

Study Of Evolution Of The Incidence And Cancer Mortality Using The Markov Chain

Lalla Hanae Alaoui, Dr. Aziz Arbai

Department of Mathematics, laboratory of algebra and theory of operators, University Abdelmalek Essaadi, Faculty of Science, Tetouan, MoroccoCity;

Department of Mathematics, laboratory of algebra and theory of operators, University Abdelmalek Essaadi, Faculty of Science, Tetouan, MoroccoCity.

Email: lala-87@hotmail.fr

ABSTRACT: Markov chains are a method among methods we can use to describe the dynamics of a disease in its different stages. While the theory and practice of modeling of Markov chains Epidemiology is not new, rather the basic application of this powerful tool can be used by non-mathematicians as well. In this work, we present the above theory by announcing the assumptions and conditions of use of Markov chains in the study of diseases and specifically breast cancer, which could facilitate the formulation of hypotheses, and understanding the evolution of the disease in a population. We focus on absorbing Markov chains in discrete time, and we are going step by step to calculate the probability of infection with breast cancer for an individual, the duration of each state and the life expectancy for infected people. These quantities are useful for characterizing the dynamics of a disease and can help researchers to explore key aspects of the evolution of the disease into a population.

Keywords: Markov chains; breast cancer; Epidemiology

1 INTRODUCTION

Au Maroc, comme dans les autres pays du Monde, le cancer du sein représente l'une des principales cibles d'actions de la lutte contre le cancer. Les estimations marocaines les plus récentes classent ce cancer en première position parmi les cancers les plus fréquemment diagnostiqués chez les femmes (36 %); il représente la première cause de mortalité féminine. Entre 30 et 45 mille personnes sont affectées chaque année par le cancer de sein selon l'information officielle émanée des statistiques 2001 de l'institut National d'oncologie Sidi Mohammed Ben Abdellah de Rabat, et semble être le premier cancer féminin avec une incidence de 60 à 90/ 100000 femmes par an. Dans le Monde, ce chiffre atteint 540 mille cas par an et près de 300 mille femmes en meurent. Par conséquent, il représente un véritable problème de santé publique, car le diagnostic est souvent tardif et la prise en charge est difficile et couteuse. L'étude de l'évolution de l'incidence et de la mortalité par cette maladie est alors essentielle et utile à plusieurs titres:

- Informer le public et les professionnels de santé de l'évolution future de cette maladie ;
- Aider les décideurs à prévoir les infrastructures nécessaires à l'accueil et à la prise en charge diagnostique et thérapeutique des patients.
- Servir de base de comparaison avec l'évolution réelle future afin d'évaluer les traitements et les actions de prévention de la maladie.

Il existe deux types de modèles pour traiter ce sujet et chacun apporte des réponses à des questions bien précises. Ils ne nécessitent pas le même degré de connaissance préalable du système pour pouvoir être appliqués de façon fructueuse. Ils sont très complémentaires et représentent un outil d'étude et d'analyse très puissant. Les modèles déterministes: Il s'agit d'établir un système d'équations, pour le quel les paramètres à l'instant initial sont déterminés par des observations disponibles et les calculs doivent pouvoir se faire d'une manière proche du temps réel. Les modèles statistiques et probabilistes: Il s'agit ici de créer un système

dans les comportements sont du même type que le système réel. Pour autant, ils ne doivent pas coïncider exactement dans le temps, mais en convergence. On s'attache ici à faire un modèle numérique dans les caractéristiques globales vont tendre en moyenne vers celle du système réel. Ce type de modèle est plutôt dédié à la simulation. On utilise dans ce cas des outils issus des mathématiques probabilistes. Ce dernier modèle est le type de modélisation que nous avons choisi, et parmi le vaste champ des mathématiques probabilistes, nous avons retenu un outil qui se répond de manière grandissante dans la modélisation faisant intervenir les phénomènes biologique ; c'est l'outil markovienne. Ils permettent d'énoncer des relations et des probabilités. Très grossièrement, on pourra dire qu'ils permettent de définir l'état d'un système à un instant t uniquement à partir de son état présent.

2 Objectif

L'étude de l'évolution de l'incidence et de la mortalité par la maladie du cancer du sein au Maroc en appliquant les chaînes de Markov absorbantes à temps discret, en calculons:

- La probabilité d'infection par le cancer de sein pour un individu ;
- La durée prévue de chaque état,
- L'espérance de vie pour les personnes malades.

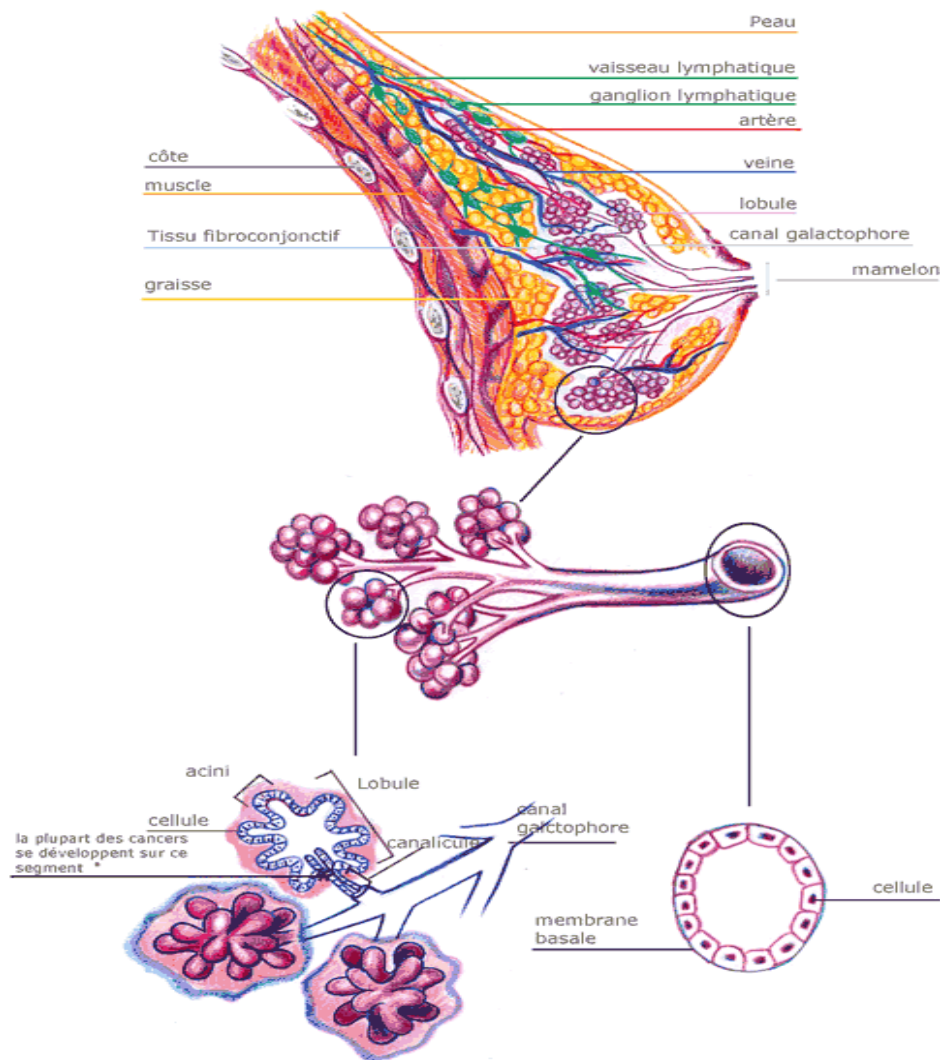
3 METHODE ET MODELISATION

3.1 Anatomies du sein

Le sein est une glande destinée à la production de lait. Il est essentiellement composé de graisse, de tissu conjonctif ainsi que de canaux et de lobes. Il existe entre quinze et vingt lobes dans le sein, qui se jettent dans des canaux galactophores, ces derniers venant aboutir au mamelon par cinq à neuf orifices. Chaque lobe est constitué de ce que l'on appelle l'unité terminale ducto-lobulaire qui est en quelque sorte la structure de base dans la production du lait. En effet, les lobules (constitués eux-mêmes d'acini) se

jettent dans un canalicule extra-lobulaire, l'ensemble de ces derniers rejoignant à la manière de petits ruisseaux les canaux galactophores que l'on pourrait comparer à des rivières. Au sein de toutes ces structures circulent les

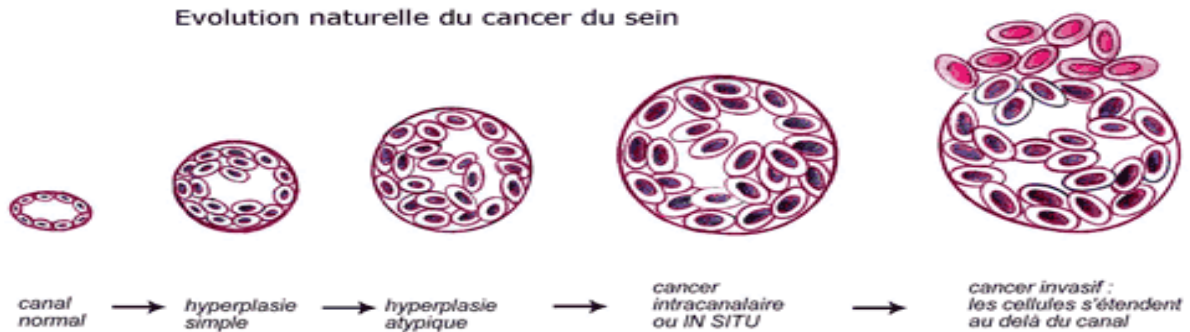
vaisseaux sanguins (artères et veines) et les vaisseaux lymphatiques. Ces derniers se drainent vers les ganglions de la chaîne mammaire interne, sous-clavière et surtout de la région axillaire.



3.2 Cancer du sein

Le cancer du sein est une tumeur maligne qui prend naissance dans les cellules du sein. Le mot « maligne » signifie que la tumeur peut se propager (métastases) vers d'autres parties du corps. Les cellules du sein subissent parfois des changements qui rendent leur mode de croissance ou leur comportement anormal. Ces changements peuvent engendrer des affections bénignes du sein, comme l'hyperplasie atypique et des kystes. Ils peuvent aussi entraîner la formation de tumeur bénigne, dans les papillomes intracanaux. Les affections et les tumeurs bénignes ne sont pas cancéreuses. Cependant, dans certains cas, des modifications dans les cellules mammaires peuvent causer un cancer du sein. Le cancer du sein se développe le plus souvent dans les cellules qui tapissent les canaux, ou tubes, qui transposent le lait des glandes au mamelon. Ce type de cancer du sein est appelé carcinome canalaire. Le cancer peut aussi se former dans les cellules des glandes productrices de lait (regroupées en lobules). Ce type de cancer porte le nom de carcinome

lobulaire. Ces deux cancers (carcinomes canalaire et lobulaire) peuvent être in situ, c'est-à-dire qu'ils demeurent confinés à leur emplacement d'origine et n'envahissent pas les tissus voisins. Ils peuvent également être infiltrant, ou invasifs, c'est-à-dire qu'ils se sont propagés dans les tissus voisins. Des types de cancer du sein, la maladie de Paget ainsi que le cancer du sein triple négatif et le cancer du sein de type basal en sont des exemples. Des types rares de cancer du sein sont entre autres de lymphome non hodgkinien et le sarcome. On peut résumer tout ce qui précède, en disant que le cancer du sein est une tumeur maligne qui, dans la plupart des cas, prend son point de départ à niveau de l'unité ducto-lobulaire. En fonction de sa topographie on parlera de cancer lobulaire (ceux qui se développent au niveau du lobule) ou de cancer canalaire (ceux qui se développent au niveau des canaux). Lorsque la tumeur reste confinée à l'intérieur d'un canal ou d'un lobule on parlera d'une lésion « in situ ». Par contre, lorsque le cancer, à un stade avancé, s'infiltré au-delà du canal ou du lobule, on parlera de lésion « invasive ».



3.3. Classification des cancers du sein

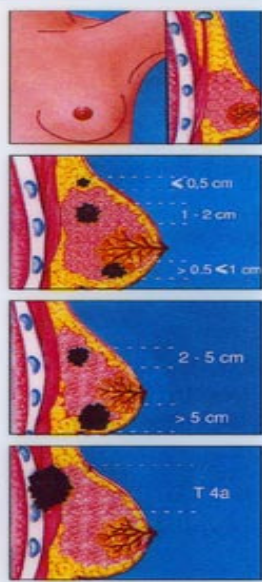
Il existe différentes possibilités de classification des tumeurs: selon l'organe, selon le tissu dont elles sont issues (anatomopathologique), selon leur aspect macroscopique, et selon leur extension. Pour évaluer l'étendue d'un cancer du sein, les médecins prennent en compte trois critères: la taille et l'infiltration de la tumeur, l'atteinte ou non des ganglions lymphatiques et la présence ou non de métastases. La taille et l'infiltration de la tumeur. Lorsque des cellules cancéreuses apparaissent, elles forment d'abord une tumeur au niveau des canaux ou des lobules du sein (carcinome in situ). Puis, progressivement, la tumeur peut traverser la paroi (appelée membrane basale) du canal ou du lobule et devenir ainsi infiltrant (on dit aussi invasive). Étudier la taille et l'infiltration de la tumeur donne donc une indication sur le degré d'évolution de la maladie. L'atteinte ou non des ganglions lymphatiques, le nombre et leur emplacement. Les cellules cancéreuses peuvent s'échapper du sein et se disséminer ailleurs. Les ganglions lymphatiques de l'aisselle (ganglions axillaires) sont les premiers à être potentiellement touchés. Lors de l'examen clinique, le médecin recherche systématiquement les ganglions anormaux en palpant les différents endroits où ils peuvent se trouver (essentiellement dans l'aisselle). Pour déterminer ou confirmer si des ganglions contiennent des cellules cancéreuses, il faut dans un second temps les analyser au microscope, après les avoir prélevés. Si des ganglions sont atteints, cela signifie que la maladie a commencé à se disséminer. Le nombre de ganglions envahis et leur emplacement permet d'en savoir plus sur le degré de propagation du cancer. La présence ou non de métastases. Les cellules cancéreuses peuvent envahir d'autres organes que les ganglions lymphatiques et y développer des métastases. Les organes les plus souvent touchés par des

métastases lors d'un cancer du sein sont le foie, les os et les poumons. Ces 3 critères – taille et infiltration de la tumeur, atteinte ou non des ganglions lymphatiques, présence ou non de métastases permettent de définir le stade du cancer selon la classification TNM de l'Union internationale contre le cancer (UICC) et de l'American Joint Committee on Cancer (AJCC), 7ème édition. TNM signifie en anglais « Tumor, Nodes, Metastasis » soit « tumeur, ganglions, métastases ». C'est le chirurgien français Pierre Denoix qui a eu le mérite de proposer cette classification (TNM) pouvant répondre à ces exigences. Son travail a débuté en 1943, pour être finalement retenu comme base de classification par le comité de nomenclature et de statistiques de l'UICC en 1953. C'est donc initialement grâce à un examen clinique minutieux que l'on devait pouvoir, en fin de consultation, résumer l'importance de la maladie en utilisant ce code. En revanche, l'idée de coder l'extension locale, régionale, ou générale est toujours à la base du système T (tumor), N (nodes), M (metastasis). D'une façon générale, on associe à ces 3 lettres des chiffres (dont la valeur augmente quand augmente la gravité) qui varie de 0 à 4 pour le T (la taille de la tumeur), de 0 à 3 pour le N (le degré d'envahissement), et sont soit 0 soit 1 pour le M (pour la présence ou non de métastase à distance). La juxtaposition de ces 3 lettres chiffrées donne une description abrégée de l'extension de la tumeur maligne. Cela conduit pour un organe donné à un grand nombre de possibilités TNM. On effectue alors un regroupement en stades que l'on définit pour qu'ils soient homogènes en durée de survie. Les cancers in situ sont toujours de stade 0, les métastatiques de stade IV.

Stade I correspond aux tumeurs les moins agressives.
Stade II est un grade intermédiaire.
Stade III correspond aux tumeurs les plus agressives.

Classification clinique

Tumeur primitive



Tx : aucune information sur la tumeur
T0 : pas de tumeur primitive

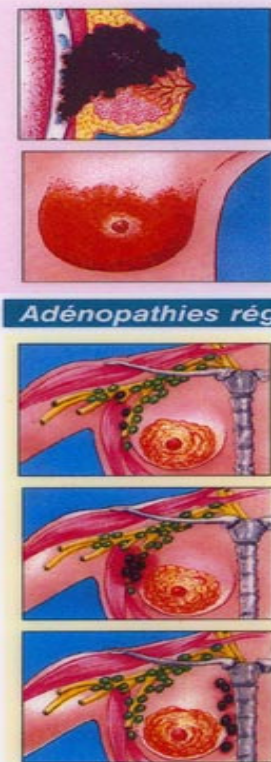
Tis : carcinome in situ

T1 : tumeur de moins de 2 cm
T1a : < 0,5 cm
T1b : 0,5 à 1 cm
T1c : 1 à 2 cm

T2 : tumeur de 2 à 5 cm

T3 : tumeur de plus de 5 cm

T4a : tumeur étendue à la paroi thoracique, quelle que soit sa taille



T4c : T4a + T4b

T4d : cancer inflammatoire

Adénopathies régionales

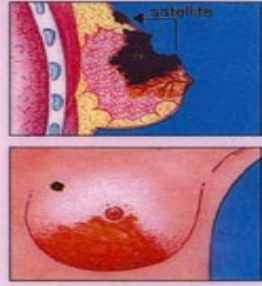
Nx : aucune information sur les adénopathies
N0 : pas d'adénopathie régionale

N1 : adénopathie homolatérale mobile

N2 : adénopathie homolatérale fixée


N3 : adénopathie mammaire interne homolatérale

Tumeur évoluée



T4b : tumeur étendue à la peau, quelle que soit sa taille : œdème, peau d'orange, ulcération, nodules internes sur le sein

Métastases à distance



Mx : aucune information sur les métastases
M0 : pas de métastase

M1 : métastase(s) à distance (y compris adénopathie sus-claviculaire)

3.4. Groupement par stade

Stade O	Tis	N0	M0
Stade I	T1	N0	M0
Stade IIA	T0	N1	M0
	T1	N1	M0
Stade IIB	T2	N0	M0
	T2	N1	M0
	T3	N0	M0
Stade IIIA	T0	N2	M0
	T1	N2	M0
	T2	N2	M0
	T3	N1, N2	M0
Stade IIIB	T4	Tous N	M0
	Tous T	N3	M0
Stade IV	Tous T	Tous N	M1

Note:

T0: pas de signe de tumeur primitive.
 Tis: carcinome in situ.
 T1: tumeur ≤2cm dans sa plus grande dimension.
 T2: tumeur >2cm et ≤5cm dans sa plus grande dimension.
 T3: tumeur >5cm dans sa plus grande dimension.

T4: tumeur de toute taille avec extension directe à la paroi thoracique.
 M0: Absence de métastases à distance.
 M1: présence de métastases à distance.*
 C'est une chaîne de Markov à cinq états.

3.5. Modélisation mathématique

Pour traiter ces données, on a choisit les chaînes de Markov, qui est une méthode stochastique parmi les méthodes qu'on peut utiliser pour écrire la dynamique d'une maladie dans ses différents stades. Tandis que la théorie et la pratique de la modélisation de la chaîne de Markov en Epidémiologie n'est pas nouvelle, plus que l'application de base de cet outil puissant peut être utilisée par les non mathématiciens aussi. C'est le mathématicien russe Andrei Markov qui, en 1902, proposa cet outil. Son objectif est de permettre d'appliquer certains calculs de probabilité à des problèmes ou la loi des grands nombres est inopérante. Il a appliqué ses chaînes à un domaine autre que les mathématiques. En 1913, il s'en est servi pour analyser les successions des lettres dans l'alphabet russe. Le même type d'applications se retrouve aujourd'hui dans les téléphones portables: les chaînes de Markov sont mises à contribution par les logiciels d'écriture de SMS, qui essaient de deviner ce que vous allez écrire. Plus sérieusement, les chaînes de Markov sont devenues un outil très employé, pour l'analyse de l'ADN, la prédiction de l'évolution des épidémies, ou pour classer l'importance des pages internet, et on retrouve d'autres applications, en mathématiques financières, météorologie, réseaux, gestion de stock..... Dans ce chapitre, nous présentons la théorie précédente et ses propriétés en annonçant les hypothèses et les conditions d'utilisation des chaînes de Markov dans l'étude de maladies et plus précisément le cancer de sein, ce qui pourrait faciliter la formulation d'hypothèses, la compréhension de l'évolution de cette maladie dans une population. Nous nous concentrons sur les chaînes de Markov absorbantes à temps discret, et nous allons étape par étape pour calculer: La probabilité d'infection par le cancer d sein pour un individu; La durée prévue de chaque état, L'espérance de vie pour les personnes malades.

3.6. Chaînes de Markov absorbantes:

Définition:

Un état est absorbant s'il forme, à lui seul, une classe récurrente.

Propriété:

Un état i est absorbant si et seulement si $R_{ii} = 1$ et $R_{ij} = 0 \forall i \neq j$

Définition:

Une chaîne de Markov est absorbante, si tous ses états récurrents sont absorbants, c'est-à-dire si chacune de ses classes récurrentes ne comporte qu'un seul état. On considère la forme canonique de la matrice de transition P

$$P = \begin{pmatrix} S & 0 \\ R & Q \end{pmatrix}$$

0 est une sous matrice nulle.

S est une sous matrice concernant les classes récurrentes de taille $(r - s) \times (r - s)$.

Q est une sous matrice concernant les classes transitoires de taille $s \times s$.

R est une sous matrice concernant les transitions de l'état transitoire à l'état récurrent de taille $(r - s) \times s$.

Dans le cas d'une chaîne de Markov absorbante, S est la matrice identité et la forme canonique de P devient:

$$P = \begin{pmatrix} I & 0 \\ R & Q \end{pmatrix}$$

Il est facile de montrer par récurrence que:

$$P^n = \begin{pmatrix} I & 0 \\ (I + Q + \dots + Q^{n-1})R & Q^n \end{pmatrix}$$

Proposition:

Soit la matrice de transition d'une chaîne de Markov absorbante, écrite sous forme canonique. Alors:

On a $\lim_{n \rightarrow +\infty} Q^n = 0$.

La matrice $I - Q$ est inversible, et son inverse vaut $(I - Q)^{-1} = \sum_{k=0}^{\infty} Q^k$

La matrice fondamentale:

Nous noterons N la matrice $(I - Q)^{-1}$, et nous l'appellerons la matrice fondamentale de la chaîne.

Alors $\lim_{n \rightarrow +\infty} P^n = \begin{pmatrix} I & 0 \\ NR & 0 \end{pmatrix}$.

On définit n_{ij} le nombre moyen de visites de l'état transitoire j avant absorption sachant que le processus débute dans l'état transitoire i est donné par: $n_{ij} = (N)_{ij} = (I - Q)^{-1}_{ij}$

Définissons, pour j transitoire, $y_n^j = \begin{cases} 1 & \text{si } X_n = j \\ 0 & \text{si } X_n \neq j \end{cases}$

Avec $\{X_n, n = 0, 1, 2, \dots\}$ une chaîne de Markov absorbante.

Notons $n_i = \sum_{n=0}^{\infty} y_n^i$ la fonction qui représente le nombre total de visites de l'état i

Théorème:

$E_i(n_i = N)$ avec i et j sont des états transitoires.

Application de La matrice fondamentale:

Théorème:

$QN = NQ = N - I$

Théorème:

On a $Var(n_i) = N(2N_{diag} - I) - N^2 = N_2$ avec i et j sont des états transitoires.

Remarque:

Si le processus est dans un état récurrent alors $t = 0$. Si le processus est dans un état transitoire alors t est le nombre de transitions nécessaire pour entrer à une classe récurrente, c'est le temps d'absorption.

Théorème:

$E_i(t) = N\xi = \tau, \quad var_i(t) = (2N - I)\tau - \tau^2 = \tau_2$ avec $i \in T$

Corollaire:

Si π est le vecteur de probabilité initiale d'une chaîne absorbante et π' constituée de s derniers composantes de π . C'est-à-dire π' est le vecteur de probabilité initial pour les états transitoires. Alors:

$$E_{\pi}(\sum_{n=1}^{+\infty} Y_n) = \pi' N$$

$$Var_{\pi}(\sum_{n=1}^{+\infty} Y_n) = \pi' N(2N_{d\bar{s}} - I) - (\pi' N)^2$$

$$E_{\pi}(t) = \pi' \tau$$

Théorème:

Dans une chaîne de Markov absorbante, la probabilité d'absorption par l'état absorbant j sachant que le processus débute dans l'état transitoire i est égale au terme b_{ij} de la matrice: $B = NR$

4. RESULTATS:

Dans notre cas on a 8 états: etat0, I IIA, IIB, IIIA, IIIB, IV et l'état de la mort. Alors on aboutit à une matrice de transition de la forme:

$$P = \begin{pmatrix} P_{11} & P_{12} & P_{13} & P_{14} & P_{15} & P_{16} & P_{17} & P_{18} \\ P_{12} & P_{22} & P_{23} & P_{24} & P_{25} & P_{26} & P_{27} & P_{28} \\ P_{31} & P_{32} & P_{33} & P_{34} & P_{35} & P_{36} & P_{37} & P_{38} \\ P_{41} & P_{42} & P_{43} & P_{44} & P_{45} & P_{46} & P_{47} & P_{48} \\ P_{51} & P_{52} & P_{53} & P_{54} & P_{55} & P_{56} & P_{57} & P_{58} \\ P_{61} & P_{62} & P_{63} & P_{64} & P_{65} & P_{66} & P_{67} & P_{68} \\ P_{71} & P_{72} & P_{73} & P_{74} & P_{67} & P_{76} & P_{77} & P_{78} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

D'où le temps moyen d'absorption est:

$$W = (I - Q)^{-1} \times \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

$$Q = \begin{pmatrix} P_{11} & P_{12} & P_{13} & P_{14} & P_{15} & P_{16} & P_{17} \\ P_{12} & P_{22} & P_{23} & P_{24} & P_{25} & P_{26} & P_{27} \\ P_{31} & P_{32} & P_{33} & P_{34} & P_{35} & P_{36} & P_{37} \\ P_{41} & P_{42} & P_{43} & P_{44} & P_{45} & P_{46} & P_{47} \\ P_{51} & P_{52} & P_{53} & P_{54} & P_{55} & P_{56} & P_{57} \\ P_{61} & P_{62} & P_{63} & P_{64} & P_{65} & P_{66} & P_{67} \\ P_{71} & P_{72} & P_{73} & P_{74} & P_{67} & P_{76} & P_{77} \end{pmatrix}$$

et la probabilité d'absorption est:

$$B = (I - Q)^{-1} \times R \quad \text{avec} \quad R = \begin{pmatrix} P_{18} \\ P_{28} \\ P_{38} \\ P_{48} \\ P_{58} \\ P_{68} \\ P_{78} \end{pmatrix}$$

5. CONCLUSION:

Dans ce travail, nous avons traité juste le coté théorique, et nous avons conclu que les chaînes de Markov nous permettent de calculer la probabilité d'absorption et le temps moyen d'absorption, mais l'étape la plus importante est l'estimation des probabilités de passage à partir des données. Il nous reste de collecter des données en observant un échantillonnage au fil du temps; pour calculer les probabilités de transitions.

REFERENCES

Ma recherche a été limitée aux travaux menés début Janvier 2014, il s'agissait d'identifier des travaux épidémiologiques sur l'évolution du cancer dans une population, et plus précisément sur les thèmes suivants:

- Application des chaînes de Markov dans l'étude de l'évolution d'une maladie.
- La différence entre la modélisation déterministe et stochastique.
- Cancer du sein.

La recherche a été limitée aux travaux menés en France, dans le reste de l'Europe et en Amérique du Nord (USA, Canada), publiés en langue anglaise ou française.

Parmi ces travaux:

- Classification TNM de l'UICC, 4ème édition 1988 (UNIONINTERNATIO1988)
- The statistician (2003): Multistate Markov models for disease progression with classification error.
- Methods in ecology & evolution (2010): A primer on the application of Markov chains to the study of wildlife disease dynamics
- Department of internal medicine, Division of Biostatistics, Virginia (2011): The dynamics of disease progression in Sepsis: Markov Modeling describing the natural history and the likely impact of effective Antisepsis Agents
- Université Joseph Fourier de Grenoble (2011-2012) Le sein: de l'embryologie à l'allaitement.