

COMMENTS ON “Rainfall in Little Havana” by Roqué

Jorge Luis Romeu

It is with much professional satisfaction that we discuss Dr. Diego Roqué’s Markov model, implemented in his paper “Rainfall in Little Havana.” For, as we have stated in previous ASCE papers (Romeu, 1996) there is much need for review, update and implementation of quantitative modeling methods in our analyses of the Cuban situation.

Roqué’s paper is a clear, clean and useful example of the use of Markov Chains in forecasting. It can be easily implemented with Cuban data, in the same manner Roqué has done in the Little Havana example. This work could even be broken down by provinces, or by geographical regions, and used in agriculture, construction and other outdoor activities that are substantially affected by rain, with obvious benefits for the Cuban economy.

However, Markov modeling can go much further. In addition to forecasting, such models can be used in the study and control of many time dependent processes, including socioeconomic and political processes. And it is in this direction that we would like to discuss and expand Roqué’s paper, in lieu of commenting on his neat math derivation.

Implementation of Markov Chains require the definition of: (i) two or more states; (ii) a transition mechanism defined by a (TPM) matrix of probabilities of change from one state to another; and (iii) a time domain with fixed increments (e.g., the system is observed by hours, days, weeks, months, etc.). In such case we obtain a “memoryless” process, where the future is dependent on the past only by way of

the present. This model allows the description of a process as a “black box” and provides, among other important performance measures: (i) the probability of being in a given state, at a given time in the future; (ii) the mean time to reach a given state; (iii) the probabilities (or mean times) of sojourning in each of the steady states; and (iv) the time to reach the system steady state. All of this Dr. Roqué has illustrated with states “rain” and “dry” in his “Rainfall in Little Havana” paper.

However, we could attempt to implement Markov models in socioeconomic contexts closer to the activities of ASCE. For example, we could define two states: one “unstable” (say dictatorial or revolutionary) and another “stable” (say democratic) for a country (say the Cuban republic during the XX Century). If we could also estimate the transition probabilities from one state to another, then we could, as Roqué did, use a Markov Chain approach. Then, we could forecast say, the probability of being in a given state on a given year (e.g., arriving at a democratic state in 2000) when starting in another given state (say, under a dictatorial regime) at some previous time (e.g., in Cuba, in 1990).

The main problem with modeling socioeconomic problems is, precisely, obtaining the data to estimate the transition (TPM) matrix. As seen from Roqué’s work, there are several years of daily rain data that allow the estimation of his TPM matrix. However, we usually do not find enough state transitions, in our mentioned socioeconomic context, to do likewise.

One alternative would be to pool data from similar countries. For example, we could assume that certain Latin American countries that share the same history and socioeconomic conditions, could be modeled in a single process by pooling their data. Unfortunately, if we pool too many countries together, however, their dissimilarities will introduce variability into the process, and the (TPM) transition estimations will suffer. Finally, we can refine this model by considering the time T , to the next transition, as a (second random exponential) variable, as opposed to the fixed daily, monthly, etc., time increments. This refined model is now an Embedded Markov Chain.

SOME USEFUL MODEL EXTENSIONS

Following the nomenclature and approach in Kalbfleisch and Prentice (1980), let us now consider T , the continuous and positive random variable “time to failure or transition.” We will denote its associated Survivor Function $P\{T>t\}$ as $F(t)$; its density function as $f(t)$ and its instantaneous failure or transition rate, at time $T=t$ (also called Hazard Function) as $\lambda(t)$. We can see that: $\lambda(t) = f(t)/F(t)$.

The Hazard Function $\lambda(t)$, which depends on the time $T=t$, can also be characterized as minus the derivative of the natural logarithm of the Survivor Function. This characterization makes the Hazard Function the centerpiece in the definition of both, the Survivor and the density functions of T . Therefore, we can now transfer our modeling efforts of variable “time to failure” T , into modeling its Hazard Function $\lambda(t)$.

First, notice that $\lambda(t)$ can be an increasing, decreasing or mixture (bathtub curve) function. In industrial reliability studies, for example, an increasing hazard occurs when failures tend to become more frequent as the (device or) process ages. Decreasing Hazard Functions occur when process failures become less frequent over time. Finally, the bathtub Hazard Function characterizes the entire process life cycle. It occurs because failures tend initially to be more frequent (infant mortality) then stabilize at a low level (useful life) and finally rapidly increase again (during the aging period).

Let us adapt this to a socioeconomic context, like the one in our Cuban example. An increasing Hazard Function can be justified in a new, revolutionary regime, whose emerging authority is initially challenged by many and can lose power. On the other hand, a decreasing hazard may be justified in a personal regime, legitimized by time and a well defined succession structure (e.g., monarchy), where authority is stable and widely accepted with passing time.

Finally, the well known and mathematically convenient constant hazard rate is obtained, in industrial applications, by curtailing the process life cycle—and hence modifying the bathtub curve. First, screen testing (weeding out infant mortality items) and then the implementation of an efficient replacement policy (items are taken out of service before they reach their aging process) leaves only the “useful life” period.

In the socioeconomic context, such bathtub hazard helps explain the entire political life cycle of a successfully consolidated dictatorship. At the onset, such regime faces strong opposition, which can bring it down, inducing a large hazard. With time, it entrenches itself (through force) crushing most of the opposition and stabilizes in power (lower, constant hazard). Finally, the hazard increases again as the dictator ages, the regime demoralizes, its traditional leaders either die or become incapacitated by old age, and a younger, better prepared generation of technocrats challenges the old guard to make changes.

Such life cycle system behavior can be currently observed in Mexico and in China. It was also characteristic of the military dictatorships of Salazar in Portugal, Porfirio Díaz in Mexico, and Stroessner in Paraguay. And it is also likely also occur in Cuba, with Castro.

Finally, political stability (constant hazard) in socioeconomic contexts results from preventing abrupt government changes, such as revolutions and military coups, as well as prolonged personal governments. It plays the same role as screening for infant mortality and good replacement policies for aging problems in the industrial setting.

In either of these two contexts (socioeconomic or industrial) constant hazard rate is what allows the application of Markov Models (which require constant transition rates to guarantee the “memoryless” property). A functional example of such processes is the Exponential model, which exhibits the classical constant Hazard Function (or failure rate).

Finally, the Two Parameter Weibull (λ , p) Model, constitutes a generalization of the Exponential. The Weibull Hazard Function is:

$$\lambda(t) = \lambda p (\lambda t)^{p-1}$$

If $p = 1$, the hazard $\lambda(t)$ is constant and the model reduces to the Exponential. However, if $p > 1$ the Hazard Function is increasing and if $p < 1$ it is decreasing in t .

REGRESSION MODELS

Let us now “open the Black Box” in the following way. Let $z = (z_1, \dots, z_p)$ be a vector of p “covariates” or process explanatory variables, and let the Hazard Function $\lambda(z)$ be:

$$\lambda(t; z) = \lambda(t) c(zb)$$

where b is also a vector, of p coefficients (i.e., weights) and the function $c(zb)$ either linear, inverse linear or exponential (i.e., $c(zb) = 1 + zb$ or $(1 + zb)^{-1}$ or $\exp(zb)$).

Now we have a Hazard Function $\lambda(t; z)$ that depends not only on the time t to failure (through $\lambda(t)$) but also on a set of device or process characteristics, defined by the vector of (explanatory) covariates z , weighted by the coefficients b . In the socioeconomic context, this would mean that T (i.e. time to transition) is now associated with a set of problem covariates that describe particular characteristics of the different countries included in the analysis, such as population, GDP, unemployment, etc.

The crux of the problem now becomes modeling this new relationship, akin to the problem of determining the parameters of a regression model. We usually select, for modeling convenience, the functional form:

$$\lambda(t; z) = \lambda(t) \exp\{zb\}$$

As a result, the conditional density function $f(t)$ of T (time to transition) is now:

$$f(t; z) = \lambda \exp\{zb\} \exp[-\lambda t \exp\{zb\}]$$

Using the transformation $Y = \text{Log}(T)$, we can eventually arrive to the following model, that expresses the log failure rate as a linear function of covariates z :

$$Y = -\text{Log}(\lambda) - zb + W$$

where W is a random variable (specifically, having extreme value distribution).

Skipping the mathematical details (that the reader can find in the reference given above) this result provides a regression model on covariates z , with coefficients b , which leads to the development of the Proportional Hazards Model (Cox, 1972).

PROPORTIONAL HAZARDS MODEL IN CUBAN SOCIOECONOMIC STUDIES

This model has been successfully and widely used in cancer studies, where it has served two very important goals. First, it has allowed the inclusion, in the model, of many different patients, with many different medical and physical conditions, thus increasing the pool of available data and leading to better estimates of the transition rates and the times to remission, to death, etc. Second, it has allowed the establishment of “risk factors,” parameters that affect (increase or decrease) the Hazard Function and consequently also the time to failure in the process (sometimes positively, other times negatively). These risk factor estimations provide (i) a relative weight for each factor analyzed and (ii) their statistical significance (or lack of significance). The latter, has proven useful in determining what are the prime (and the secondary) factors affecting the cancer process, providing a better understanding of it and ultimately some degree of control over its course.

In the Cuban socioeconomic context, we propose using this approach to study and understand the factors and subprocesses associated with the current Cuban situation. We can apply this modeling approach with certain economic and social factors known to effect the political stability. Then, we can analyze and quantify their contribution, sign and statistical signif-

icance, as described above, and use them to help redress the state of the nation.

To find enough data to implement this approach we first need to define a subset of appropriate Latin American countries. Then, a set of covariates such as GDP, inflation rate, unemployment, current account, etc. would be defined. Then, we also need to define the set of states (e.g., democracy, dictatorship) we are interested in studying, with the transition modes to go from one to the other (e.g., revolution, coup, foreign intervention, election). We then collect, from the countries selected, for the historical period included in the analysis, the time to state change (say in years or months) and the corresponding values of the covariates. Finally, we use the Proportional Hazards model with the above defined times and covariates and obtain estimates of their values, signs and statistical significance.

There is yet another statistical alternative to the above mentioned approach. It is also of the regression class, though not related to reliability modeling. It consists in implementing a Discriminant Analysis.

It may be possible to divide the Latin American countries into three groups. One group will be comprised of those countries considered unambiguously as “positive.” A second group is comprised of those considered unambiguously as “negative.” Finally, there is a third group composed of those countries for which we do not have a clear cut position or evaluation. We can again measure specific factors or variables (say, size, GDP, inflation, unemployment, etc.) and implement a Discriminant Analysis using them. Such approach would also yield the degree of influence, sign and statistical significance of the factors sought. But this is material for a future, separate paper for ASCE.

To conclude, we believe that Dr. Roqué, with his well developed example of the use of Markov Chains to study environmental problems, has provided an opportunity for ASCE researchers to review the wealth and potential of the Markov, Reliability and Proportional Hazards models, in the study of Cuban socioeconomic problems.

REFERENCES

- Cox, D. R. (1972). “Regression Models and Life Tables (with discussion).” *Journal of the Royal Statistical Society, Series B*. Vol. 34, pp. 187-220.
- Kalbfleisch, J. and R. Prentice (1980). *Statistical Analysis of Failure Time Data*. New York: Wiley.
- Romeu, J. L. (1996). “Comments on the Future Phases of the Cuban Economy.” *Cuba in Transition—Volume 6*, pp. 317-319. Washington: Association for the Study of the Cuban Economy.