaamonde Liste and Barreiro Felpeto [4] on how to use it for various analyses.

Now back to the problem at hand on Markov processes. In this case the R package to be used (markovchain) is not yet incorporated into Rcmdr. So, we explain in detail how to use (see also Spedicato, Kang y Yalamanchi [12]) , first, the code to repeat the analysis in this article, and second, to adapt it to any similar problem that the reader may face. In the code below, the hashtag/pound sign # indicates that a comment follows on that line. Lines without the # sign are the actual code.

This code below analyzes the chemotherapy only treatment, "chemo". To calculate the results of chemotherapy treatment with Bevacizumab, the same code would be used except "chemo" would be replaced with "chemoplusbeva", and the transition probabilities and costs corresponding to the second option of chemotherapy plus Bevacizumab would replace the numbers for chemotherapy alone.

```
# Install the library. Once installed, do not re-execute the statement
install.packages ("markovchain")

# Load the library for use in the current session.
```

```
# This statement must be executed at the beginning of each session
library(markovchain, pos=4)

# Create the Markov chain using the "Markov chain" package
# Specify the names of the states of the chain (in quotes)
# After, enumerate transition probabilities, by rows (byrow = TRUE)
# Indicate dimension of the matrix (nrow = 5)
# And finally, optional, the name you want to give the Markov chain
chemo <- new("markovchain", states = c("R", "CL", "P", "CG", "M"),transitionMatrix = matrix(data = c(0.8671,
+0.0024, 0.1270, 0.0035, 0, 1, 0, 0, 0, 0, 0, 0, 0.8623, 0, 0.1377, 0, 0, 0.9, 0.1, 0, 0, 0, 0, 0, 1), byrow = TRUE,
+nrow = 5), name = "Chemo")

# Verify that the transition matrix is correct, because it is printed in matrix format
chemo

# Draw a Markov diagram
plot(chemo)

# Draw another nicer Markov diagram nicer in black and white (in particular the one in Figure 1)
# For this you need to install another library. Once installed do not re-execute the sentence
install.packages("diagram")

# Load the library, to use it in this session
library(diagram, pos = 4)

# And now ask R to plot the graph
plot(chemo, package = "diagram", box.size = 0.04)

# Save the graphic in PNG format to a location on your computer
dev.copy(png,'C:/investigacion_operacional/figura_1.png')
dev.off ()

# Calculation of the fundamental matrix
Q_chemo = matrix (data = c(0.8671, 0.0024, 0.1270, 0.0035, 1, 0, 0, 0, 0, 0, 0.8623, 0, 0, 0, 0.9, 0.1), byrow =
+TRUE, nrow = 4)
I4 = diag(4)
F_chemo = solve(I4-Q_chemo)
F_chemo

# Expected time for absorption, whereas response begins in the response state
t_chemo F_chemo% *% = c(1,0,0,0)
cat( "Expected Lifetime =") + print(t_chemo)

# Calculate the cost of treatment, as does the TreeAge software
# Costs are per month in each state of the chain
cost = cbind(c(524,262,809,4157,0))
utility = c(1,0.75,0.5,0.5,0)
x <- c(1,0,0,0,0)
cost0 = x%*%cost
Cost = 0
for (i in seq(1:60)) {
x = x*chemo
Rewardeff = utility*x
Rewardcost = cost0%*%Rewardeff
Cost = sum(Rewardcost) + Cost
}
cat ("Cost of treatment =") + print(Cost)
```

6. CONCLUSION

Sto Tree and R with the markovchain package are good alternatives to the relatively expensive TreeAge software for Markov decision tree cost-effectiveness analysis.

REFERENCES

[1] BHARGAVA, H. K., SRIDHAR, S., and HERRICK, C. (1999): Beyond spreadsheets: Tools for building decision support systems. **IEEE Computer,** 3 (3), 31-39.

[2] BECK, J.R. and PAUKER, S.G. (1983): The Markov process in medical prognosis. **Medical Decision Making**, 3(4), 419-58.

[3] GOUBERMAN, A., and SIEGLE, M. (2014): Markov reward models and Markov decision processes in discrete and continuous time: Performance evaluation and optimization. In: Remke, A. and Stoelinga, M. (Eds.) **Stochastic Model Checking. Rigorous Dependability Analysis Using Model Checking Techniques for Stochastic Systems**, 156-241, Springer, Berlin Heidelberg.

[4] GUISANDE GONZÁLEZ, C., VAAMONDE LISTE, A. and BARREIRO FELPETO, A. (2011): **Tratamiento de datos con R, Statistica y SPSS**, Ediciones Díaz de Santos, Madrid.

[5] HAZEN, G.B. (2002): Stochastic trees and the StoTree modeling environment: Models and software for medical decision analysis, **Journal of Medical Systems**, 26(5), 399-413.

[6] HOWARD, R. A. (1971): **Dynamic Probabilistic Systems**, Vol. 1-2, Wiley, New York.

[7] MENN, P. and HOLLE, R. (2009): Comparing three software tools for implementing Markov models for health economic evaluations, **Pharmaeconomics**, 27(9), 745-753.

[8] MINION, L. E., BAI, J., MONK, B. J., KELLER, L. R., ESKANDER, R. N., FORDE, G. K., CHAN, J.K. and TEWARI, K. S. (2015): A Markov model to evaluate cost-effectiveness of antiangiogenesis therapy using bevacizumab in advanced cervical cancer, **Gynecologic oncology**, 137(3), 490-496.

[9] NORBERG, R. (1995): "Differential Equations for Moments of Present Values in Life Insurance," **Insurance: Mathematics and Economics**, 17(2), 171-180.

[10] RUBIO-TERRÉS, C. (2000): Introducción a la utilización de los modelos de Markov en el análisis farmaeconómico. **Revista Farmacia Hospitalaria**, 24(4), 241-247.

[11] SOTO ÁLVAREZ, J. (2012): **Evaluación económica de medicamentos y tecnologías sanitarias: Principios, métodos y aplicaciones en política sanitaria**, Springer Healthcare Iberica, Madrid.

[12] G.A. SPEDICATO, T.S. KANG y S.B. YALAMANCHI (2015): The markovchain Package: A Package for Easily Handling Discrete Markov Chains in R, Available at https://cran.r-project.org/web/packages/markovchain/vignettes/an_introduction_to_markovchain_package.pdf, The Comprehensive R Archive Network (CRAN), Last accessed 15-3, 2016.

[13] **StoTree v.7 manual** (2015), Available at http://users.iems.northwestern.edu/~hazen/Documentation StoTree 7 (3).pdf, Last accessed 15-3, 2016.