ABSORBING MARKOV CHAIN FOR BREAST CANCER

Lalla Hanae Alaoui^{1 §}, Mohamed El Merouani²

^{1,2}Department of Mathematics and Statistics
University Abdelmalek Essaadi at Tetouan
Tetouan, 93000, MOROCCO

Abstract: Breast cancer is the most common type of cancer in women and ranks first in the world. Every year, more than one million new cases are diagnosed worldwide, which represents 30% of new cases of female cancer in industrialized countries and around 14% in developing countries.

The objective of our study was to detail an absorbing Markov chain model to study the evolution of breast cancer disease. We applied the results and the properties of the absorbing Markov chain that we accomplished in our last article, we consider a transition matrix estimated by the data of 780 patient, and then we created an algorithm to calculate the average expected duration to stay in each state and the probability of absorption using *Matlab* software.

AMS Subject Classification: 60J10

Key Words: breast cancer, absorbing Markov chain

1. Introduction

This study is essential firstly to inform the public and health professionals of the future evolution of this disease and then to help the decision-makers to provide the necessary infrastructure for the reception and the diagnostic and therapeutic support of patients. Finally, it will serve as a basis for comparison with the actual future evolution in order to evaluate the treatments and the prevention of the disease.

To determine the different stages of breast cancer disease, the TNM classifi-

Received: November 19, 2017

© 2018 Academic Publications

[§]Correspondence author

cation of the Union for International Cancer Control (UICC) and the American Joint Committee on Cancer (AJCC) were considered. TNM stands for “Tumor, Nodes and Metastasis”. It was the French surgeon Pierre Denoix who had the merit of proposing this classification. Its work began in 1943 to be finally adopted as a classification basis by the UICC Statistics Committee in 1953, [1].

In this manuscript, we limited our study to 4 stages of the disease varying from the less aggressive case to the most aggressive case, adding two states: Free of cancer and Dead. We leave open the possibility of adding other states and sub-reports for future research.

2. Breast Cancer Classification

Breast cancer is a malignant tumor that originates in breast cells, and that can spread to other parts of the body. To assess the extent of breast cancer, physicians consider three criteria: tumor size and infiltration, whether or not lymph nodes are affected, and whether or not metastasis is present.

2.1. The Size and Infiltration of the Tumor

When cancer cells appear, they first form a tumor in the channels or lobules of the breast. Then, gradually, the tumor can pass through the wall of the canal or the lobule and thus become infiltrating.

2.2. Whether the Lymph Nodes are Affected or Not

Cancer cells can escape from the breast and spread elsewhere, touching the lymph nodes of the armpit in the first rank.

2.3. The Presence or Absence of Metastases

Cancer cells can invade other organs than lymph nodes and develop metastases. The organs most commonly affected by metastasis in breast cancer are the liver, bones and lungs.

3. Mathematical Modelisation

The last classification, called TNM by the International Union Against Cancer, makes it possible to define the stages of breast cancer. It means “Tumor, Nodes,

and Metastasis”, [1]. Generally, these letters are associated with numbers ranging from 0 to 4 for T (tumor size), 0 to 3 for N (degree of invasion) and are either 0 or 1 for M (presence or absence of metastasis). The evolution of the tumor can be grouped by 4 stages in the following table: We limited our study to the 4 stages of the disease vary from the less aggressive case to the most aggressive case, adding two states: Free of cancer and Dead. Our objective is to calculate the probability of being absorbed (the absorption probabilities) and the mean time to absorption of breast cancer data. The data are collected at Dr. Anderson’s hospital and instituted from the University of Texas with observation of 780 patients. These data are analyzed at the beginning by McBride and al. to develop a stochastic model of mortality by breast cancer and estimate the proportion of patients cured. In another study, Toskos and Oguztoreli [2] worked with the same data to estimate the transition probabilities. In our study we will use these transition probabilities to construct a transition matrix to calculate the probability and mean time to absorption.

3.1. Transition Matrix

Let n_{ij} denote the number of patients who were in state i in period $t - 1$ and are in state j in period t . We can estimate the probability of an individual being in state j in period t given that they were in state i in period $t - 1$, denoted by P_{ij} , using the following formula:

$$P_{ij} = \frac{n_{ij}}{\sum_j n_{ij}}. \quad (*)$$

Thus, the probability of transition from any given state i is equal to the proportion of individuals that started in state i and ended in state j as a proportion of all individuals in that started in state i . This estimator $(*)$ is a maximum-likelihood estimator that is consistent by biased, with the bias bending toward zero as the sample size increases. Thus, it is possible to estimate a consistent transition matrix with a large enough sample, [3]. Toskos and Oguztoreli, for example, provide estimates of transition matrix.

	stage1	stage2	stage3	stage4	Free of cancer
stage1	0.9451	0.0062	0.0000	0.0000	0.0334
stage2	0.0000	0.9695	0.0116	0.0000	0.0176
stage3	0.0000	0.0000	0.1611	0.0160	0.0109
stage4	0.0000	0.0000	0.0000	0.1451	0.0030
Free of Cancer	0.0047	0.0000	0.0000	0.0000	0.9927

In this matrix we do not have 6th stage (Dead), we go step by step to calculate a transition matrix, we need to calculate $P_{16}, P_{26}, P_{36}, P_{46}, P_{56}$. The probability distribution for the Markov chain, is such that:

$$P(x_t | x_{t-1}, x_{t-2}, x_{t-3}, \dots, x_0) = P(x_t | x_{t-1}), \quad \forall t.$$

This probability distribution can be written as

$$P(x_0, \dots, x_t) = P(x_0) \Pi_1^t P(x_t | x_{t-1}).$$

If $x_t = \delta_j$ and $x_{t-1} = \delta_i$, then we can write

$$P(x_t = \delta_j | x_{t-1} = \delta_i) = P_{ij}(t) = P_{ij}, \quad \forall t.$$

This formulation assumes that the Markov process is stationary. Proceeding under this assumption, we can arrange the transition probabilities P_{ij} , into an (6×6) transition probability matrix $P = [P_{ij}]$, which has the following properties:

$$0 \leq P_{ij} \leq 1 \quad \sum_{j=1}^6 P_{ij} = 1 \quad \text{for } i = 1, 2, \dots, 6.$$

The summation condition above implies that the row sums must equal to one. Finally, the last row of the transition matrix is solved by using the following equation:

$$P_{iR} = 1 - \sum_{j=1}^{R-1} P_{ij}.$$

Then we can write

$$P = \begin{pmatrix} 0.9415 & 0.0062 & 0.0000 & 0.0000 & 0.0334 & 0.0189 \\ 0.0000 & 0.9695 & 0.0116 & 0.0000 & 0.0176 & 0.0013 \\ 0.0000 & 0.0000 & 0.1611 & 0.0160 & 0.0109 & 0.8120 \\ 0.0000 & 0.0000 & 0.0000 & 0.1451 & 0.0030 & 0.8519 \\ 0.0047 & 0.0000 & 0.0000 & 0.0000 & 0.9927 & 0.0026 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

3.2. Algorithm

The algorithm we have created is as follows:

```
% Let us give the absorbing Markov chain characterized by
the two matrices Q and R,
We compute three matrices describing the behavior of
the Markov chain.
```

```

%
% Entries:
Q
R
%
%
% The first matrix to compute is the fundamental matrix N.
% N (i, j) gives the expected number of visits in transient state j,
% before absorption, from transient state i.
% The fundamental matrix N depends only on the matrix Q.
%  $N = I + Q + Q^2 + Q^3 + Q^4 + \dots$ 
%
% We obtain the simple formula of the fundamental matrix N.
% By observing that, if we multiply both sides by (I-Q),
% We obtain the equation  $N * (I-Q) = I$ ,
% Then  $N = (I-Q)^{-1} = \text{inv}(I-Q)$ ,
% With inv is the function of the inverse of a matrix.
%
%
S = size (Q);
N = s (1);
I = eye (n);
N = inv (I - Q)
%
% Now we calculate the expected number of transition before absorption,
% Start with each transient state.
% Let W (i) be the expected number of transition before absorption,
% W is a column vector; it is formed by the sum of the rows of N.
%
M = ones (n, 1);
W = N * m
%
Let us now calculate the probability of absorption at each k states.
% Let B (i, j) be the probability of absorption at state j given that
% we have started from the transition state i.
% As
 $(I, j) + (Q^3 * R) (i, j) + \dots$ 
Then  $B = (I + Q + Q^2 + Q^3 + \dots) * R = N * R$ .
%

```

B = N * R
%

We then calculated the three matrices N, W and B.

3.3. The Results

N =

28.8412	5.8628	0.0811	0.0015	146.2151
10.8075	34.9838	0.4837	0.0091	134.5184
0.2425	0.0493	1.1927	0.0223	3.0185
0.0652	0.0132	0.0002	1.1697	0.8111
18.5690	3.7747	0.0522	0.0010	231.1248

W =

181.0017
180.8025
4.5254
2.0594
253.5216

B =

1.0000
1.0000
1.0000
1.0000
1.0000

4. Conclusion

We conclude that the average time of absorption is 181 months starting from the first stage, 180 months starting from the second stage, 4 months for the third, 2 months for the fourth and 253 for the last stage. From the results, we deduce the validity of the Markov model for estimating the average time of death starting from any stage. This model can be used in other situations more important and more delicate, as for example to compare two treatments tested

on two people who have the same conditions of the disease. Then, according to the results, the researchers can decide which is the most effective treatment. We can also go further by adding other stages of the disease, and even apply it to other types of cancer diseases.

References

- [1] American Joint Committee for Cancer Staging, *Manual for Staging of Cancer* (2002), 221-236.
- [2] C.P. Tsokos, M.N. Oguztoreli, A probabilistic model for breast cancer survival data, *J. Comput. Math. Applic.*, **14** (1987), 835-840.
- [3] M.T. Jones, Estimating Markov transition matrices using proportions data, In: *An Application to Credit Risk* (2005).