A Markov Chain Model for Covid-19 Survival Analysis

Jorge Luis Romeu, Ph.D.
https://www.researchgate.net/profile/Jorge_Romeu
http://web.cortland.edu/romeu/; Email: romeu@cortland.edu

1.0 Introduction

The present *Markov Chain analysis* is intended to illustrate the power that Markov modeling techniques offer to Covid-19 studies. It is part of *our pro-bono collaboration to the American struggle against Covid-19*, whereby retired professionals would provide input, based on our long experience. You can read our *Proposal for Fighting Covid-19 and its Economic Fallout* in: https://www.researchgate.net/publication/341282217_A_Proposal_for_Fighting_Covid-19_and_its_Economic_Fallout

We have previously written *An Example of Survival Analysis Applied to Covid-19 Data*, found in https://www.researchgate.net/publication/342583500_An_Example_of_Survival_Analysis_Data_Applied_to_Covid-19, also *Multivariate Statistics in the Analysis of Covid-19 Data* and *More on Applying Multivariate Statistics to Covid-19 Data*, both of which can also be found in: https://www.researchgate.net/publication/341385856_Multivariate_Stats_PC_Discrimination_in_the_Analysis_of_Covid-19 and, as the already cited, also in our *ResearchGate* web page: https://www.researchgate.net/publication/342154667_More_on_Applying_Principal_Components_Discrimination_Analysis_to_Covid-19 These latter statistical methods provide useful tools for classification of states, regions, counties etc., according to levels of infection and other metrics.

In addition, we have written a tutorial on the use of *Design of Experiments (DOE) Applied to the Assessment Covid-19*. It provides an example of a tool for assessing and controlling appropriate levels of infection in states and regions. It can also be found in our ResearchGate web page: https://www.researchgate.net/publication/341532612_Example_of_a_DOE_Application_to_Coronavarius_Data_Analysis We have written *an evaluation of* the results of 25 years *off-shoring* tens of *thousands American jobs*, and the *impact* this has had *on US preparedness* to fight the Coronavarus Pandemic, found in: https://www.researchgate.net/publication/341685776_Off-Shoring_Taxpayers_and_the_Coronavarus_Pandemic And we have written a short study on the use of *reliability methods in the design and operation of ICU* units, that can be found in: https://www.researchgate.net/publication/342449617_Example_of_the_Design_and_Operation_of_an_ICU_using_Reliability_Principles

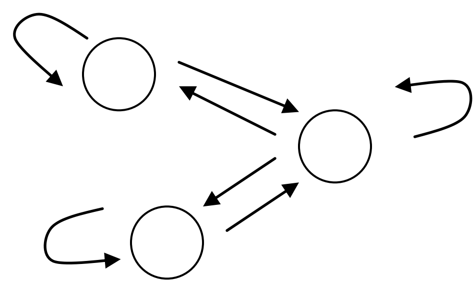
In this article we model the trajectory of Covid-19 infected patients into an ICU, and up to their death, using a Markov Chain. We start by considering a simple three-element state space. We then include additional states, to account for more complex situations.

Then, using such models, *we obtain* (1) **the probability of death** of a Patient; and **(2) their expected time to death,** using their sojourns in the different states.

2.0 A Simple Markov Chain

Let *X(T)* a *Markov Chain* over a *three-element state space*: (0) Non-infected, (1) Infected and (2) Hospitalized population. *Markov equations and a state diagram* for this model, are given below:

$$
\begin{aligned}
p_{01} &= P\{X(T)=1 \mid X(T-1)=0\} = q_{01} \\
p_{10} &= P\{X(T)=0 \mid X(T-1)=1\} = p_{10} \\
p_{12} &= P\{X(T)=2 \mid X(T-1)=1\} = q_{12} \\
p_{21} &= P\{X(T)=1 \mid X(T-1)=2\} = p_{21} \\
p_{ii} &= P\{X(T)=i \mid X(T-1)=i\} = 1-\sum_{j\neq i} p_{ij}
\end{aligned}
$$



Consider every *Markov Chain transition* X(T) at time T, as an independent trial corresponding to a transition from its current state **i** into its next state **j** = 0, 1, 2, having *probability of success* $\mathbf{p_{ij}}$

The *Transition Probability Matrix P* for this *Markov Chain* model is given below:

$$
P = \begin{array}{cccc} States & 0 & 1 & 2 \\ 0 & p_{00} & p_{01} & p_{02} \\ 1 & p_{10} & p_{11} & p_{12} \\ 2 & p_{20} & p_{21} & p_{22} \end{array} = \begin{array}{cccc} States & 0 & 1 & 2 \\ 0 & 1-q_{01} & q_{01} & 0 \\ 1 & p_{10} & 1-p_{10}-q_{12} & q_{12} \\ 2 & 0 & p_{21} & 1-p_{21} \end{array}
$$

Rows must add to one (probability is unit because the system is always in one of its three states). And, if we want to find the probability $p_{ij}^{(n)}$ of ending in some state **j** after 'n' steps, given that we started in some state **i** of the system, we *raise matrix P to the power 'n'* and look at entry $\mathbf{p_{ij}}$ of the resulting matrix P^n. In our present analysis we will model two settings as Markov Chains:

Matrix MatSimpGood

N.I.	**Inf.**	**Hosp**
0.95	0.05	0.0
0.10	0.70	0.2
0.00	0.30	0.7

This is an *efficient system* because *transition probability* From States 0 to 1 (infection) is 0.05, and probability of remaining in the Hospital (State 3) is 0.70; both smaller. As a result, X(T) will have the *steady state probabilities*:
$\boldsymbol{\pi = (\pi_0, \pi_1, \pi_2) = (0.545455 \quad 0.272727 \quad 0.181818)}$

Matrix MatSimpBad

N.I.	**Inf.**	**Hosp**
0.90	0.1	0.00
0.12	0.7	0.18
0.00	0.2	0.80

This is an *inefficient system* as the transition probability From States 0 to 1 (infection) is 0.1, and probability of remaining in the Hospital (State 3) is 0.80, both larger. As a result, X(T) will have the *steady state probabilities*:
$\boldsymbol{\pi = (\pi_0, \pi_1, \pi_2) = (0.387097 \quad 0.322581 \quad 0.290323)}$

Steady state distribution $\boldsymbol{\pi}$ represents the *Long-run percent* of cases in each of the system states, as well as the *rates* at which said Markov Chain X enters such states.

$$
\Pi = Limit_{T\to\infty}(\Pr ob\{X(T)=0\}, \Pr ob\{X(T)=1\}, \Pr ob\{X(T)=2\}) = (\Pi_1;\Pi_2;\Pi_3)
$$

The corresponding *Long-run Times* $\boldsymbol{T_i}$ for each state **i**, can be *interpreted* as the average *time between two successive visits* to said state **i**. They are computed as: $T_i = 1/\pi_i;\ i = 0,1,2$

For Efficient System: $T_0 = 1/\pi_0 = 1/0.545 = 1.834$; $T_1 = 1/\pi_1 = 3.667$; $T_2 = 1/\pi_2 = 5.50$;

One of the *most important uses of modeling activity* is to *compare performances for different states of affairs and courses of action*, by *using system* steady state *performance measurements*. *George Box*, an eminent statistician, *said* that *all models are wrong, but some models are useful*.

In the *inefficient case*, when *infection rates are 10%* instead of CDC suggested 5%, there is a *higher percent of patients hospitalized, or having a higher rate of entering the Hospital* (*29% instead of 18%).* This shows how, letting the *infection rate increase above the suggested upper bound of 5%, results in saturating the Health Care system* with too many patients.

A similar situation occurs with *Times between two successive visits to a state **i***. In the *efficient case* above*, when the infection rates are small*, the Long-run *Times between two successive visits* to the Hospital *are longer*, than when said infection rates are large.

Table #1: comparisons of the (above) systems performance measures:

Case (Rates)	**Long-run**	**Not Infected**	**Infected Home**	**Hospitalized**
Efficient (5%)	Probabilities	0.545	0.273	0.182
Efficient (5%)	Times Between	1.834	3.667	5.50
Inefficient (10%)	Probabilities	0.387	0.322	0.290
Inefficient (10%)	Times Between	2.583	3.099	3.444

Verify above how, when infection rates are kept at levels suggested by the Health Authorities (i.e. below 5%) the performance measures for the *Efficient System* are significantly better than the same performance measures for the *Inefficient System*, when infection rates increase to 10%.

In the next section we analyze the Covid-19 situation using a more complex Markov Chain. This new Chain is defined (1) over a larger (five element) state space, (2) includes an *Absorbing state*, and (3) patients will die (in the previous, simple model, every patient improves; no patient dies). These three features will also introduce several technical differences.

3.0 A More Complex Markov Chain

Define now a *Markov Chain over five states*: (1) *Non Infected* (in the General Population); (2) *Infected* (isolated at home); (3) *Hospitalized* (after becoming ill); (4) in the *ICU* (or ventilators); and (5) *Dead* (absorbing state; no possible return). Its *Transition Matrix P* is given below:

Pop;	Infect	Hosp	ICU	Dead	
0.93	0.07	0.00	0.00	0.00	(1); Non-Infected population
0.05	0.80	0.10	0.05	0.00	(2): Infected (isolated at home or hospital)
0.00	0.15	0.80	0.05	0.00	(3): Hospitalized (after becoming ill);
0.00	0.00	0.05	0.80	0.15	(4): In the hospital ICU (or ventilators);
0.00	0.00	0.00	0.00	1.00	(5) Dead (absorbing state; no return)

The *unit time* is, as before, *a day*. *Transitions* refer to changes observed from say *one morning to the following morning*. The daily probability of a person becoming infected is 7% (93% remain uninfected). No other transition is possible from this state. Daily recover probability of Infected persons, without hospitalization, is 5%; 10% get sick enough to need hospitalization (5% are so sick that they are placed directly in the ICU); and 80% remain infected at home. A 15% of those hospitalized, improve, and are sent home for further cure; 80% remain hospitalized, 5% become so sick that need to be interned in the ICU unit. A 5% of those in the ICU improve and are again returned to the general ward; 80 % remain for another day in the ICU; 15% die.

This is a more realistic model (also, more somber) and closer to situations in Italy or NYC at the height of their Covid-19 crisis. It provides more flexibility in the modeling activity. It also has a significant difference with the previous section Markov Chain. That one, modeled a *recurrent process*, where every state could, directly or indirectly, reach any other state (i.e. *nobody died*, but was able to recover). The present Markov Chain models a *non-recurrent process* that through a number of *Transient states*, eventually *leads to an Absorbing State (Death*) *of no return*.

Because the state space contains both transient and recurring (absorbing) states, there is no point in obtaining a steady state solution. Instead, we obtain (1) the *long run probabilities of dying*, as well as (2) the *expected times to die*, both of these starting at states 2, 3 and 4, respectively.

To achieve this, we obtain from P the sub-matrix of all transient states, by deleting the row and column corresponding to absorbing state. The remaining matrix, denoted "Q", is given below:

Matrix Q of Transient States

Pop;	Infect	Hosp	ICU	
0.93	0.07	0.00	0.00	(1) Non Infected (in the General Population);
0.05	0.80	0.10	0.05	(2) Infected (isolated at home);
0.00	0.15	0.80	0.05	(3) Hospitalized (after becoming ill);
0.00	0.00	0.05	0.80	(4) in the hospital ICU (or ventilators);

We then subtract matrix Q from the Identity Matrix, yielding (I-Q), and invert this latter one:

Matrix inverse $(I-Q)^{-1}$ of Transient States: Pop; Infect.; Hosp.; ICU;

Populat	Infectd	Hospital	ICU
26.1905	16.6667	10.0000	6.66667
11.9048	16.6667	10.0000	6.66667
9.5238	13.3333	13.3333	6.66667
2.3810	3.3333	3.3333	6.66667

The Potential matrix contains the *Long-Run Sojourns* (average number of visits) to each of the states of said matrix columns, when starting from the states of the matrix rows. For example, the *average number of days* an *uninfected* person *spends* isolated and *Infected at home*, when they were initially infected, is *16.67 days*. The number of *days spent in the hospital* after being sent, is 10 days. The *number of days* a person *spends in an ICU/Ventilator* unit, before passing away or dying is *6.67 days*. *Adding these numbers up*, we obtain *the average number of days* it takes for a person, initially infected, to be isolated at home, then hospitalized, and finally in the ICU and/or Ventilator, *before passing away (dying*). This average is: 16.67+10+6.67 = 33.34 days. We can do likewise with all other rows, and obtain the average times to death, from all transient states:

Table #2: average times to death from all transient states

Starting State for any Individual	**Average Time to Pass Away (Die)**
From Initial Time of Consideration	26.19 + 16.66 + 10.00 + 6.66 = 59.52 days
From the Time of Infection/Isolation	16.67 +10 + 6.67 = 33.34 days
From the Time of Hospitalization	=10.00 + 6.66 = 16.66 days
From the Time of entering ICU/Ventilator	= 6.66 days
Average Time, before becoming infected	26.19 days

We now calculate the probabilities of a patient dying, starting from any of the transient states. We calculate these probabilities for a horizon of two, four, eight or sixteen days (considering that the death has occurred after said person started from the considered transient state).

Results are obtained using the Markov Chain property:

$$P^m = P*P \ldots P \qquad \text{(m times, once for each day considered)}$$

Table #3: probabilities of a patient dying, starting from any of the transient states

Starting	State	Two Days	Four Days	Eight Days	Sixteen Days
From Initial	**(Healthy)**	0.000	0.002	0.018	0.098
From Time	**of Infection**	0.007	0.036	0.118	0.282
From Time	**Hospitalization**	0.007	0.038	0.127	0.307
From Time	**ICU/Ventilator**	0.270	0.444	0.636	0.780

We can see how the probability of dying in sixteen days or less, except for those in an ICU or in a Ventilator, is relatively low. It is worth mentioning that, using our model, in the long-run (in a very large number of days) the entire population is wiped out by Covid-19 (see matrix F below).

We now calculate *matrix F*, yielding the probability of a person ever reaching any Markov Chain state, especially the *absorbing state of dying*, given that such person starts in any of the previous transient states (healthy, infected, hospitalized, ICU). Notice that it is impossible to reach any of the transient states (probability zero) after having reached the Absorbing State of Death.

Table #4: *Matrix F* yields the probability of a person ever reaching any state from another

	Healthy	Infected	Hospitalized	ICU/Vent.	Death
Healthy	0.96	1.0	0.75	1.0	1.0
Infected	0.45	0.94	0.75	1.0	1.0
Hospitalized	0.36	0.8	0.92	1.0	1.0
ICU/Vent.	0.09	0.2	0.25	0.85	1.0
Death	0.0	0.0	0.0	0.0	1.0

For example, the *probability of eventually becoming Infected*, or going into an ICU/Ventilator, or Dying, *when starting Healthy*, is Unit (e.g. a *sure event*). On the other hand, the *probability of* going back to the Hospital ward, or *going back home* to isolation, or becoming Healthy again, *when starting from an ICU/Ventilator* Unit, is *very low* (at most 0.25); while the *probability of* staying on the ICU/Ventilator is 0.85, and that of *eventually Dying is Unit* (e.g. *the sure event*).

Matrix F shows the *importance of avoiding a Long term evolution of a High Infection rate*.

4.0 Mathematics

Markov Chains is an advanced statistical topic, with involved mathematical background. In addition, its calculations are not generally included in statistical packages, as it occurs with regression, principal components, etc. We need software that handles matrix algebra, including transposing, inverting, adding and multiplying matrices. We next give a summary explanation, and we remit the reader to the textbooks of Cinlar, Taylor, and others, in the Bibliography.

In Section 2.0 we discussed an *irreducible, recurrent Markov Chain* $\{X_n; n >= 0\}$ that illustrates a simple Coronavarus model (everyone gets cured; nobody dies). Such model provides the steady *Steady State distribution* $\boldsymbol{\pi}$, which allows (1) performance comparison of alternatives, and (2) the *calculation of the logistic needs of health care units*, based upon the number of infected patients.

Steady state probabilities $\boldsymbol{\pi}$ can be obtained (1) by *raising Transition Matrix P to the power 'n'*:

```
0.95  0.05  0.0           0.545455  0.272727  0.181818
0.10  0.70  0.2     =>    0.545455  0.272727  0.181818
0.00  0.30  0.7           0.545455  0.272727  0.181818
```

Or (2) by solving the linear equations: $\boldsymbol{\pi P = (\pi_0, \pi_1, \pi_2)P = \pi;}$ with $\boldsymbol{\Sigma\ \pi_i = \pi_0 + \pi_1 + \pi_2 = 1}$

In Section 3.0 we presented a *more complex* and realistic, (and also pessimistic) *Markov Chain* $\{X_n; n >= 0\}$. Said Chain is *not irreducible*; on the contrary, it has a set of *Transient States*, leading to an absorbing state (Death). We *cannot obtain Steady state probabilities* for such a model. But *we can calculate the absorption times* (*Deaths*) *and its probabilities* by subtracting matrix Q from the Identity, then inverting. The important section of *Potential Matrix R* is $(I-Q)^{-1} = S$, which provides the *Sojourns*, or average number of times that different Transient states are visited. Adding them up we get the *average time to Death*, starting at different Transient states.

Matrix S in turn provides the relevant part of *Matrix F*, of the *probabilities of ever reaching the absorption state (Death*) from every Transient state. Such probabilities are defined as *Zero*, if between Recurrent and Transient states (e.g. *impossible*, *to leave* the absorbing state of Death). They are defined as *Unit, the sure event*, if between Transient and Recurrent states, meaning that everyone will eventually die, independently of their initial state. Finally, we obtain probabilities between all Transient states (i,j), which are obtained through formulas based on Matrix S, of R:

$$F(i, j) = R(i,j) / R(j,j); \qquad\qquad F(j,j) = 1 - 1 / R(j,j);$$

All above-mentioned *performance measures* help calculate Logistics required for the health care facilities. Sometimes we forget that, as important as the care for sick patients may be, the basic Logistics that enable such care to take place are also very important, and need to be established.

For example, health care professionals require Protective Personal Equipment. How many of each? Where will we get them from? How will we store and distribute them? When will there be a vaccine, a treatment, or enough testing kits? Who will produce and distribute them? What is the supply chain? Where are the funds to support all these operations, coming from?

Stats and math models help us answer such questions, and evaluate different alternatives.

<u>5.0 Discussion</u>

Results from the present Markov Chain models complement those results obtained in previous Survival Analysis ones. Different models address different aspects of the Covid-19 problem.

Notice how, for the *simple model in Section 2.0*, keeping *the infection* rate below 5% was *key* to avoiding an overload of healthcare facilities. Thence, the importance of wearing masks, of social distancing, testing, and contact tracing, all of which contribute to *keep infection rates down*.

For the *more complex model of Section 3.0* we obtained *a crucial finding*: *prolonged and high infection rates* will eventually *get most everybody sick*. The *more vulnerable* population *will die*, and we will *reach heard immunization at a very high human cost*. Sweden tried this approach, at the start of the Pandemic, but later changed it because of the large number of deaths it produced.

Probabilities of becoming infected and eventually dying, starting from different states, are very pessimistic, as shown in Table #3 (*unit* probability *is the sure event*). *Section 3 model reflects the terrible situations arisen in Europe* (and in New York City) at the height of Covid-19 infections. Had these cities not implemented the drastic measures they put forth, which were able to reduce their infection rates, the Covid-19 results would have been catastrophic.

Probabilities of dying, starting from different states and considering different time horizons (in Table 4), as well as *times to absorption* (Death) starting *from different states*, help *estimate* the *usage of critical equipment* (e.g. ventilators) by patients, thus helping *estimate the quantities required* to provide adequate patient care, *as well as*, in the ultimate instance, *triage rules*.

Again, stats and math models help answer such questions, and compare different alternatives.

<u>6.0 Conclusions</u>

The Markov Chain models developed in this paper, as well as its Transient and Recurrent states, transition rates, etc. were all based on our readings about Covid-19. This author is not a public health specialist, but a statistician. In spite of our efforts, we were unable to obtain real data from any health care organization. The statistical models were built using our professional experience. Our intention is to place them in the hands of health specialists, so they can redo them with more accurate information.

In spite of all this, there are several results from our specific Covid-19 Markov Chain models that are useful. They have to do with steady state distributions, from the model in Section 2.0, and with probabilities and times to absorption, from the model in Section 3.0.

By changing the transition rates in the recurrent Markov Chain of Section 2.0 we can study how they affect the steady state distribution (or the percent population) in each state, and *compare* the efficient and inefficient *models*. This may help *establish the largest acceptable infection rate*. Times between Sojourns help *estimate* the *required sizes of the facilities* that will treat patients.

The model in Section 3.0 is more realistic. The probabilities of ever reaching every other state, and the number of Sojourns in them, may help *find logistic parameters* to ensure a sufficiently large and well-equipped health care facility.

Finally, *combining* the results from the present *Markov Chain Model* with those in our *previous Survival Model*, may provide doctors with more objective rules to establish Triage procedures, may such procedures ever become necessary.

Bibliography

Taylor, H. and S. Karlin. An Introduction to Stochastic Modeling. Academic Press. NY. 1993.

Cinlar, E. Introduction to Stochastic Processes. Prentice Hall. NJ. 1975.

Heyman, D. and M. Sobel. Handbooks in Operations Research and Management Science. *Vol. 2: Stochastic Models*. North Holland. Amsterdam. 1990.

Romeu, J. L. RAC START: Vol. 11, No. 6: *Availability*. https://web.cortland.edu/matresearch/AvailabilitySTART.pdf

Romeu, J. L. RAC Journal: First Quarter; 2000: *A Discussion of SW Reliability Modeling Problems*. https://web.cortland.edu/matresearch/SWRelArtRomeuRAC1ST_Q2000.pdf

Romeu, J. L. QR&CII Tutorial Series, Vol. 1, No. 2*: Infant Mortality in Reliability Studies for the Weibull*. https://web.cortland.edu/matresearch/InfantMortStudyQR&CII.pdf

About the Author:

Jorge Luis Romeu retired Emeritus from the State University of New York. He was, for sixteen years, a Research Professor at Syracuse University and is currently an Adjunct Professor of Statistics. Romeu worked for many years as a Senior Research Engineer at the Reliability Analysis Center (RAC), an Air Force Information and Analysis Center operated by IIT Research Institute (IITRI). Romeu received seven Fulbright assignments: in Mexico (3), the Dominican Republic (2), Ecuador, and Colombia. He holds a doctorate in Statistics/O.R., is a C. Stat Fellow of the Royal Statistical Society (RSS), a Member of the American Statistical Society (ASA) and of the American Society for Quality (ASQ). Romeu is a Past ASQ Regional Director, and holds Reliability and Quality ASQ Certifications. He created and directs the *Juarez Lincoln Marti Int'l. Education Project* (https://web.cortland.edu/matresearch/) which is dedicated to support higher education in Ibero-America. Readers can find all our work on Coronavarus (and other subjects) in our *Quality, Reliability and Continuous Improvement Institute* (QR&CII) web page:

https://web.cortland.edu/romeu/QR&CII.htm