# Modélisation aléatoire de l'activité des Lymphocytes T Cytotoxyques

Claire Christophe


# THÈSE

**En vue de l'obtention du**

## DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

**Délivré par :** *l'Université Toulouse 3 Paul Sabatier (UT3 Paul Sabatier)*

---

**Présentée et soutenue le** *2 décembre 2014* **par :**

CLAIRE CHRISTOPHE

---

## Modélisation aléatoire de l'activité des Lymphocytes T Cytotoxiques

---

**JURY**

| | | |
|---|---|---|
| VINCENT BANSAYE | Professeur, Ecole Polytechnique | Rapporteur |
| PATRICK CATTIAUX | Université Paul Sabatier | Directeur |
| NICOLAS CHAMPAGNAT | INRIA Nancy - Grand Est | Examinateur |
| SÉBASTIEN GADAT | Université Capitole | Directeur |
| MARC LAVIELLE | INRIA Saclay - Île-de-France | Président |
| MICHÈLE THIEULLEN | Université Pierre et Marie Curie | Examinatrice |
| SALVATORE VALITUTTI | INSERM CPTP Toulouse | Examinateur |

---

**École doctorale et spécialité :**
*MITT : Domaine Mathématiques : Mathématiques appliquées*

**Unité de Recherche :**
*Institut de Mathématiques de Toulouse*

**Directeur(s) de Thèse :**
*Patrick Cattiaux* et *Sébastien Gadat*

**Rapporteurs :**
*Vincent Bansaye* et *Adeline Leclercq Samson*

# Remerciements

Tout d'abord je souhaite remercier mes directeurs de thèse, Patrick, Sébastien, notamment pour le sujet de thèse, qui inspire de nombreuses chansons, telle que "Ah ! Qu'est ce qu'on est serré au fond de cette tumeur ! Chantent les cellules, chantent les cellules", et j'en passe des meilleures. Plus sérieusement, je vous remercie pour votre enthousiasme et l'ambiance de travail décontractée. Sébastien, avec qui j'ai fait mes premiers pas en recherche, je te remercie pour ton écoute, ta disponibilité et pour m'avoir plusieurs fois remonté le moral ; souvent je sortais des rendez vous avec le sourire.

Je remercie chaleureusement Adeline Leclercq Samson et Vincent Bansaye d'avoir rapporté cette thèse avec minutie et précision. Je remercie aussi Marc Lavielle, Michèle Thieullen et Nicolas Champagnat pour avoir accepté de faire partie de mon jury.

Un grand merci à la chaleureuse *famiglia* de Salvatore Valitutti pour son accueil en ses terres qui, il y a 3 ans, m'étaient complètement inconnues. Merci d'avoir répondu à mes nombreuses questions. Grâce à vous, il me semble que j'ai un deuxième chez moi (rassurez vous, même si j'ai manié une fois la pipette, je ne suis pas prête à vous remplacer aux manips).

Plus spécialement, je remercie Salvatore pour son enthousiasme, sa disponibilité, les nombreuses dis-cussions, mais aussi pour son accent qui permet de faire un saut spatio-temporel nous transportant en une seconde au soleil, face à la mer, sur une terrasse à manger une *pizza*. Merci à Magda pour son organisation qui n'oublie personne, à Régis pour le "Bonjour Claire" (quoique), à Roxana qui n'oublie jamais d'être féminine, à Fanny et Nico pour avoir installé une bonne ambiance au labo, il ne nous restait plus qu'à vous imiter ; et aussi pour les nombreux repas où vous me livriez les secrets de l'INSERM. Merci à Eric, Loïc et Zilton pour les discussions math bio. Ah Sabina, je me souviens encore des pâtes au pois chiches, Cécile, mes papilles frissonnent encore. Enfin, merci à Javier et à tous les petits, pour les bons moments passés dans l'aquarium.

Je souhaite remercier Jean-Michel qui, en envoyant Hélène au Chili dès le mois d'octobre de sa première année de thèse, a déclenché la magnifique collection de carte postale du bureau 201 (souvent imitée, jamais égalée). Merci à tous ceux qui ont contri-

bué à cette collection !

Merci à Bertrand pour les nombreuses discussions aussi bien mathématiques que sur la vie de labo, pour m'avoir expliqué plein de trucs, même si je ne les comprends pas tous encore très bien. Merci à Sanae pour les moments de détentes et pour avoir supporter les nombreuses discussions matheuses. Merci à Malika pour sa bonne humeur, à Mélanie "L'organisatrice", à Benoit pour nous rappeler qui sont les vrais méchants. Merci les Claires pour les nombreuses discussions ; il faut que je vous avoue, vous m'avez vaincue car malgré mes multiples tentatives, Claire 2 et Claire 3 n'a absolument pas pris au labo. Merci aux anciens doctorants qui nous ont montré le chemin : Tibo et son enthousiasme, Thibaut et ses drôles d'aventures, Paul, Julie,... Merci Gaëlle et Loïc pour les escapades non toulousaines.

Bien sûr, je remercie mes co-bureaux. Je commence par mes co-burettes de 3 ans, Hélène et Magali. Tout d'abord, je remercie vos oreilles, en effet, vous m'avez beaucoup beaucoup beaucoup écoutée, quand "de temps en temps il pouvait arriver qu'exceptionnellement vous m'entendiez râler un tout petit peu". Puis je remercie vos yeux, qui ont relu des mails, des résumés... merci encore ! Merci à Anne-Claire ma fournisseuse officielle de rajouts. Enfin les nouveaux co-bureaux Brendan, Clément, Jonathan et Sylvain qui ont apporté, comment dire, une touche masculine au bureau. Merci Anne-Charline, pour les nombreuses discussions sur tout et rien et sur les maths aussi, ton bureau était presque mon deuxième bureau. C'est à ce titre, suivant l'adage bien connu "les co-bureaux de mes co-bureaux sont mes co-bureaux", que je remercie Fabien et ses desserts parfaits, le théâtre et l'art théâtral, Mathieu pour les interminables discussions (et même pas sur le cancer), Anaïs pour avoir fêter Pâques juste avant Noël, Benjamin pour les supers randos et Julien, grâce à toi je peux réécouter des cassettes dans ma voiture !

Enfin, Je remercie aussi tous les autres doctorants, aussi bien Espiens que Mipiens ou Picardiens, qui contribuent à une super ambiance au labo.

Merci au staff administratif, Marie-Laure, Delphine, Françoise, Marie-Line..., qui facilitent la vie au quotidien dans le labo, merci aux informatologues, toujours disponibles. Merci aussi pour nous avoir montré les dessous du labo les midis.

Je remercie ceux que j'ai oubliés, il n'y a pas de doute, ils se reconnaîtront.

Il est enfin venu le temps de remercier les amis. Merci à Marianne et Aurore, Yannick et Sophie, Rémi, pour vos récits de voyages, qui m'ont fait rêver et m'ont permis de m'évader un peu de Toulouse, merci pour les cartes postales ! Merci à Christelle et François, Vanessa, Florbela et Rémi, je suis ravie de voir votre enthousiasme pour les maths que vous enseignez. Merci à Charlotte, Cisco, Coco, pour les nombreuses rigolades.

Merci Guigui, pour les randos à toutes les saisons, ton écoute, tes conseils, tes

histoires, bref, pour le goût myrtille.

Je remercie les nombreux relecteurs par morceaux, qui font peut-être de cette thèse une des plus lues !

Je dédie cette thèse à ma famille.

Merci à mes parents sur qui je peux toujours compter, aux bons moments mais aussi aux moments de doutes, votre écoute m'aide beaucoup ; Trespécoul est pour moi, un lieu reposant et ressourçant. Merci à Sophie, ma jum's. Tu n'as jamais douté, tu avais même peur de me perdre si je faisais des maths trop théoriques ("déjà qu'on ne te comprend pas tout le temps"). Merci aussi pour les voyages pendant la thèse, pour ses moments de détente mais aussi pour avoir supporté parfois mon stress. Merci à Julien mon frérot et à Coline, (ma plus "vieille amie", je te mets dans la famille, et j'en suis ravie !), même si ça fait un moment que nous n'avons plus eu l'occasion de combattre des fantômes, vous comptez beaucoup pour moi. Merci à Marlène, pour son écoute inlassable, pour le voyage près-thèse inoubliable. Je ne doute pas que le voyage post-thèse sera tout aussi inoubliable. Merci à mes grands-parents qui, du haut de leur 84 et 88 ans, sont restés toujours aussi jeunes et cools.

# Table des matières

# Introduction

Ce travail de thèse propose une étude des propriétés probabilistes et statistiques de la dynamique entre des cellules immunitaires, plus spécialement des Lymphocytes T Cytotoxiques (CTL) et un nodule tumoral. Il se situe en étroite collaboration avec l'équipe 1043 de Salvatore Valitutti de l'INSERM à l'hôpital Purpan, qui a contribué aux données expérimentales et à la rigueur immunologique.

Dans un premier temps, nous présentons le contexte biologique dans lequel se place cette collaboration, ainsi que les problématiques qui en découlent. Nous présentons ensuite les modèles mathématiques développés afin de répondre aux questions soulevées.

## Nodules tumoraux et système immunitaire

Les sous-sections ci-dessous apportent les connaissance sur les phénomènes biologiques qui, partant d'une cellule saine, conduisent à la formation d'un nodule tumoral. Puis nous présentons les moyens mis en place par le système immunitaire pour combattre un nodule tumoral. Enfin, nous exposons les problématiques que nous avons abordées.

### Formation d'un nodule tumoral

**De la cellule saine à la cellule tumorale** Les cellules ont diverses fonctions, suivant leur place dans l'organisme. Cependant, toutes les cellules sont programmées pour se multiplier puis mourir. Cette mort cellulaire programmée, suite à un processus normal d'autodestruction, est dite **apoptose**. Le processus conduisant une cellule, dite **mère**, à grossir pour ensuite donner naissance à deux cellules identiques, dites **filles**, est appelé le **cycle cellulaire**. L'ADN contenu dans les chromosomes du noyau de la cellule, transmet à la cellule les informations, comme ceux de se diviser ou de mourir. L'ADN est composé de gènes, qui parfois subissent une altération, appelée une mutation. Dans certains cas, ces mutations peuvent conduire à une transmission d'ordres anormaux.

C'est au cours du processus lent d'accumulation des mutations, qui confère de nouvelles propriétés aux cellules, que progressivement, une cellule saine se transforme alors

en une cellule cancéreuse. Ce développement s'apparente au processus darwinien. **Six marqueurs fondamentaux**, distincts et complémentaires, ont été identifiés comme étant présents dans la plupart des cancers humains. Voici une brève description de ces marqueurs (pour plus d'informations nous renvoyons à [HW00]).

**Autonomie de croissance.** La prolifération des cellules est normalement la conséquence de signaux stimulant la croissance : une cellule cancéreuse a la faculté de croître indépendamment de ces signaux.

**Insensibilité à la non prolifération** L'organisme dispose de mécanismes pour arrêter la croissance cellulaire : une cellule cancéreuse a la faculté d'ignorer ces signaux antiprolifératifs.

**Manquement à l'apoptose.** Une cellule cancéreuse a la faculté d'ignorer les signaux d'autodestruction envoyés par le noyau : elle évite l'apoptose.

**Réplication à l'infini.** Une cellule saine ne peut se diviser qu'un nombre fini de fois, mais une cellule cancéreuse n'est plus soumise à cette limite.

**Induction de l'angiogenèse.** L'angiogénèse est un processus consistant à créer de nouveaux vaisseaux sanguins, apportant ainsi des nutriments et de l'oxygène à la population cellulaire et contribuant ainsi à sa prolifération. Les cellules cancéreuses ont la capacité de réaliser l'angiogenèse.

**Creation de métastases.** Les cellules cancéreuses ont la capacité de migrer d'un organe (ou tissu) à un autre organe (ou tissu). Elles forment alors de nouvelles lésions analogues, appelées métastases.

Nous venons de voir que les cellules tumorales ont, entre autres, la capacité de proliférer. Dans les tissus, ces cellules cancéreuses s'agglomèrent ce qui forme une tumeur, dit aussi un nodule tumoral. Nous développons dans le prochain paragraphe une description concise de cette formation.

**Le nodule tumoral** Un nodule tumoral désigne un amas de cellules. Le terme de tumeur maligne est employé pour désigner des tumeurs constituées de cellules cancéreuses. Le développement d'une tumeur maligne est habituellement caractérisé par trois étapes.

La première étape est la croissance du nodule en captant les nutriments et l'oxygène nécessaires à la division cellulaire dans son microenvironnement. A partir d'un certain nombre de cellules, autrement dit, une certaine taille du nodule, il est composé de cellules qui ont des statuts différents. Cela est du à la structure tridimensionnelle de la tumeur. Celle-ci rend l'accès aux nutriments et à l'oxygène pour une cellule, d'autant plus restreint que cette dernière est loin du bord du nodule. Il est généralement admis que le nodule est divisé en trois parties :

– Un **corps nécrotique** situé au centre du nodule, qui est constitué de cellules **nécrotiques** à savoir des cellules mortes ou mourantes, des suites de dégâts cellulaires. Cette mort cellulaire est différente de l'apoptose qui est une mort

programmée.
- Une **partie quiescente**, qui est composée de cellule qui ont stoppé leur croissance.
- Une **partie proliférative** se trouve en superficie du nodule, elle est constituée de cellules cancéreuses qui se divisent et contribuent ainsi à la croissance du nodule.

La deuxième étape du développement d'une tumeur maligne est la vascularisation, grâce à la mise en place de l'angiogénèse, qui donne accès aux nutriments et à l'oxygène à d'avantage de cellules.

La dernière étape du cancer est la phase métastatique, durant laquelle la tumeur colonise l'organisme.

## La réponse immunitaire

Pour lutter contre le cancer, la médecine actuelle a développé différents types de traitements. Le premier traitement est la **chirurgie** qui consiste à enlever la tumeur. Le deuxième traitement est la **radiothérapie**, qui par exposition de la tumeur à des rayonnements, empêche sa croissance et entraine sa destruction. Le troisième traitement est la **chimiothérapie** qui repose sur l'administration de molécules chimiques, permettant, quelle que soit la localisation de la tumeur, soit d'empêcher la prolifération des cellules tumorales, soit de les détruire. Le quatrième traitement, l'**immunothérapie**, consiste à administrer des molécules d'origine chimique ou biologique, qui ont pour but de stimuler la réponse immunitaire.

La radiothérapie et la chimiothérapie ne sont pas ciblées sur la tumeur, elles touchent aussi des cellules saines, provoquant ainsi des effets secondaires du traitement. L'immunothérapie, par contre, est beaucoup plus ciblée. L'**immunologie** est la discipline qui étudie les mécanismes de défenses de l'organisme vis-à-vis des substances reconnues comme étrangères, comme les cellules cancéreuses. Une partie de la recherche actuelle en immunologie a pour vocation d'attaquer la tumeur en réduisant les effets secondaires [CF13].

Dans de le cadre de cette thèse, nous nous sommes intéressés à des problématiques provenant de l'immunothérapie, notamment sur l'activité des Lymphocytes T Cytotxiques (CTL). Ces cellules jouent un rôle central dans le fonctionnement du système immunitaire, en détruisant les substances reconnues comme étrangères à l'organisme ; les cellules tumorales en sont un exemple [DL10].

**Fonction cytolytique d'un CTL** C'est l'activité permettant à un CTL de tuer une cellule cible. Elle est déclenchée via la reconnaissance d'un antigène spécifique lié au CMH (Complexe Majeur d'Histocompatibilité), présent à la surface des cellules cibles, par le récepteur spécifique des CTL : le TCR (le Récepteur des Cellules T). Ainsi, la sécrétion de molécules cytolytiques, stockées dans des granules intracellulaires des

CTL, est déclenchée, conduisant à la perforation de la membrane [1] de la cellule cible puis sa mort. Cette aire de contact très organisée entre le CTL et sa cible, où a lieu la sécrétion de molécules cytolytiques, est dite **synapse immunologique** [VE05].

Les cellules tumorales, tout comme les cellules infectées par un virus, présentent l'antigène tumoral ou viral lié au CMH, reconnu par les CTL. Ceux-ci peuvent donc les détruire.

Nous venons de voir qu'un CTL est capable d'éliminer les cellules tumorales, et en ce sens, il accomplit son rôle au sein du système immunitaire, qui consiste à éliminer les cellules pathogènes. Cependant, certaines tumeurs parviennent à échapper à cette surveillance, et continuent à se développer. Ce manque d'efficacité s'explique par le double rôle joué par le système immunitaire, qui est détaillé dans le paragraphe suivant.

**Immunoediting** C'est le processus décrivant les changements survenus sur un nodule tumoral sous la pression du système immunitaire. Ces changements sont caractérisés par le passage progressif d'un nodule dominé par le système immunitaire, à un nodule tumoral hors de contrôle du système immunitaire. Ce processus est caractérisé par trois phases successives, voir [SOS11] pour plus de détails.

**Elimination** Nous parlons aussi d'**immunosurveillance**. C'est la phase pendant laquelle le système immunitaire se met en place : détecte la tumeur et détruit des cellules tumorales entrainant la décroissance du nodule.

**Equilibre** La tumeur est alors composée de cellules qui ont échappé au système immunitaire et qui sont sélectionnées pour croître. Cette phase est la plus longue, et à l'image de la sélection darwinienne, de nouvelles versions de cellules tumorales, qui sont de plus en plus résistantes au système immunitaire, apparaissent. Dans cette phase la croissance de la tumeur est contrôlée par le système immunitaire.

**Echappement** Les cellules tumorales sélectionnées ont ouvert une brèche dans les défenses immunitaires de l'organisme. Elles ont la capacité de contourner la reconnaissance ou même leur destruction par les CTL, certaines cellules perdent la molécule CMH, ce qui les rendent invisibles aux CTL.

Non seulement, les cellules tumorales deviennent de plus en plus résistantes aux CTL, mais en plus elles "relarguent" dans le microenvironnement des molécules qui entrainent la défaillance des CTL à éradiquer la tumeur [Cre+13]. Nous parlerons d'**"exhaustion"** des CTL.

Ce phénomène est aussi décrit dans [HW11]. Il est considéré comme un des deux nouveaux marqueurs fondamentaux caractéristiques des cancers (ils s'ajoutent aux six premiers, introduits dans le paragraphe "de la cellule saine à la cellule tumorale"). Le second nouveau marqueur fondamental est une dérégulation du métabolisme de la cellule tumorale, pour s'adapter aux ressources disponibles. Ces deux derniers soulignent

1. Enveloppe délimitant la cellule et qui la sépare son environnement.

la capacité exceptionnelle de mutation et d'adaptation des cellules tumorales.

## Cadre de la thèse et problématiques

Nous nous intéressons au premier stade du développement de la tumeur, à savoir une tumeur ni vascularisée ni métastatique. La tumeur est alors en pleine croissance. Nous supposons qu'elle est d'assez grande taille pour avoir une structure divisée en 3 parties, et que dans son microenvironnement se trouve des CTL capables de détruire des cellules tumorales. Dans le cadre de cette thèse, nous nous plaçons au niveau de l'étape d'élimination conjuguée au processus d'immunoediting.

**Problématiques** Ce stade d'une tumeur, plutôt précoce, reste cependant digne d'intérêt, car il se situe à une étape clé à comprendre. En effet, comme nous venons de le voir, un CTL est efficace contre une cellule tumorale, mais peut perdre son efficacité face à un nodule tumoral. Cela conduit à une hyper croissance de la tumeur via la vascularisation, et à la colonisation de tissus ou d'organes sains par la tumeur ; entrainant un diagnostic plutôt pessimiste quant à la survie d'un patient.

Nous souhaitons alors comprendre quels sont les mécanismes, mis en place par les CTL pour détruire une cellule tumorale, qui sont défaillants face à un nodule tumoral. En effet, comprendre ces mécanismes peut être la clé de nouvelles stratégies thérapeutiques.

Plus formellement, nous voulons déterminer les paramètres entrant en compte dans l'interaction CTL/nodule tumoral et, chercher ceux qui sont pertinents et, ceux qui favorisent la réponse immunitaire. Aussi, nous voulons comprendre comment la modification de ces paramètres peut optimiser la réponse immunitaire.

**Choix de la modélisation mathématique** Il est possible de réaliser des expériences *in vivo*[2], par exemple, en implantant dans des souris des tumeurs. C'est un processus long, lent et couteux. De plus, il peut être assez difficile expérimentalement de faire varier les paramètres, et donc de quantifier leurs effets sur le développement de la tumeur.

Il est aussi possible de réaliser des expériences *in vitro*[3]. En effet il existe des protocoles permettant de réaliser des tumeurs dans des puits. Cela consiste à placer des cellules tumorales dans un puits contenant du milieu (liquide riche en nutriments), puis à retourner le puits à l'envers. Ainsi, par la gravité, les cellules tombent dans le creux de la goute de milieu et forment un nodule. Pour étudier l'interaction avec des CTL, une fois formé, le nodule est placé dans un plaque de culture en forme de $U$ avec des CTL, voir la Figure 1. Par contre, ce modèle ne permet pas d'appréhender une interaction tridimensionnelle entre les CTL et un nodule tumoral, car les CTL, eux

2. Expériences pratiquées sur un organisme vivant.
3. Expériences pratiquées dans des laboratoire, en dehors d'un organisme vivant.

aussi, tombent au fond de la goute, ils ne sont pas répartis de manière homogène dans le milieu.

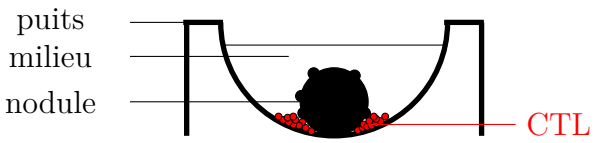



FIGURE 1 – Schéma d'expérience dans un puits, *in vitro*, d'un nodule et de CTL.

Ainsi, il est difficile d'appréhender cette interaction par des expériences biologiques. Pour pallier cette difficulté, nous proposons des modèles mathématiques ayant pour rôle de compléter le volet expérimental, en proposant des études numériques et des études statistiques et probabilistes. Nous présentons trois modèles étudiés pendant cette thèse.

# PARTIE 1 : **Un modèle agent centré pour déterminer les paramètres pertinents dans l'interaction entre des CTL et un nodule tumoral**

Nous nous plaçons dans le contexte présenté ci-dessus, à savoir, un nodule tumoral sous la pression du système immunitaire confronté à la présence de CTL dans son microenvironement. Nous nous plaçons également dans le cadre de l'immunoediting.

Nous souhaitons définir les paramètres pertinents de cette interaction, et déterminer ceux qui favorisent la réponse immunitaire.

**Modélisation mathématique** Un modèle agent centré bi-dimentionnel sur grille est proposé pour décrire cette interaction. Ce choix de modèle, bien qu'étant une simplification, offre une représentation claire. De plus, il permet une implémentation assez simple et une facilité de modification, ce qui permet de bien comprendre le rôle de chaque paramètre dont dépend le système. En outre, des données expérimentales, provenant d'observations biologiques acquises en deux dimensions, ont fait l'objet d'études statistiques, afin d'estimer les échelles des paramètres.

Nous supposons que les déplacements des CTL, la croissance des cellules tumorales et le temps d'élimination d'une cellule tumorale par un CTL, ne sont pas déterministes, mais suivent des lois de probabilités que nous préciserons.

**Contributions** Nous avons mis en évidence deux paramètres, qui jouent un rôle important dans cette interaction et qui peuvent favoriser la réponse immunitaire.

PARAMÈTRE 1 : **Le déplacement des CTL** Dans un premier temps, une marche aléatoire symétrique est présumée pour le déplacement des CTL.

En vue d'une immunothérapie, dans un second temps, nous proposons une marche aléatoire biaisée. Plus précisément, nous suggérons une immunothérapie consistant à produire des CTL, qui une fois sur le bord du nodule, sont capables de relarguer des chimiokines [4], attirant ainsi les autres CTL dans leur direction. Nous montrons que diriger les CTL améliore la réponse immunitaire.

Paramètre 2 : **Le nombre de cellules cibles éliminées par un CTL** Nous retrouvons dans notre modèle que plus ce nombre est grand, meilleure est la réponse immunitaire.

Nous avons également mis en évidence, qu'au contraire, la **réduction du temps mis par un CTL pour éliminer un cellule cible**, ne joue pas un rôle important dans l'interaction CTL/nodule tumoral.

Dans chacune des deux parties qui suivent, nous étudions plus spécifiquement un de ces deux paramètres.

Pour la partie 2, nous restons dans ce cadre : un nodule tumoral sous la pression du système immunitaire. Nous souhaitons quantifier, par un système d'équations différentielles, le bénéfice apporté, par l'orientation des CTL, à la réponse immunitaire.

Dans la partie 3, pour étudier l'activité cytolytique individuel des CTL, afin de déterminer le nombre de cellules cibles détruites par un CTL, nous changeons de cadre. Nous nous plaçons dans le cas où un CTL est confronté à un grand nombre de cellules cibles.

## Partie 2 : Un système d'EDO pour quantifier l'importance de l'attraction des CTL

Dans cette étude, nous supposons aussi un nodule tumoral confronté à la réponse immunitaire, c'est-à-dire avec des CTL dans le microenvironnement de la tumeur. Notons que, tant que le CTL ne touche pas le nodule, il ne détruit pas de cellules tumorales. Ainsi le temps mis par un CTL pour atteindre le nodule, autrement dit son déplacement, est un paramètre non négligeable. Dans cette partie, nous souhaitons comprendre quel peut être l'avantage d'une immunothérapie consistant à diriger les CTL vers le nodule.

**Modélisation mathématique** Afin de comprendre l'importance de la trajectoire des CTL, nous proposons deux dynamiques aléatoires pour modéliser leur déplacement : un mouvement Brownien dans le cas d'un déplacement non dirigé, et sous immunothérapie, un Processus d'Ornstein-Uhlenbeck. Ce dernier traduit l'attraction des CTL vers le nodule.

4. petites protéines, dont la fonction la plus étudiée est l'attraction : chimiotaxie

Au vue du grand nombre de cellules tumorales dans le nodule, nous proposons une évolution déterministe du nodule, qui vient d'un modèle microscopique probabiliste. Elle dépend du nombre de CTL sur le bord du nodule. Ce dernier est notamment déterminé par les temps d'atteintes du nodule par les CTL.

Afin de simplifier le problème, nous nous ramenons à un problème de dimension 1, en considérant d'une part le processus de Bessel, et d'autre part, le processus d'Ornstein-Uhlenbeck radial, qui représentent la loi de la norme des deux processus précédents.

**Contributions** *D'un point de vue mathématique.* Nous donnons une approximation de la probabilité du temps d'atteinte d'un point $R$, différent de 0, alternativement, par un processus de Bessel ou un processus d'Ornstein-Uhlenbeck radial.

Nous mettons en évidence l'existence de distribution quasi-stationaire pour le processus de Bessel et le processus d'Ornstein-Uhlenbeck radial évoluant dans un intervalle de $\mathbb{R}$, absorbés en un des bords et rejetés en l'autre. Grâce à l'algorithme Fleming-Viot, nous obtenons une estimation numérique de la première valeur propre des générateurs associés à ces deux processus.

Nous donnons des solutions numériques au modèle proie-prédateur proposé, révélant des phénomènes de transitions de phases.

*D'un point de vue biologique.* A rayon de tumeur fixé, nous mettons en évidence l'existence d'une **transition de phase** dans le nombre de CTL nécessaires pour éradicaquer le nodule, et que cette transition se produit avec moins de CTL dans des conditions d'attraction des CTL vers le nodule.

Le système d'équations développé dans cette partie, étant rapide d'implémentation, pourrait être utilisé pour des pratiques cliniques. C'est-à-dire, après avoir calibrer les paramètres du système d'équations aux données observées d'un patient, il serait possible d'administrer un traitement optimal et personnalisé.

## Partie 3 : **Un modèle statistique décrivant la fonction cytolytique des CTL**

Bien que l'activité cytolytique des CTL soit étudiée depuis de nombreuses années par les biologistes, la question du nombre de cellules cibles éliminées par un CTL reste en suspens. Au vue de l'importance que cela peut apporter à la réponse immunitaire, de nouvelles expériences biologiques ont été réalisées. Pour cela, un CTL a été confronté à un nombre relativement grand de cellules cibles (une dizaine), pendant douze heures. Le nombre de cibles éliminées a été enregistré, et une grande variabilité a été observée : un CTL peut tuer de 0 à 12 cibles.

Dans cette partie, nous répondons à la question suivante : est-ce que cette variabilité provient de l'activité cytolytique d'une population homogène de CTL ou, au contraire,

provient-elle d'une population hétérogène, dont les paramètres sont à estimer.

**Modélisation statistique** Les observations biologiques étant des temps de mort, nous proposons des modèles mettant en jeux des lois de Poisson pour décrire cet échantillon. L'hétérogénéité du nombre peut alors provenir d'une seule loi ou plus profondément d'un mélange de sous-populations. En statistique, le mélange de lois est habituellement utilisé pour modéliser le fait qu'une population est divisée en plusieurs sous-populations.

Notons qu'un mélange de lois de poisson est caractérisé par un nombre de lois composant le mélange : **la taille du mélange ou la taille du modèle** et par ce que nous appelons les **paramètres du mélange**, à savoir, la proportion de chaque loi et la valeur des paramètres des lois. Nous verrons que la valeur maximale de ces paramètres des lois de Poisson, que nous appelons **taille maximale des paramètres**, joue un rôle important pour l'estimation statistique.

**Contributions** *D'un point de vue mathématique.* Nous mettons en place un algorithme $EM$ pour des mélanges finis de lois de Poisson. A taille de mélange fixée, grâce à l'algorithme $EM$, nous obtenons des estimateurs des paramètres d'un mélange. Par une étude numérique, nous mettons en évidence que cette estimateur se détériore si la taille maximale des paramètres des lois de Poisson est grande.

Dans le cadre de la sélection de modèle, nous proposons un critère pénalisé non asymptotique pour les modèles de mélanges de lois de Poisson. Ce critère pénalise la taille des modèles, mais aussi la taille maximale des paramètres des lois. Ainsi, minimiser ce critère pénalisé permet d'obtenir des estimateurs consistants. Nous suggérons un algorithme permettant d'implémenter ce critère pénalisé à des observations.

Afin d'obtenir le critère pénalisé, nous donnons une majoration de l'entropie à crocher pour l'ensemble des mélanges finis de lois de Poisson puis pour l'ensemble des mélanges quelconques de lois de Poisson.

*D'un point de vue biologique.* Nous avons mis en évidence l'existence de **deux sous-populations de CTL**. Une sous-population représentant un tiers de la population totale et composée de CTL qui tue en moyenne 6,4 cellules cibles en 12h, et une seconde sous-population qui détruit en moyenne 2,8 cibles en 12h.

# Part I

# Agent based model to describe the interaction between Cytotoxic T Lymphocytes and tumor nodule

# Chapter 1

# Identification of parameters which improve the CTL efficacy during cancer immunoediting [1]

A major constraint in natural and therapeutically induced immune responses against cancer is that, under the selective pressure of the immune system, tumor cells undergo changes that allow them to escape immune surveillance (cancer immunoediting). Moreover, cytotoxic T lymphocytes (CTL), that are the major component of the anti-tumor immune response, are known to loose efficacy when entering tumor microenvironment (CTL exhaustion). Mathematical modeling of CTL/tumor interaction must take into account these important constraints to immune surveillance in order to provide numerical data relevant to physiopathology. We present here a random dynamical particle interaction model of CTL/melanoma cell interaction that takes into account cancer immunoediting and CTL exhaustion. The model allows to test tunable parameters influencing the balance between CTL efficacy and increasing tumor cell resistance and provides estimation of the probabilities of tumor eradication.

Our results reveal that a bias in CTL motility that induces a progressive attraction of individual cells towards a few scout CTL that have detected the tumor, and the capacity of CTL to kill a large number of target cells before being exhausted, are crucial parameters in allowing tumor nodule eradication. Our results highlight unprecedented aspects of immune cell behavior the might inspire new CTL-based therapeutic strategies against tumors.

1. Ce chapitre fait l'objet d'une soumission, écrite en collaboration avec Sébastien Gadat, Salvatore Valitutti, Loic Dupré, Patrick Cattiaux. Les expériences biologiques ont été réalisées par Magda Rodrigues, Anne-Elisabeth Petit et Sabina Muller

## 1.1 Introduction

CTL destroy virally infected cells and tumor cells via the secretion of lytic molecules stored in intracellular granules [DL10]. CTL are key components of the anti-cancer immune response and it is therefore crucial to study in depth, and possibly enhance, their biological responses against tumors [AG13]. Accordingly, therapeutic protocols designed to potentiate CTL responses against tumor cells are currently at the frontline of cancer clinical research [CM13]. The molecular mechanisms of tumor recognition by CTL and the biological responses of CTL against tumors have been thoroughly investigated. However, since CTL/tumor cell interactions are highly dynamic, it is crucial to define the cell motility and interaction parameters that might influence CTL efficacy against tumor cells and tumor eradication.

Limited tumor site accessibility by CTL and intrinsic tumor cell resistance to CTL attack are major limits to CTL-mediated immune surveillance [Gaj+13]. Moreover, the tumor micro-environment can progressively affect CTL function, leading to CTL exhaustion [Cre+13]. Finally, it has been demonstrated that the adaptive immune response plays a dual role in cancer. While CTL can control tumor growth by destroying tumor cells, the selective pressure of the immune system promotes tumor progression by selecting tumor variants that are fit to survive in an immunocompetent host. Such a process is defined as cancer immunoediting [SOS11].

The design of immunotherapies that could bypass cancer immunoediting is presently a major conundrum in cancer clinical research. Experimental approaches aiming at recapitulating the complex evolving balance between CTL efficacy and the resistance of a multicellular tumor nodule are technically difficult to perform. Moreover, the integration of the parameters influencing this balance, in the context of hypothetical human tumors, leads to a very complex dynamical predator/prey system [Bab12].

Here we propose a mathematical model based on experimental measurements that provides a comprehensive view of the kinetic parameters of such a predator/prey system. We present an accurate random dynamical model and numerical simulations which describe the competition between spherically growing tumors and a clonal population of CTL. We focus on melanoma since in this neoplastic disease, tumor-associated antigens have been described and are known to elicit CTL-based immune responses that are counteracted by immunoediting processes in the tumor microenvironment [Boo+94; Spr+13]. Moreover various CTL-based adoptive transfer therapies are currently under evaluation for melanoma patients [KJ13].

We consider here only planar interactions. This is a simplification, yet it has the advantage of allowing a clear representation of cellular interactions and a rapid numerical implementation of the model in order to vary different parameters of CTL function and tumor growth either individually or simultaneously. Moreover the 2-D model allows to incorporate *in vitro* microscopy data that were acquired in 2-D. Based on biological

observations showing that the kinetic evolution of tumor growth and CTL displacement follow a random dynamics [Har+12; CPB08], we adopt a random modeling of CTL displacement, killing capacity and cancer immunoediting.

Our results show that in a context of randomly occurring immunoediting: i) CTL attraction towards scout siblings having detected the tumor and ii) CTL ability to kill a large number of target cells before becoming exhausted, are crucial parameters allowing early productive CTL/tumor collisions and tumor eradication.

## 1.2 Description of the model

The model presents a random evolution of tumor growth, immunoediting, CTL displacement and killing capacity and uses multiple simulations to compute the probability of success/loss in tumor eradication (Table 1.1 shows the parameters used in the model, see at the end of this section).

### 1.2.1 Cell behaviours

Before a sharp description of each theoretical evolution, we first enumerate the several randomized behaviors taken into account in our model.

- Since we aim to describe the competition between the tumor nodule and the CTLs, we consider a longitudinal section of the tumor nodule evolving on a compact two-dimensional domain $\Omega$ in parallel with a population of motile CTLs.
- Each CTL evolves according to a random evolution in the domain and is reflected on its boundary. This simplification is legitimate since it describes the fact that even if some CTL can exit the domain, some other ones can also reach. Hence, this shortcut roughly means that we will consider a population of CTL of constant size. When a CTL encounters a tumor cell, it automatically stops its walk over and starts the killing process of the tumor cell, which is also random according to our model.
- All along the computational time, a CTL becomes "exhausted" within the tumor environment when it reaches a maximal number of killing rounds (5, 10, 20 tumor cells etc.). Hence, an exhausted CTL can kill no more tumor cells. Exhausted CTL also evolve in $\Omega$ according to a random walk, bouncing back both at the frontier of the domain and when contacting the tumor mass. CTL exhaustion is irreversible.
- The tumor is a nodule of 2 different types of tumor cells: the duplicative ones constitute the proliferative shell of actively dividing cells, and the central core of non-proliferative cells (See Figure 1.1A). This concentric (but not necessarily

exactly spherical structure) mimics the structure of tumor clusters as defined by confocal laser scanning microscopy (See Figure 1.1B). Without any intervention of any CTL, the proliferative shell will follow a randomized growing branching process.

– In the model we implement the possibility that, after each division, tumor cells could acquire mutations that might make them progressively more resistant to CTL attack (as a part of an immunoediting process) or even invisible to CTL (mimicking MHC Class I loss described in tumor cells during immunoediting, [SOS11]). Hence, for the acquisition of resistance, daughter tumor cells can become stochastically and progressively more resistant than its mother cell.

### 1.2.2 Discretization of the time and of $\Omega$

CTL and tumor cells are spatially parameterized by their coordinates $(x, y)$ on a discrete grid (See Figure 1.1A), where the step sized movement $\Delta$ of each CTL is

$$|(x, y) - (x + 1, y)| = |(x, y + 1) - (x, y)| = 12.5\mu m = \Delta.$$

It corresponds to the mean diameter of a melanoma cell as measured by confocal laser scanning microscopy.

The evolution will be discretized according to a time step $\delta_t$, so that the time sequence $(t_n)_{n\geq 0}$ is equally spaced: $\delta_t = t_n - t_{n-1}$. The time step $\delta_t$ is strongly related to the velocity of CTL, since it is the time necessary for one CTL to move from its position to one of its 1-neighborhood. In order to obtain an immediate velocity of each CTL that corresponds to the one observed on real data, we fix the time-step $\delta_t$ to be the mean time necessary for one CTL to cover a distance that corresponds to $\Delta$. The mean recorded velocity is about $\boldsymbol{v} = 8.66\mu m/min$, then $\delta_t$ is fixed to $1.44min$.

### 1.2.3 Tumor nodule growth

In the proliferative shell, tumor cells are divided with a constant rate $\lambda$. It traditionally yields a continuous exponential distribution for changing state, in our case, to move from one cell to two cells. According to our time discretization, such a probabilistic distribution need to be slightly modified. The distribution of the division time is thus converted to a randomly geometrical time $\mathcal{G}(\lambda^*)$ which is the equivalent distribution for discrete time. The average measured time of duplication of melanoma cells in vitro is 16.66hours (1000 minutes). This is used to fix the tumor division rate parameter via $\lambda^* = \lambda \cdot \delta_t$ where $\frac{1}{\lambda} = 1000$. The model assumes that the probability for

a tumor cell $i$ to divide at a time $t_n$ is

$$P(T^i_{Div} = t_n) = \lambda^*(1-\lambda^*)^{n-1}.$$

The tumor nodule growth is also described by the parameter $E$ namely the thickness of the proliferating shell. It means that per unit time, the dividing rate of tumor cells in the proliferative shell (corresponding to the cells for which the distance from the nodule frontier is lower than $E$ ) is $\lambda^*$. Each division yields the addition of a new tumor cell. Two cases are possible:

1. Some free location is available in the immediate neighborhood ($\sqrt{2}$ -neighborhood) of the mother cell and the new cell chooses its place among all the free neighborhood;
2. The new cell chooses an occupied location in its $\sqrt{2}$-neighborhood and pushes another cell located in the proliferating shell towards the exterior of the nodule.

Both cases of such dynamics are illustrated in Figure 1.1 *A*. After each division event, the proliferative shell is instantaneously updated.

**Immunoediting** As pointed above, tumor cells could acquire mutations and then become resistant to CTL lysis. After each duplication of a tumor cell, a daughter cell can become progressively more resistant with a rate $p_{res}$. In this case CTL will kill daughter cells within a time longer than that required to kill mother cells. The model postulates several levels of resistance $R \in \{1, 2, 3, \ldots\}$. A tumor cell with $R = 1$ is a tumor cell which did not increase its resistance upon division. A tumor cell with $R_d = R_m + 1$ is a tumor cell which increases its resistance upon one division ($R_d$ is the resistance of the daughter cell, $R_m$ is the resistance of the mother cell). As a result, each tumor cell acquiring resistance upon division requires on average twice as long as its mother cell to be killed. In the model, a tumor cell becomes invisible to CTL with the rate $p_{inv}$. See Table 1 for the values of $p_{res}$ and $p_{inv}$.

### 1.2.4 CTL function

**Movement of the CTL** The number of CTL ($N_{CTL}$) is fixed for each evolution, assuming that cells entering the field replace CTL leaving the field or dying. For CTL displacements, positions are parameterized by their planar coordinates at time $t_n$: $z^i_n := (x^i_n, y^i_n)$, for $i = \{1, ..., N_{CTL}\}$ and these coordinates still belong to the regularly spaced grid. Let $n^i_{free}$ be the number of free 1-neighbors of the CTL $i$, without other interaction with tumor nodule or another CTL, each CTL evolves according to a symmetric random walk moving from one time to another to its $n^i_{free}$ free 1-neighbors (see Figure 1.1C). Equation (1.1) describes the displacement probability of each CTL

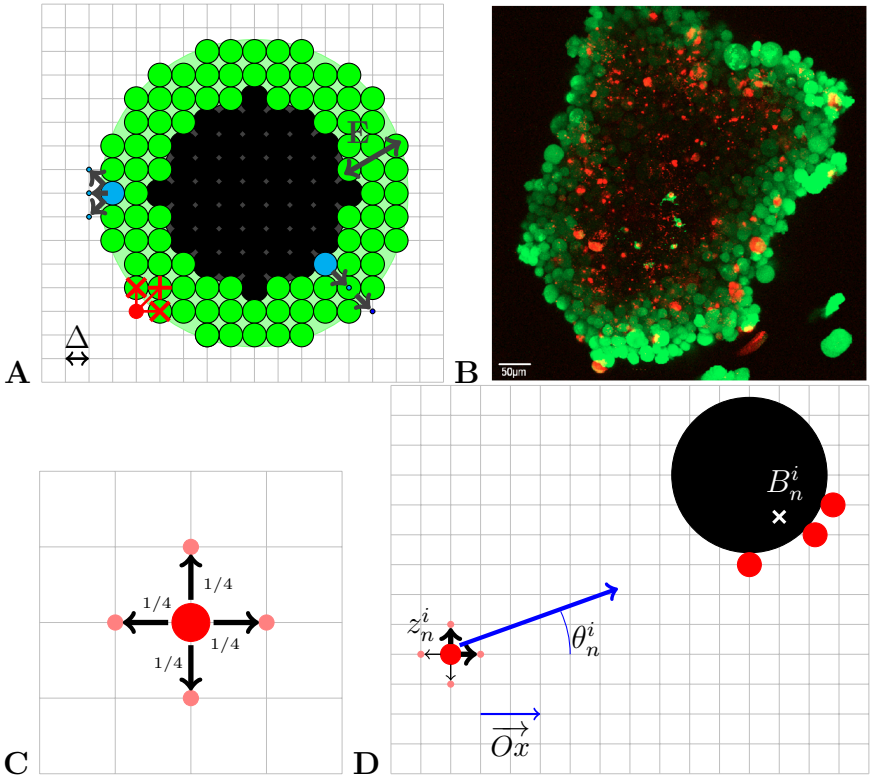



Figure 1.1: **Graphical representation of the model.** $(A)$ Synthetic planar structure of the nodule, fitted to the grid. Green: cells in the proliferative shell. Black: non proliferative and necrotic cells. Red: CTL killing its $\sqrt{2}$-neighbourhood tumor cells. Blue: two tumor cells attempt to divide. On the left, a tumor cell is dividing towards one of the three free available positions (black arrows). On the right, there is no free location available at the time of division, the dividing cell pushes its close neighbours towards the exterior of the nodule. $(B)$ Picture of a melanoma nodule. Green: proliferative cells, red: dead cells. $(C)$ Random Walk on the 2D grid of one CTL. Red: position of the CTL. Black arrows: admissible displacements. $(D)$ One CTL at position $z_n^i$ is attracted by 3 scout sibling CTLs (in red) that have hit the tumor nodule (black sphere). The barycenter $B_n^i$ is the target position for $z_n^i$. The random walk of CTL on position $z_n^i$ is reinforced towards $B_n^i$ (blue arrow), please see material and methods section for further details.

$i$ in the grid if one 1-neighbors are free, let $p_1, .., p_{n^i_{free}}$ be the $n^i_{free}$ free positions

$$P\left[z^i_{n+1} = p_j\right] = \frac{1}{n^i_{free}}, \quad \text{where } j \in \{1, ..., n^i_{free}\}. \tag{1.1}$$

Other wise, if there is no free 1-neighbor, the CTL remains on its place.

**CTL killing** A single CTL detects the nodule as follows: when one CTL hits the frontier of the nodule, the CTL stops instantaneously and kills the tumor cells in its immediate $\sqrt{2}-$neighborhood. A tumor cell are also killed according to a random

geometrical time. The mean measured time in minutes for detection of apoptosis induction in melanoma cells in vitro [Car+09] is $\frac{1}{\mu} = 26.4min$. To obtain the inverse of the mean time required for killing one target cell according to the discretized time used in simulations it is necessary to compute $\mu^* = \mu \cdot \delta_t$. Hence, the probability that a CTL $i$ kills a tumor cell is

$$\mathbb{P}(T^i_{kill} = t_n) = \mu^*(1-\mu^*)^{n-1}.$$

After one killing step of a tumor cell, one CTL restarts a symmetric random walk according to the rules defined in 1.1. A CTL becomes "exhausted" when it reaches a maximal number of killing rounds denoted $\kappa$ in the sequel.

### 1.2.5 Biased displacement for CTL

In the above describtion, the model postulated a scenario in which CTL are not preferentially directed towards the growing tumor. In these conditions, only CTL randomly colliding with a cognate tumor cell exhibit cytotoxicity, then only a low percentage of CTL participate to the killing. As a consequence, the CTL population, in its whole, mostly ignores the growing tumor. In the attempt to optimize the number of CTL collisions with tumor cells, we introduced a bias for CTL movement. The updated model postulates that when a CTL collides with the tumor nodule, it releases chemo-attractants that guide the trajectories of other CTL towards the nodule (see Figure 1.1$D$ for graphical illustration).

We thus consider a self-interacting particle system, inspired from population dynamics already described in other biological fields (e.g. ant colonies [YYS13]). The principle can be described in this way: when one (or several) CTL hits the boundary of the nodule, other CTL may be attracted to the position of this scout CTL if they are located close enough (the distance between scout cells and attracted ones should be less than a parameter $D$). In these conditions, the model postulates that a given CTL $i$ (whose position $z^i_n$ does not belong to the frontier of the nodule at time $t_n$) is attracted towards the barycenter of the scout CTL $B^i_n$ (see Figure 1.1$D$). The attraction towards the barycenter is provided by the global strength parameter $\nu^i_n$ that belongs to $[0, \nu]$. This parameter $\nu^i_n$ depends on the distance between $B^i_n$ and $z^i_n$, and on the number of scout CTL, denoted $K^i_n$. The model postulates a fixed maximal scout CTL number $K$ (corresponding to 10 CTL) and a maximal attraction strength $\nu$. To define the impact that the attraction of CTL towards the tumor nodule might have on tumor eradication, we let vary the parameter $\nu$ in the different numerical simulations. Finally, we use the

following formula

$$\nu_n^i = \left(1 - \frac{|z_n^i - B_n^i|}{D}\right) \times \min\left\{1; \frac{K_n^i}{K}\right\} \times \nu, \tag{1.2}$$

which stands that the maximal attraction number is reached when $K_n^i = K = 10$ CTL and the attracted CTL $z_n^i$ is very close to the barycenter $B_n^i$.

Let $n_{free}^i$ be the number of free 1-neighbors of the CTL $i$. Thus, the position of $z_{n+1}^i$ follows the mixture law

$$1 - \nu_n^i \cdot \textit{symmetric random walk} \quad + \quad \nu_n^i \cdot \textit{ drifted attraction towards } B_n^i,$$

conditioned to $n_{free}^i \neq 0$, otherwise the CTL $i$ remains on its place. The drift is defined by $\theta_n^i$ the planar angle between the two vectors $\overrightarrow{B_n^i z_n^i}$ and $\overrightarrow{Ox}$ (as shown in Figure 1.1D). If $n_{free}^i \neq 0$, let us denote $\epsilon(a)$ the algebraic sign of any real number $a$, then the probability to go from $z_n^i$ to 1-neighbor, namely $q_i$, is

$$\begin{aligned}
q_1 &= P\left[z_{n+1}^i = z_n^i - (\epsilon(\cos\theta), 0)\right] = \frac{\frac{1-\nu_n^i}{4}}{q_1 + ... + q_4} \\
q_2 &= P\left[z_{n+1}^i = z_n^i - (0, \epsilon(\sin\theta))\right] = \frac{\frac{1-\nu_n^i}{4}}{q_1 + ... + q_4} \\
q_3 &= P\left[z_{n+1}^i = z_n^i + (\epsilon(\cos\theta), 0)\right] = \frac{\frac{1-\nu_n^i}{4} + \nu_n^i \cos(\theta)^2}{q_1 + ... + q_4} \\
q_4 &= P\left[z_{n+1}^i = z_n^i + (0, \epsilon(\sin\theta))\right] = \frac{\frac{1-\nu_n^i}{4} + \nu_n^i \sin(\theta)^2}{q_1 + ... + q_4}.
\end{aligned} \tag{1.3}$$

### 1.2.6 Random dynamic agent based model or differential equations?

Computational works on biological dynamical systems consider either infinitesimal agent based model described by a randomized particles system or partial differential equations that describe the chronological evolution of a macroscopic number of interest (density of CTL, size of the tumor nodule,...). Indeed, these two alternative ways of describing the biological system are generally equivalent. On the one side it is easier to amend the dynamical model with particles systems owing to its flexibility, on the other side, differential equations are easier to handle from a mathematical point of view.

Let us first discuss on our choice of dynamical agent based model. First, agent based models are slightly more natural than differential approaches since they mimic more closely the nature of the real phenomenon. The proportion density of the CTL has no real existence although CTLs are positioned into the systems and evolve randomly.

In this work, we aim to describe a possible interacting effect in the population of CTL due to chemotactism for a specific subpopulation of CTL. This phenomenon can be handled with random particles without any difficulty, by the addition of a drift to the symmetric random walk described by (1.1). By no means this chemotactism can be described so easily with P.D.E., recent advances in this direction relies on Keller-Segel models (see for instance [KS71]) but these models are far from being trivially solved by P.D.E. Especially, these models needs to be enhanced to implement a chemotactism towards a specific zone of the population of CTL. We are currently working on an improvement of a Keller-Segel type description of our particular chemotactism with the view to obtain more simpler equations.

| | param | description | value |
|---|---|---|---|
| CTLs | $v^\star$ | CTL displacement velocity | $8.66\mu m/min$ |
| | $\mu^{\star o}$ | Inverse of the time required for killing one target cell | 0.055 |
| | $\kappa^o$ | Number of tumor cell killed by a single CTL | |
| | $N^o_{CTL}$ | CTL number at the beginning of the simulation | |
| Tumor | $\Delta^\star$ | Melanoma cell diameter | $12.5\mu m$ |
| mass | $L^\diamond$ | Melanoma nodule length at the beginning | $300\mu m$ |
| | $\lambda^{\star}$ | Inverse of the time required for tumor cell division | 0.00144 |
| | $E^\star$ | Thickness of the proliferative shell of the tumor mass | $2\Delta$ |
| | $p^\diamond_{res}$ | Probability of tumor cell to become more resistant to the CTL attack at each division | $\frac{1}{2}$ |
| | $p^\diamond_{inv}$ | Probability of tumor cell to became "invisible" to CTL at each division | $\frac{1}{2}\left(\frac{1}{2}-\frac{1}{res}\right)$ |
| Biased motility | $D^\diamond$ | Maximal distance allowed for CTL attraction (if distance between CTL and nodule $\geq \sqrt{D}$, CTL are not attracted) | $350\mu m$ |
| | $K^\diamond$ | Number of scout CTLs generating a maximal attraction | 10 |
| | $\nu^o$ | Maximal level of attraction | |

Table 1.1: **Parameters used in simulations.** $^\star$ estimated parameters, $^o$ varying parameters, $^\diamond$ arbitrary parameters.

## 1.3 Parameter estimations

In this section we detail the estimator used to compute the parameters indicated by a star in Table 1.1. For parameters $v$ and $\Delta$, mean of the observation is used, we give a sharp description of the estimator used for parameter $\mu^*$, $\lambda^*$ and $E^*$.

### 1.3.1 Estimation of $\mu$: killing rate

We compute an estimation of $\mu$ in an independent context of our random dynamical model. Experimental measurement of the time required for melanoma cell killing by CTL were employed to compute $\mu$ estimation [Car+09]. For a number of $N$ observations, we observe several lysis duration $(t^i_{lysis})_{i\in\{1,...,N\}}$, and standard estimation on exponential dustibution yields an estimator of the rate $\hat{\mu}_N^{-1}$, which is the inverse of the mean time of lysis. To obtain the rate according to our time step $\delta_t$, we then compute $\hat{\mu}^*_N$ as

$$\hat{\mu}^*_N = \frac{1}{\hat{\mu}_N}\delta_t = N\left(\sum_{i=1}^{N} t^i_{lysis}\right)^{-1} \times \delta_t.$$

### 1.3.2 Estimation of $\lambda$: division rate

The estimation of the division rate $\lambda$ of tumor cells is slightly more complex than the estimation of $\mu$ owing to the nature of available biological observations. Measurements were performed in culture conditions in which tumor cells grew in 2-D without forming cellular spheroids. In these conditions, the replication of the tumor cells evolves without any geometric constraints. The number of cells was kept fixed at time 0 (denoted $n_0$) and counted at time $T$, so that we obtained this final number $n_T$. Value of tumor cell growth obtained in $N$ individual culture plates corresponds to the observation of $(n^1_T, ..., n^N_T)$.

The model can be seen as a Pure Birth Process (Yule-Furry process) from a probabilistic point of view: when one tumor cell is replicated, the population increases of one individual cell. If we denote $X_t$ the number of cells at time $t$, the probability that $\{X_t = n\}$ given by $P_n(t) = \mathbb{P}(X_t = n)$ satisfies

$$P_n(t+h) = P_n(t)\,(1-\lambda n h) + P_{n-1}(t)\,(n-1)\,\lambda h + o(h).$$

Since the probability that one birth occurs during the interval $[t, t+h]$, up to the condition $\{X_t = n\}$, is $n\lambda h + o(h)$. Then $(1 - n\lambda h) + o(h)$ is the probability that no division appears during $[t, t+h]$ if $\{X_t = n\}$. Then $(P_n(t))_{t\geq 0, n\in\mathbb{N}}$ satisfies the differential system

$$P'_n(t) = -n\lambda P_n(t) + (n-1)\lambda P_{n-1}(t). \tag{1.4}$$

Since the starting number of cells is $n_0$, we have *i.e.* $P_{n_0}(0) = 1$, and for all $n \neq n_0, P_n(0) = 0$. One may check that the unique solution of (1.4) is given by

$$P_n(t) = C^{n-n_0}_{n-1} e^{-\lambda n_0 t}(1 - e^{-\lambda t})^{n-n_0}. \tag{1.5}$$

Hence, $X_t$ follows a negative binomial distribution $\mathcal{NB}(n_0, e^{-\lambda t})$ initialised with $n_0$ cells

and a succeed parameter $e^{-\lambda t}$.

We aim to estimate $\lambda$ and by denoting $p = e^{-\lambda T}$, we can use the Maximum Likelihood Estimation (M.L.E.) to compute $\hat{p}_N$, and then $\hat{\lambda}_N$. Using our observations $(n_T^1, \dots n_T^N)$, the log-likelihood is given by

$$l(n_0, p, n_T^1, \dots n_T^N) = \log\left(\prod_{i=1}^{n} P_{n_T^i}(T)\right).$$

We use now (1.5) to obtain

$$\begin{aligned} l(n_0, p, n_T^1, \dots n_T^N) = \sum_{i=1}^{N} \log((n_T^i + n_0 - 1)!) - \sum_{i=1}^{N} \log(n_T^i!) - N\log((n_0 - 1)!) \\ + N n_0 \log(p) + \sum_{i=1}^{N} n_T^i \log(1-p). \end{aligned}$$

We can optimize the last expression with respect to $p$ to obtain that

$$\hat{p}_N = \arg\max_{0 \le p \le 1} l(n_0, p, n_T^1, \dots n_T^N) = \frac{n_0}{n_0 + \sum_{i=1}^{N} \frac{n_i^T}{N}}.$$

Since we have set $p = e^{-\lambda T}$, we deduce that $\hat{\lambda}_N$ is given by

$$\hat{\lambda}_N = T^{-1} \log\left(1 + \sum_{i=1}^{N} \frac{n_T^i}{N n_0}\right).$$

As in the estimation of $\mu$, to obtain the rate according to time step, we then compute $\hat{\lambda}_N^* = \hat{\lambda}_N \times \delta t$.

### 1.3.3 Estimation of $E$: thickness of the proliferative part

The size of the proliferative part is an important feature for the growth of the tumor nodule. Hence, $E$ should be carefully estimated and we assume in the sequel that $E$ is constant over the time evolution of the nodule. In order to fix a plausible value of $E$, the global diameter of the nodule was observed over the time. We denote $(d_t)_{0 \le t \le T}$ such a diameter from the initial time $t = 0$ to the ending one $t = T$ (see Figure 1.3).

Again, the number of cells in the proliferate shell can be included in a stochastic paradigm of Birth & Death process. We denote $N_t$ the number of cells in the proliferate shell at time $t$ and $Q_n(t) = \mathbb{P}(N_t = n)$, and we aim to obtain a differential equation similar to (1.4) which takes into account the nodule structure as well as the size of several shells. The planar structure is given in Figure 1.1 $A - B$ and the situation is summarized in the synthetic Figure 1.2.

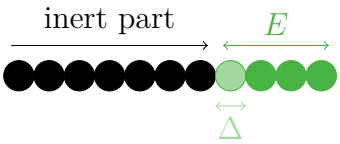



Figure 1.2: **Schematic representation of the inert and proliferative part of a tumor nodule.** The proliferative part is of size $E$, each cell has a size $\Delta$.

In particular, the probability for any chosen cell to belong to the immediate neighbourhood of the inert part is $\frac{\Delta}{E}$ according to Figure 1.2, where $\Delta$ is the size of individual tumor cells (this last approximation is valid as soon as the radius of the tumor is somewhat larger than $E$). Moreover, $n\lambda h$ is the probability that one cell of the proliferative part is divided over $[t, t+h]$ up to the condition that $N_t = n$, and if one division occurs, such a probability should be balanced by $\left(1 - \frac{\Delta}{E}\right)$ (probability that the chosen divided cell does not belong to the frontier of the inert part). This last probability is important since when the chosen cell is near to the non-proliferative shell, one of the two daughter cells enters directly in the non-proliferative shell and in such a case, the number $N_t$ is kept constant. Similarly, the probability to obtain one birth in $[t, t+h]$ up to the condition $\{N_t = n-1\}$ is $(n-1)\lambda h\left(1 - \frac{\Delta}{E}\right) + o(h)$ and we obtain the differential infinitesimal equation:

$$\begin{aligned} Q_n(t+h) = Q_n(t)\left[(1 - n\lambda h) + n\lambda h\frac{\Delta}{E}\right] \\ + Q_{n-1}(t)(n-1)\lambda h\left(1 - \frac{\Delta}{E}\right) + o(h). \end{aligned} \tag{1.6}$$

We can thus deduce the dynamical evolution of $Q_n$:

$$Q_n'(t) = (n-1)Q_{n-1}(t)\tilde{\lambda} - nQ_n(t)\tilde{\lambda},$$

where $\tilde{\lambda} = \lambda\left(1 - \frac{\Delta}{E}\right)$. To estimate $E$, it is sufficient to obtain $\tilde{\lambda}$ and an estimation of $\tilde{\lambda}$ can be easily obtained through the first moment of $N_t$. More precisely, let $\Gamma(t) = \mathbb{E}(N_t)$ the mean number of cells in the proliferative part. Using (1.6), we get

$$\begin{aligned} \Gamma'(t) &= \sum_{n\geq 1} nQ_n'(t) \\ &= \sum_{n\geq 1}(n-1)^2 Q_{n-1}(t)\tilde{\lambda} + \sum_{n\geq 1}(n-1)Q_{n-1}(t)\tilde{\lambda} - \sum_{n\geq 1} n^2 P_n(t)\tilde{\lambda} \\ &= \sum_{n\geq 0} n^2 Q_n(t)\tilde{\lambda} + \sum_{n\geq 0} nQ_n(t)\tilde{\lambda} - \sum_{n\geq 1} n^2 Q_n(t)\tilde{\lambda} \\ &= \Gamma(t)\tilde{\lambda}. \end{aligned} \tag{1.7}$$

The solution of Equation (1.7) is

$$\Gamma(t) = \Gamma(0)e^{\tilde{\lambda}t}. \tag{1.8}$$

The mean number of cells $N_t$ in the tumor proliferation part is of course not observed over the time, but we can use the observed sequence $(d_t)_{0\leq t\leq T}$ of the diameter of the nodule to obtain $\Gamma(t)$, since $\Gamma(t)$ is related to $d_t$ by the following formula

$$\Gamma(t) = \underbrace{\frac{d_t^2}{\Delta^2}}_{\text{number of cells in the nodule}} - \underbrace{\frac{(d_t/2-E)^2}{(\Delta/2)^2}}_{\text{number of cells in the inert part}} = \frac{4E(d_t-E)}{\Delta^2}. \tag{1.9}$$

Using this last equation at time $t = 0$ and $t = T$, and using (1.9) in (1.8), $E$ should satisfy:

$$\frac{4E(d_t-E)}{\Delta^2} = \frac{4E(d_0-E)}{\Delta^2}e^{\tilde{\lambda}t} \iff E = \frac{\Delta\lambda t}{\lambda t - \log(d_t-E) + \log(d_0-E)}. \tag{1.10}$$

It is then possible to numerically solve this equation, but as we can only consider entire parts of cell sizes, we choose $E^* = \Delta\left\lceil\frac{E}{\Delta}\right\rceil$, where $E^*$ is the size of the proliferative part in our simulations. We obtained with our data $E^* = 2\Delta$. As a goodness of fit testing for our value $E^*$, Figure 1.3 compares the experimental observed diameter with the evolution of the theoretical diameter $(d_t)_{0\leq t\leq T}$ obtained with (1.9) (for $E^* = \Delta$ and $E^* = 2\Delta$), for all $0 < t < T$, $d_t = \frac{\Delta^2\Gamma(t)}{4E} + E$.

### 1.3.4 Experimental parameters measurement

The mean velocity of using antigen specific CTL clones displacement (8.66 $\mu m/min$) was measured by time-lapse microscopy using a confocal laser-scanning microscope (either a Zeiss LSM-510 or a Zeiss LSM-710 microscope, Zeiss Germany). Measurement were performed on non-stimulated CTL that were free of moving on poly-D-lysine coated LabTeck-chambers.

The mean time required for killing a melanoma cells was estimated by measuring in a significant number of CTL/target cell conjugates the time occurring between the initial CTL/target cell contact and the beginning of blebbing in target cells, as reported in I. Caramhalo et al. by tiem-lapse confocal microscopy [13]. Experimental measurement of the mean time required for melanoma cell killing by CTL (∼26 minutes) were employed to compute $\mu$ estimation.

Melanoma cell mean diameter was estimated on paraformaldehyde fixed melanoma cell spheroids using a confocal microscope.

The time required for melanoma cell division was measured in standard melanoma cell 2-D cultures by counting the cell number at fixed time intervals over a total period of three days.

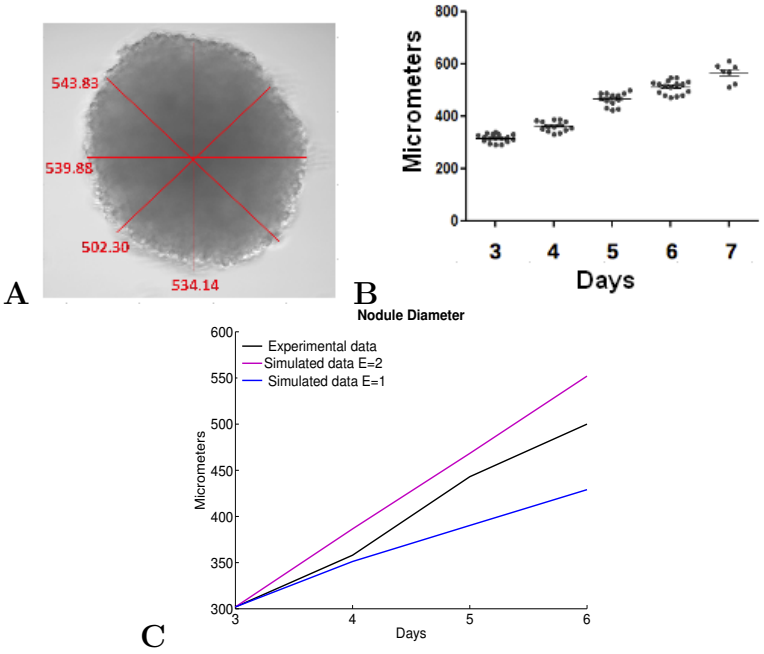



Figure 1.3: **Measurement of the diameter of tumor nodules over time.** $(A)$ and $(B)$ Experimental results. $(C)$ Comparison of experimental measurements of the diameter of nodules over time (black line) and simulated measurement of nodules over time with $E = 2$ tumor cells (magenta) and with $E = 1$ tumor cell (blue).

Melanoma cell spheroids were generated using the 'hanging-drop' method. Briefly 1000 cells/25$\mu l$ were seeded in the wells of a Terasaki culture plate. The plate was then inverted to form suspended droplets. The measurement of the diameter of the spheroids was performed using a Zeiss LSM-710 microscope using a 20x objective. The mean diameter of a spheroid was calculated by calculating the mean value of 4 diameters measured using the LSM software (Zeiss). To visualize the alive/dead fraction of cells in the spheroid at day 6 of culture the spheroids were incubated for 2 hours with 5nM Ethidium-1-HomoDimer (Eth-1-HD, red) and 1 $\mu$ M calcein (green) at 37oC. Staining was visualized using a Zeiss LSM-710 using and 20x ojective.

## 1.4 Algorithms

The goal of this section is to present an algorithm allowing to perform numerical simulations of the model detailed in section 1.2. First, we describe the general algorithm, then we detail functions created in order to implement it.

### 1.4.1 General algorithm

First of all, we give definition of terms used in the algorithm to characterize cells:

– **invisible tumor cell** = tumor cell with the highest level of resistance, they are invisible to CTL, they cannot be killed by a CTL;
– **exhausted CTL** = CTL which has killed $\kappa$ tumor cells;
– **active CTL** = non-**exhausted CTL** touching visible tumor cells;
– **available place** = place which is not occupied by any cells.

**Algorithm I.1:** general algorithm

**Input**: parameter liste
**Initialisation:**- nodule formation - random positioning of CTL - define CTL which are *active_CTL* at time 0;
**for** $t$ = *current time (3 days)* **do**
    CTL functions;
    **for** *each CTL* **do**
        **if** *CTL are not active* **then**
            **Function:** CTL displacement
        **else**
            **Function:** Active CTL killing
        **end**
    **end**
    Exhausted CTL function;
    **for** *each exhausted CTL* **do**
        **Function:** Exhausted CTL displacement;
    **end**
    Tumor functions;
    **for** *each tumor cell* **do**
        **Function:** Tumor cell division
    **end**
    **for** *each invisible tumor cell* **do**
        **Function:** Invisible tumor cell division
    **end**
**end**

Each cell is defined by its **status**: active CTL, exhausted CTL, CTL (if it is not active or exhausted), tumor cell (called sometime by visible tumor cell) or invisible tumor cell, and by its **position** in the grid (or coordinate). According to its status a cell is also defined by its **characteristic**. CTL and Active CTL are characterized by the number of tumor cells already killed $\boldsymbol{k}$, and tumor cells are characterized by its resistance $\boldsymbol{R}$. All this information are denoted by cell **data**.

### 1.4.2 Function created for the general algorithm

Let $\boldsymbol{i}$ be the loop parameter. Next functions are described for the cell $i$.

**CTL functions**

A CTL moves (as a biased or symmetric random walk according to there is or not attraction). An active CTL has a killing function.

**Algorithm I.2: Function:** CTL displacement

**Input**: $i$, CTL data, active CTL data, exhausted CTL data, tumor cell data, invisible tumor cell data

**if** *There is at least one available place on its 4 neighbors* **then**

1. Without attraction;
   Choose uniformly from available places, one of them;
2. With attraction;
   Define the distribution describing the biased: computing angle $\theta^i$ and attraction strength defined in (1.2), for each available place;
   Choose one of them according to the distribution given in (1.3)

**else**

Remains on its place

**end**

**Output**: Active CTL data

Recall $\kappa$ the maximal number of tumor cells that one CTL can kill, $R$ the tumor cell resistance and $k$ the tumor cell number yet killed by the CTL. Let $\boldsymbol{nb_{kill}}$ be the number of cell killed at the same time by the CTL $i$. We assume $nb_{kill} \leq 3$, since it seems difficult that a CTL can kill more than 3 target cells simultaneously.

**Algorithm I.3: Function:** Active CTL killing

**Input**: $i$, active CTL data, tumor cell data

$nb_{kill} := 0$;

**for** *Each tumor cell in contact with active CTL* $i$ **do**

$X \sim \mathcal{B}(\mu^*/R)$;

**if** $[\kappa - k > 0]$ *and* $[nb_{kill} < 3]$ *and* $[X = 1]$ **then**

Delete the tumor cell;

$k \leftarrow k + 1$;

$nb_{kill} \leftarrow nb_{kill} + 1$

**end**

**end**

**if** $\kappa - k = 0$ **then**

active CTL $i$ becomes exhausted CTL

**end**

**Output**: Tumor cell data, active CTL data, exhausted CTL data

**Exhausted CTL functions**

An exhausted CTL has only one function: move according to a symmetric random walk.

**Algorithm I.4: Function:** Exhausted CTL displacement

**Input**: $i$, exhausted CTL data, active CTL data, CTL data, tumor cell data, invisible tumor cell data

**if** *there is at least one available place on its 4 neighbors* **then**
  choose uniformly from available places, one of them
**else**
  remains on its place
**end**

**Output**: Exhausted CTL data

**Tumor cell function function**

A tumor cell has only one function, to proliferate. But, only tumor cell (visible or invisible) in the proliferative part, close enough to the edge of the nodule, can duplicate. In addition, a visible tumor cell in contact with an active CTL cannot divide. Hence we distinguish tumor cell invisible division from tumor cell visible division.

**Algorithm I.5: Function:** Tumor cell division

**Input**: $i$, tumor cell data, invisible tumor cell data, active CTL data, exhausted CTL data, CTL data

$X_1 \sim \mathcal{B}(\lambda^*)$;
**if** *[tumor cell $i$ is in the proliferative part] and [tumor cell $i$ is not close to an active CTL] and* $[X = 1]$ **then**
  Choose uniformly one place (preferentially available) in its neighborhood;
  If this place is occupied, translate cells affected by this modification
**end**
$X_2 \sim \mathcal{B}\left(p_{inv}\left(\frac{1}{2} - \frac{1}{R}\right)\right)$;
**if** $X = 1$ **then**
  tumor cell $i$ becomes invisible
**else**
  $X_3 \sim \mathcal{B}(p_{res})$;
  **if** $X = 1$ **then**
    $R \leftarrow R + 1$;
    tumor cell $i$ becomes more resistant
  **end**
**end**

**Output**: Tumor cell data, invisible tumor cell data, exhausted CTL data, active CTL data, CTL data

The duplication of invisible tumor cell is easier than the one of visible tumor cell, owing to the duplication conditions are more simple. In addition an invisible tumor cell is at the top of the tumor cell evolution.

**Algorithm I.6: Function:** Invisible tumor cell division

**Input**: $i$, invisible tumor cell data, tumor cell data, exhausted CTL data, active CTL position and CTL data

$X \sim \mathcal{B}(\lambda^*)$;

**if** *[invisible tumor cell $i$ is in the proliferative part] and* $[X = 1]$ **then**

  Choose uniformly one place (preferentially available) in its neighborhood;

  if this place is occupied, translate cells affected by this modification

**end**

**Output**: Invisible tumor cell data, tumor cell data, exhausted CTL data, active CTL data, CTL data

The general algorithm and the functions are implemented in MATLAB. As we see, the advantage of this model is an easy implementation, though taking account all rules entering in the model, and it is easy to amend these rules. The above mentioned assets allow a straightforward understanding of the relevance or not of parameters intering in the interaction.

## 1.5 Results

Having established the basic parameters and an algorithm of the model, we employed Monte-Carlo simulations model to estimate the probability of success of CTL in eradicating the tumor nodule. We compute this probability varying parameter such that the number of CTL $N_{CTL}$, the time required to kill one tumor cell $\mu^*$, the attraction strength $\nu$ in biased displacement case of the CTL and the number of targets, $\kappa$, that a CTL can kill before become exhausted.

### 1.5.1 Influence of early productive CTL/tumor cell collisions

In a first approach, we investigated the number of collisions between CTL and tumor cells that were required to grant tumor eradication. For this analysis we considered a set of numerical simulations performed by varying the number of CTL ranging from 600 to 1100. Numerical simulations showed that, for CTL success, it was important that a minimum number of CTL/tumor cell productive collisions would occur during the early time points of CTL/tumor cell dynamic confrontation. More precisely, in our mathematical simulation lasting 72 hours, the number of collisions during the first 5 hours (defined as early collisions) was significantly larger in the case of CTL success than in the case of CTL incapacity to eradicate the tumor (Figure 1.4 *A*). Figure 1.4 *A*

also shows that, the average number of early collisions corresponding to CTL inefficacy was ∼112 collisions/5 hours, while ∼137 collisions/5 hours corresponded to success in tumor eradication.

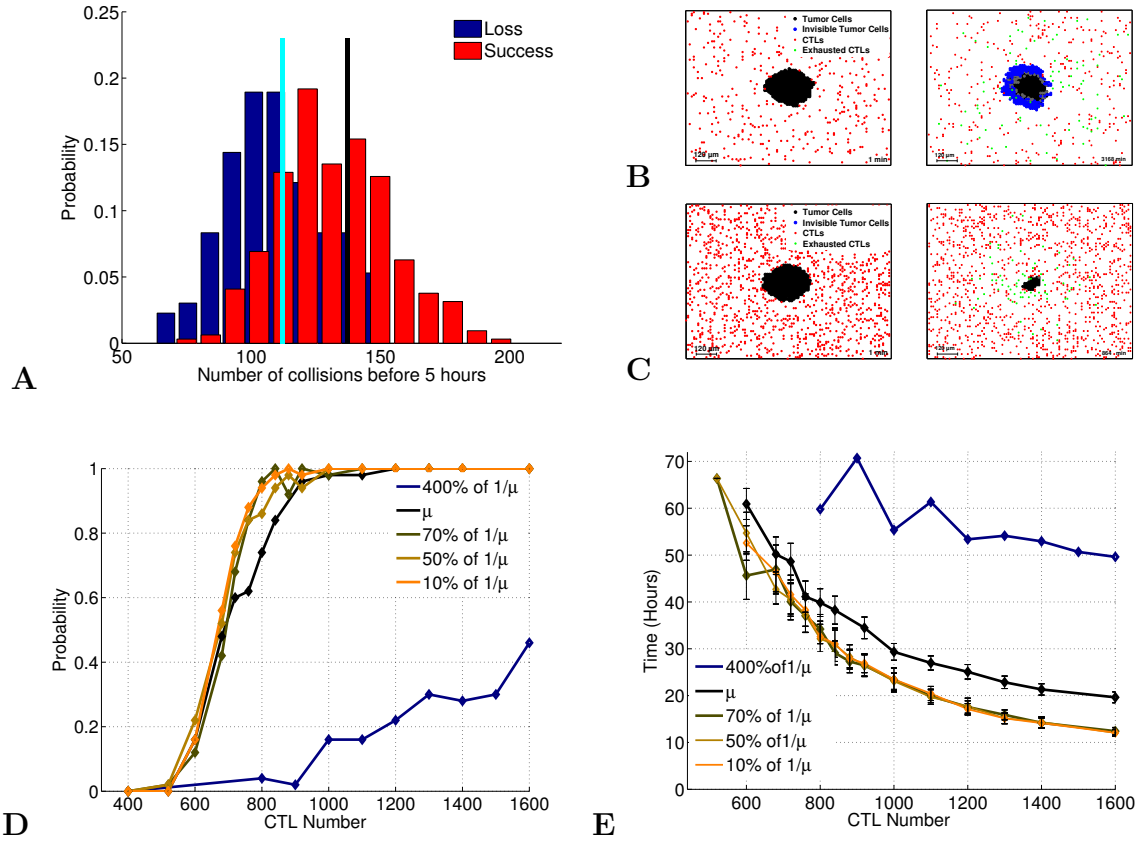



Figure 1.4: **The probability of tumor nodule eradication increases with the increase of CTL number.** $(A)$ Empirical distribution of the number of early collisions (CTL killing during the first 5 hours) in the lost scenario (blue) and its respective mean number of collisions (cyan bar). Same computations in the case of victory scenario (red) and its respective mean number of early collisions (black bar). $(B)$ Snapshots taken at times $t = 1min$ and $t = 3168min$ of the evolution of the CTL population as well as the size of the nodule for 400 CTLs. A large amount of tumor cells become invisible (blue). $(C)$ Snapshots taken at times $t = 1min$ and $t = 864min$ showing the evolution of the CTL population as well as the size of the nodule for 1200 CTLs. Invisible tumor cells have not been generated and the CTL population eradicates the tumor nodule. $(D)$ Estimation of the victory's probability of the CTL population in tumor nodule eradication. Black line: estimated probability with respect to the number of CTLs whithout any modification of the mean killing time $1/\mu$. Blue line: the mean killing time is augmented of 4 times. Other lines: same evolution when the mean killing time is decreased by $10\%, 50\%$ and $90\%$. $(E)$ Mean time needed to eradicate the tumor nodule for the CTL population, results are presented as means $\pm$ standard deviation of 50 numerical simulations.Black line: estimated time whithout any modification of the mean killing time (other lines: with increase or decrease of the mean killing time, as indicated in $(D)$).

Taken together the above results indicate that in a cancer immunoediting scenario, the larger the number of early collisions between CTL and target cells, the higher the probability of tumor eradication.

### 1.5.2 Influence of population size

In order to define the parameters that might ensure a sufficient number of early collisions leading to tumor eradication, we initially varied the size of the CTL population. We varied the CTL number for two reasons. First, because it has been previously described that the number of available killer cells is a crucial parameter in defining the success of a killer cell population over tumors [Bud+10]. Second, because it has been thoroughly experimentally demonstrated that the capacity to eliminate tumor target cells increases with the increase of the size of the CTL population. In line with these reported data, our numerical simulations showed that the probability of success in tumor nodule eradication was improved with the increase of CTL number while the mean time required for nodule eradication decreased (Figure 1.4 $B - E$ ).

To investigate if the increase in the number of killer cells has an impact on the number of early CTL/target cell productive collisions, we measured the distribution of the number of collisions during the first 5 hours under different conditions. This analysis showed that when the number of CTL was increased from 400 to 1600 a sharp increase of the productive early collisions was observed (Figure 1.8 $A$, compare black and cyan histograms, mean values of 80 collisions/5 hours versus 186 collisions/5 hours).

An alternative way to graphically represent the impact of CTL number on the probability of success in tumor nodule eradication is shown in Figure 1.5 in which three simulations performed with high (1200, A), intermediate (400, B) and low (200, C) CTL number are presented. The plots represent the number of cells in the tumor nodule, the number of killed tumor cells, the number of exhausted CTL and the number of invisible tumor cells over time.

They show that, while at high and low CTL number the balance CTL success and tumor resistance is rapidly in favor of CTL or of tumor respectively (A and C), at intermediated cell numbers a more complex behavior is observed (B). A first phase in which CTL manage to reduce tumor cell number is followed by a period of equilibrium in which the size of tumor nodule remains stable. These phases are followed by a phase in which the number of tumor cells starts to grow in parallel with the random generation of tumor cells not visible by CTL. Interestingly, these three phases in tumor/CTL confrontation observed at intermediate CTL number are reminiscent of the three phases described in cancer immunoediting: elimination, equilibrium and escape [SOS11].

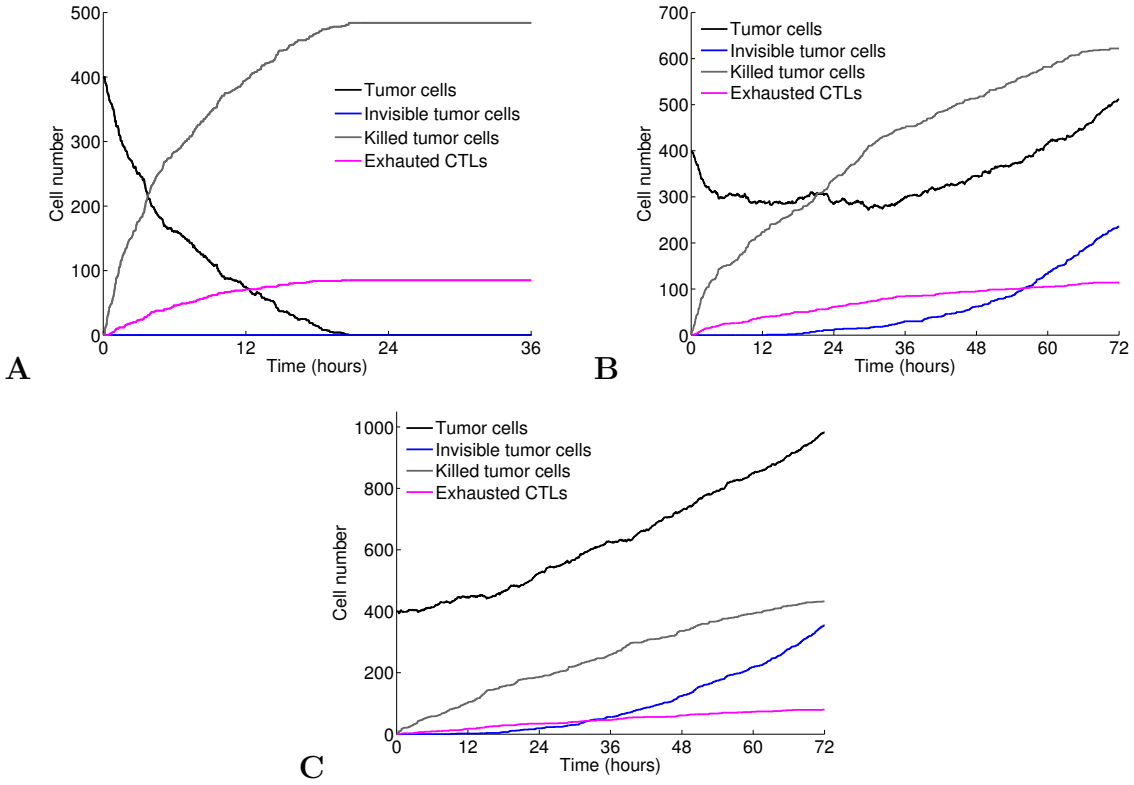



Figure 1.5: **Evolution of the number of cells over time as a function of the initial CTL number**. (A) 200 CTLs are present at the beginning of the simulation. (B) 400 CTLs are present at the beginning of the simulation. (C) 1200 CTLs are present at the beginning of the simulation.

### 1.5.3 Non influence of reducing the time required for killing

We next investigated the impact that CTL time of killing (defined as the time required by a single CTL to annihilate a target cell) might have on tumor eradication. To this end, numerical simulations in which the time required to kill target cells was reduced to less than 10% of the experimentally measured time were performed [Car+09]. Interestingly, a sharp reduction of the time required for killing of individual tumor cells (down to 3 minutes from the initial 26.4 minutes) did not significantly affect neither the efficacy of tumor nodule eradication by CTL nor the mean time required for nodule eradication (Figure 1.4 $D - E$). Conversely, when the time of killing was increased of about 4 times (up to about 2 hours) a clear effect on the efficacy of CTL-mediated cytotoxicity was observed.

The distribution of the initial productive collision, when the time required to kill a target cell was reduced of 30% or of 90%, increase with a similar moderate from the initial time (compare cyan and green histogram to black histogram in Figure 1.8 $B$). This indicates that it is not possible to surpass an upper limit of productive collision number by modulating the time of killing (mean collision value $\sim 130$).

Taken together, the above results support the finding that the number of productive contacts between CTL and tumor cells is an important parameter influencing the success of a CTL population. They show that while an augmentation of the time required for killing of individual target cells strongly affects CTL efficacy, the sharp reduction of the time required for target cell killing does not significantly affect the probability of CTL success in tumor eradication.

### 1.5.4 Influence of biased random walk for CTL displacement

Numerical simulations based on the above-described biased movement show that the probability of success in tumor nodule eradication significantly increases with the increase of CTL attraction towards the scout cells. As shown in Figure 1.6, with a relatively low number of CTL, the increase of the attraction of CTL towards scouts having detected the tumor nodule, strongly augments the probability of tumor nodule eradication. For instance for 400 CTL with no attraction there was no chance of success (see Figure 1.6 *A*). While a relatively moderate attraction of $\nu = 0.1$ guarantees almost 100% probability of success. Lower attraction strengths (ranging between 0.01 to 0.075) still yielded significant impact on the probability of tumor eradication. Figure 1.6 *A* also shows that attraction reduces the range between the minimal number of CTL required for 100% eradication and the maximum number of CTL failing in tumor eradication. These results indicate that attraction sharpens the phase transition in the critical number of CTL that are determinant for tumor eradication. At last, the stronger the attraction $\nu$, the earlier is the phase transition of the probability of CTL success. The mean time required for nodule eradication decreased accordingly (Figure 1.6 *B*).

To verify whether the increase of CTL attraction would result in an augmentation of productive early collisions, we measured the distribution of the number of collisions during the first 5 hours in conditions in which a moderate attraction was applied. As shown in Figure 1.8 *C* the distribution of early collisions for a given number of CTL (400 CTL) was sharply increased in the presence of attraction (cyan histogram, mean value of 188 collisions/5 hours) when compared to the distribution of the same number of CTL with no attraction (black histogram).

Taken together, the above results show that in a cancer immunoediting scenario the rapid recruitment of killer cells helps to reach a sufficient number of early productive collisions leading to tumor eradication.

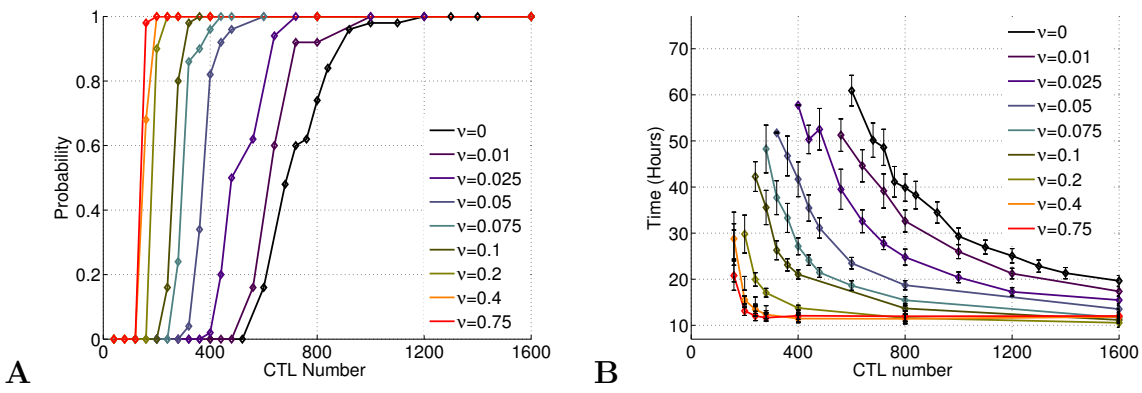



Figure 1.6: **The capacy of scout CTLs to attract other CTLs strongly enhances tumor nodule eradication.** ($A$) Evolution of the probability of the nodule eradication when an attraction strenght is added to the (default) symetric random walk dynamic, number going from 0.01 to 0.75 indicate increasing attraction. ($B$) Mean time of nodule eradication when attraction is added to the default symetric random walk dynamic of an increasing number of CTLs. Results are represented as mean $\pm$ standard deviation of 50 numerical simulations.

### 1.5.5 Influence of reducing CTL exhaustion

An additional parameter that can influence killer cell efficacy against tumors is the phenomenon of exhaustion by which CTL entering the tumor microenvironment receive inhibitory signals that make them progressively less fit in their cytotoxic function. In the model described above we introduced CTL exhaustion by imposing a limit to the maximum number of targets that an individual CTL could kill (this number was defined as 5). In order to investigate the impact that CTL exhaustion could have on the probability of tumor nodule eradication, we let vary the maximal number of target cells that could be killed by an individual CTL. This maximal number will be referred as $\kappa$ in our numerical simulations. As shown in Figure 1.7 $A$, by using the default parameters already used in Figure 1.6 with no attraction ($\nu = 0$), the increase of $\kappa$ from 5 to 10 yields a significant increase of the probability of success with 400 CTL (tumor eradication going from 0 to $\sim$ 95%). Moreover, further increasing the number of target cells killed by a single CTL to 20 and 100, results in an additional increase in tumor eradication probability with less and less required CTL. Numerical simulations also showed that the mean time required for tumor eradication was significantly shortened with the increase of the number of target cells killed by a given CTL (Figure 1.7 $B$).

Figure 1.8 $D$ shows that augmentation of the number of target cells that a CTL could kill resulted in an augmentation of productive early collisions. In fact the distribution of early collisions for a given number of CTL (400 CTL) able to kill 10 target cells was sharply increased (cyan histogram, mean value of 215 collisions/5 hours) when compared to the distribution of the same number of CTL able to kill 5 target cells (black histogram).

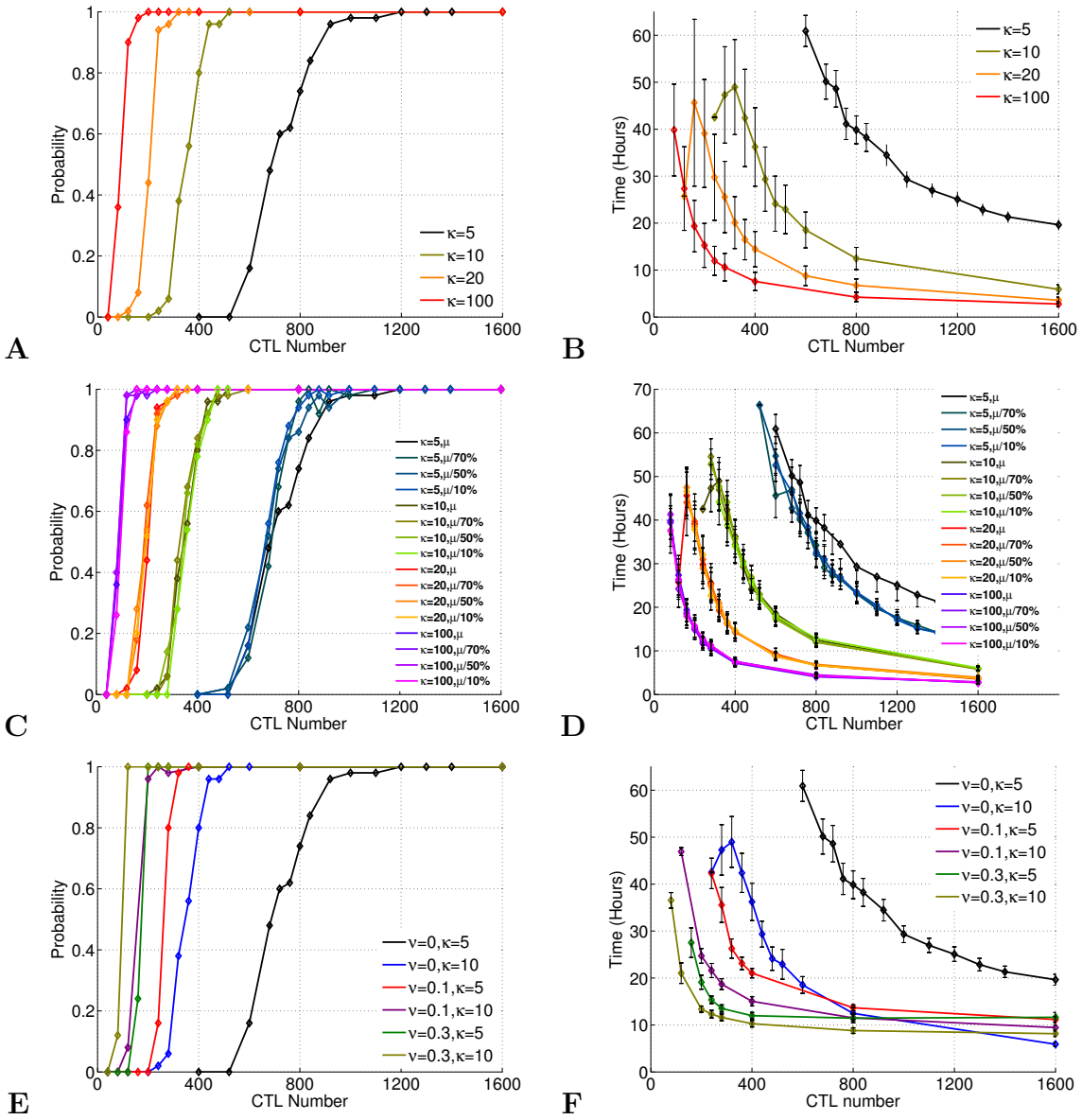



Figure 1.7: **Capacity of CTLs to kill an increasing number of target cells correlates with killing efficacy.** $(A)$ Evolution of the probability of success of the CTL population when the number of target cells killed by a given CTL increases. $(B)$ Evolution of the mean time of success with several parameters of exhaustion. $(C)$ Absence of synergy between reduced exhaustion and time required for killing on the probability of success in tumor nodule eradication. $(D)$ Absence of synergy for the mean time of success. $(E)$ Synergy between reduced exhaustion $\kappa$ and attraction $\nu$ for the probability of success. $(F)$ Synergy between reduced exhaustion $\kappa$ and attraction $\nu$ for the mean time of success. Results are represented as mean $\pm$ standard deviation of 50 numerical simulations.

Interestingly, when the number of killable target cells was varied together with the time required for killing of individual target cells (variation of $\mu$) no synergy between

the two parameters was observed. As shown in Figure 1.7 $A$ and $C$, while the increase of target cells killed by an individual CTL had an important impact on probability of tumor eradication, the reduction of time required for killing did not substantially improve the probability of success. Similar results were obtained when the time required for tumor nodule eradication was measured. While the number of target cells killed by a given number of CTL strongly influenced the time required for tumor eradication, the killing time did not have a major impact on this parameter (Figure 1.7 $B$ and $D$).

Together the above results show that the capacity of CTL to kill an increasing number target cells before becoming exhausted is strongly related to the probability of tumor eradication.

### 1.5.6 Synergy between CTL attraction and reduced CTL exhaustion

To verify whether the augmentation of the number of target cells killed by a given CTL and the attraction of CTL towards the tumor nodule could synergize resulting in a large number of early CTL/tumor productive collisions, we investigated the probability of CTL victory in conditions in which both the number of killed target cells and the CTL attraction were increased. As shown in Figure 1.7 $E$, when the number of killable target cells and the CTL attraction were increased simultaneously, a dramatic increase in the probability of CTL victory was observed, for relatively low attraction and a limited augmentation of the number of killed target cells. Interestingly the phase transition in the number of CTL required to achieve 100% eradication was sharper as compared to the situation in which only attraction was present. The time to eradicate tumor nodules was reduced accordingly (Figure 1.7 $F$).

Figure 1.8 $E$ shows that this combination of effects results in an augmentation of the number of collisions during the first 5 hours. Indeed, the distribution of early collisions for a given number of CTL (400 CTL) was sharply increased (cyan histogram, mean value of 242 collisions/5 hours) when compared to the distribution of the same number of CTL able to kill 5 target cells with no attraction (black histogram).

The above results show that the number of early productive collisions between CTL and their tumor target cells can be efficiently enhanced by increasing in parallel CTL attraction and the number of target cells killed by each CTL.

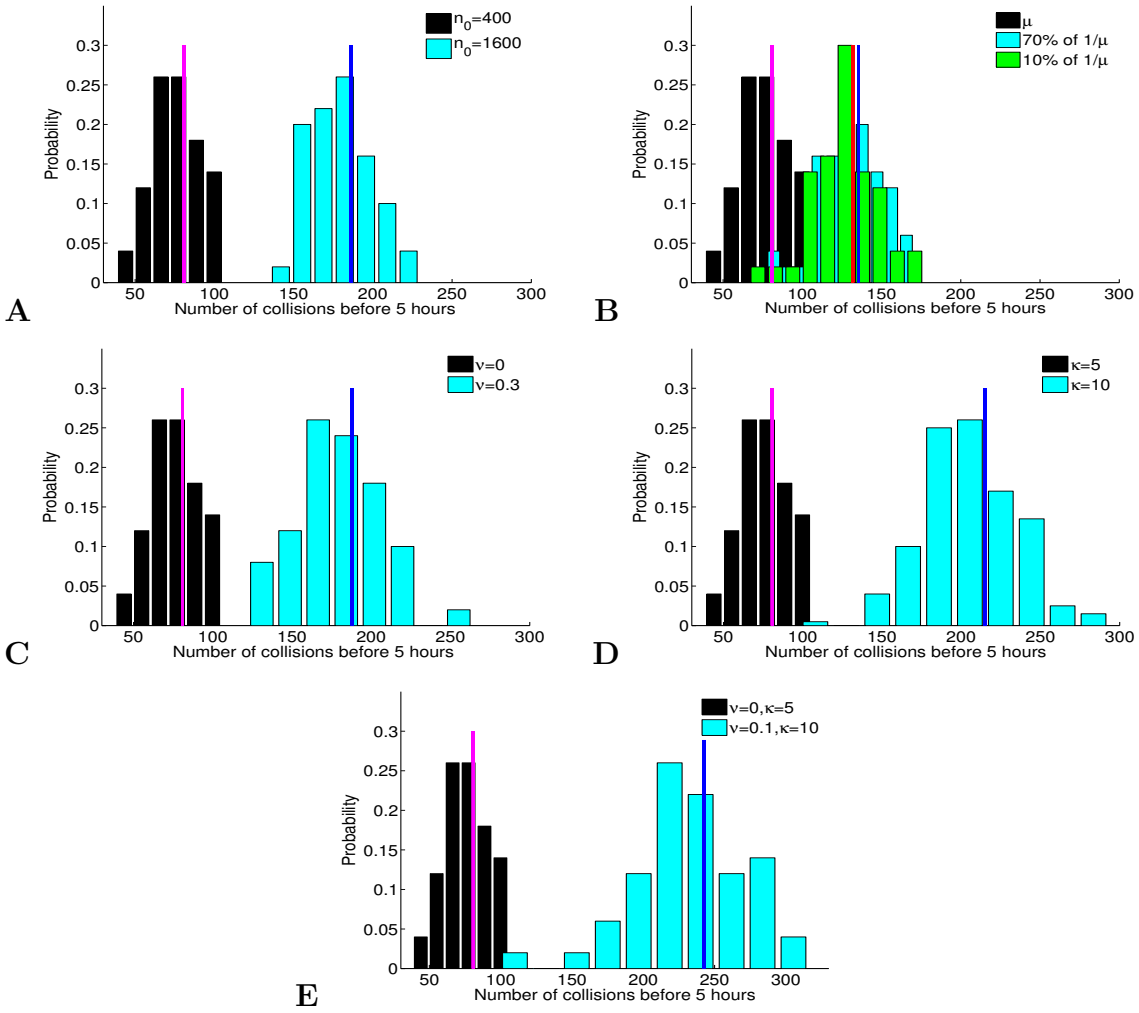



Figure 1.8: **Distribution of the number of early collisions with different parameters.** $(A)$ Empirical distribution of the number of early collisions (CTL killing during the first 5 hours) with a low number of CTL in Black (400 CTLs) and a large number of CTLs in Cyan (1600 CTLs). $(B)$ Empirical distribution with an experimental measured killing time $\mu^{-1}$ in Black. Empirical distribution with a shortest time decreased of 30% in Cyan, and of 90% in Green. $(C)$ Empirical distribution with no attraction ($\nu = 0$) in Black and with attraction ($\nu = 0.3$) in Cyan.$(D)$ Empirical distribution with number of cells killed by each CTL is $\kappa = 5$ before exhaustion in Black and with reduced exhaustion ($\kappa = 10$) in Cyan. $(E)$ Empirical distribution number of cells killed by each CTL is $\kappa = 5$ before exhaustion and no attraction ($\nu = 0$) in Black. Same distribution in Cyan when $\nu = 0.1$ and $\kappa = 10$.

## 1.6 Conclusion and Perspectives

In the present work, we applied mathematical modeling to dissect the multiparametric confrontation between CTL and a growing tumor nodule that undergoes immunoediting. We report that the success of a CTL population in tumor eradication mainly depends on the rate of CTL/tumor nodule productive collisions occurring during the initial period of CTL/tumor confrontation. In this view, we show that, by keeping the number of CTL constant, two major CTL functional responses are identified as critical for tumor eradication: an increased capacity of CTL to kill multiple target cells before undergoing a functional exhaustion and a bias in CTL motility inducing their attraction towards the tumor nodule.

**CTL efficacy against one tumor cell, and CTL inefficacy against tumor mass** It is well established that CTL are very efficient killer cells, that can rapidly annihilate target cells expressing a very small number of specific antigenic determinants on their surface [Far+03; Pur+04]. Moreover, CTL are known to be able to kill outnumbering target cells either simultaneously (via a mechanism named multiple killing [Wie+06]) or sequentially by bouncing from one target to another [PTSV87].

However, the rapidity and efficacy of CTL cytotoxic function does not apply to cancer immuno surveillance [Gaj07]. In the tumor microenvironment, the balance between CTL efficacy and tumor resistance to CTL attack might be in the favor of tumor escape because of: i) the progressive CTL inactivation in the immune-suppressive tumor microenvironment [Cre+13]; ii) the acquisition of resistance by tumor cells via the process of immunoediting [SOS11]. Recent results show that these two processes are interconnected. The set up of multiple checkpoints in cancer microenvironment (such as regulatory T cell recruitment, expression of PDL-1, etc) aiming at inactivating CTL has been shown to be provoked by CTL themselves [Spr+13].

Although mathematical modeling of CTL attack of tumor cells has been previously performed [GR09; MDP06], a mathematical model that globally describes the whole cellular dynamics during tumor nodule growth and immunoediting is missing. In this work, we provide a stepping-stone to address this complex issue. We propose a sharp random dynamical system based on quasi-real parameters that allows to visualize and quantify the probability of success of the whole CTL population.

**Inefficiency of the killing rate reduction** We provide the unexpected observation that the decrease of the time required to kill a target cell by an individual CTL does not strongly improve the performance of the entire CTL population. This result can be explained by the fact that the time of the experimentally measured tumor cell duplication is much longer than the mean time of killing. Since these two time frames do not belong to the same scale, reducing the killing time has only a minor impact on tumor eradication. Conversely, when we strongly increased the time required to kill

target cells, this manipulation had a strong impact on CTL efficacy because tumor cell duplication and time required for killing became closer.

It is tempting to speculate that, in the context of an in vivo confrontation between CTL and tumor cells, variations within a scale of minutes of the time required for killing should not strongly affect the success of an immunotherapy. Conversely, a major obstacle to an efficient immunotherapy is the generation of tumor cell variants enduring for an extremely long time CTL attack or even becoming completely resistant to CTL-mediated cytotoxicity. Thus, successful immunotherapies are likely to be those avoiding the generation of such variants by eradicating the tumor before substantial immunoediting.

**Initial productive collision** Our numerical simulations show that the initial rate of productive collisions between CTL and tumor nodules is a corner stone in tumor eradication. This observation is by itself not fully surprising since it has been previously shown by both mathematical modeling [GBDB11] and experimental approaches [Bud+10] that increasing the ratio between effectors and target cells augments cytotoxicity. Nevertheless, our results allow to illustrate that, in a context of cancer immunoediting, a large number of CTL succeeds in tumor eradication because they can make a large number of early contacts with target cells. As consequence, tumor cells are annihilated before they can accumulate mutations making them progressively resistant or even invisible to CTL.

Numerical simulations show that augmenting the number of target cells killed by each CTL before getting exhausted or *and* biasing CTL random walk towards the tumor strongly increase the probability of CTL success. This striking success is mainly due to the augmentation of the number of early collisions avoiding immunoediting and CTL exhaustion. We also show that increasing multiple killing before exhaustion and directed CTL displacement towards tumor can synergize to achieve even stronger and faster tumor eradication. It is tempting to speculate that this scenario might be amplified in a 3-D situation, since 3-D random collisions are likely to be rare.

**Discussion on CTL population in its whole** Numerical simulations thus reveal an interesting and previously undescribed critical behavior of CTL in tumor eradication: guiding CTL trajectories towards tumor nodules is crucial for efficient detection of a tumor by the CTL population as a whole.

It has been previously proposed that T cells might be viewed as a type of sensory cells and that a multitude of T lymphocytes, with different specificities, can behave as sensory organs [Dav+07]. In the present work we extend this notion by proposing that a tumor-specific CTL population can be viewed as a sensory organ that, in its whole, ignores the developing tumor nodule because of a perception bias.

In cognitive sciences, the biased competition theory of perception postulates that

mental processes can bias the visual perception of objects by prioritizing one object in the visual field instead of another. By analogy to the visual cortex, the CTL population analyzed in this study would be naturally biased to detect pathogens presented by professional antigen presenting cells and would therefore tend to ignore self-antigens expressed by indolently growing tumor cells. The addition of a bias in CTL displacement compensates for this perception defect and makes the entire CTL population more successful in tumor detection and eradication.

A thorough characterization of the observed sharp phase transition in CTL responses when attraction is present, is relevant to better understand the collective functional behavior of CTL populations. The capacity of individual T cells to provide all-or-nothing responses has been thoroughly documented both at the molecular and at the cellular level [Nae+07; Hua+13] and has been illustrated by mathematical modeling of TCR-associated signaling pathways [Das+09]. Here we show that, in conditions of CTL directed migration towards the tumor, the entire CTL population responds in a digital fashion for tumor eradication. We thus extend the notion of digital T cell responses from the individual T cell level to the entire T cell population for a given complex response. Such a digital response behavior of a T cell population further reinforces the analogy between a whole T cell population and a sensory organ.

**Biological perspectives** CTL motility biasing towards scout cells has a biological justication in that CTL are known to store into cytoplasmic vesicles the CTL-attracting chemokine CCL5 and to release it rapidly after TCR stimulation [Cat+04]. Moreover our unpublished experimental results, inspired by the present model, indicate that localized activation of CTL against a tumor nodule favors the recruitment of other CTL towards the nodule (Sabina Müller [2] unpublished observations). Further research is required to define the biological relevance *in vitro* and *in vivo* of CTL recruitment towards a tumor nodule by scout sibling. Nevertheless, it is interestingly to note that our computational study not only stems from experimental measurements, but also suggests precise experimental strategies that will be interesting to develop in the future.

**Mathematical developments** An interesting aspect of our approach is that the mathematical model and the biased competition theory of CTL migration can be extended to different predator/prey dynamical systems as well as colonies of self-interacting animals (ants, sheep, etc.) in which the success of the population depends on biased displacement and self-interacting behaviors of particles.

The mathematical model raises several challenging questions such as the theoretical estimation of the probability of success (with respect to $\nu, \mu, \lambda$, etc) and the mathematical characterization of the phase transition (e.g. a digital or all-or-nothing response)

2. Member of Salvatore Valitutti's team

observed in our simulations with increasing number of CTL in conditions in which CTL displacement is biased towards the tumor nodule.

**Conclusion** We propose a biased competition theory of CTL function in which the trajectories of individual CTL are guided by scout siblings resulting in efficient detection of tumor cells. Such a mechanism favors the success of the entire CTL population that, like a sensory organ, responds to biased stimuli for efficient signal/noise discrimination. The capacity of individual CTL to kill multiple targets either simultaneously or sequentially synergizes with CTL biased displacement in tumor eradication. Our results suggest that, to be successful, therapeutic strategies based on CTL adoptive transfer should aim at potentiating these two synergistic functions of CTL in an attempt to prevent CTL exhaustion and cancer immunoediting.

# Part II

# Macroscopic study derived from coupled system of differential equations

# Chapter 2

# Stochastic study of CTL/nodule interaction

Biologists have established that a Cytotoxic T Lymphocytes can kill tumor cells. But in context of tumor mass, CTL becomes inefficient. To bypass this issue, in Chapter 1 we propose to direct CTL toward CTL that are on the border of the nodule. To have a better understanding of the role of such CTL attraction, we propose here a new model of differential system describing the interaction between CTL en tumor nodule. This model is adapted for alternatively two differents CTL displacements: a self-governing displacement or a biased displacement toward the nodule.

## 2.1 Introduction

The biological setting of this part is the same as the previous chapter, namely, a tumor mass confronted to an immune response. Even if the ability of the CTL to destroy a tumor cell is proved [DL10; Car+09], in this framework, tumor mass circumvents immune recognition [SOS11], via the immunoediting process.

One reason of the CTL inefficiency, highlighted in the previous chapter, is the accessibility of the CTL to the nodule, which depends on CTL displacements. To bypass this difficulty, an immunotherapy is proposed. This therapy consists to generate CTL that are able to release chemio-attractant, when it is in contact with a tumor cell. This allows the attraction of the CTL toward the nodule.

In order to measure the impact of a such immnunotherapy, we develop a system of differential equations that mimics this interaction. More precisely, as there is a high number of tumor cell, we consider a deterministic development of the tumor mass. The equation describing the growth, comes from a random microscopic model.

The growth depends on the number of tumor cells that are actively dividing. The decrease of the nodule relies on the number of CTL on the nodule frontier, those called

scout CTL. This last number is described by the random hitting time of the nodule by a CTL, which depends on the CTL displacement.

Then, to model the impact of the therapy suggested, we propose two dynamical trajectories for the CTL. A brownian motion models the CTL displacement without treatment, and an O.U. process models the CTL displacement under simulated immunotherapy.

Such a model relies on a few number of parameters and allows an easy numerical resolution. Then, we could imagine to apply the results to personalized medicine. That is to say a calibration of the parameters using clinical data of individual patient, and a personal calibration of the immunotherapy that leads to tumor eradication.

Firstly, we give a detailed description of the system of differential equation describing the CTL/tumor nodule interaction. As a stochastic approach is used to describe the CTL displacement, we give an equation on the expectation of the number of scout CTL. This equation depends on the distribution of the first time for a CTL to reach the nodule, which is studied in a third section. The last section is devoted to the proof of the main theorems.

## 2.2 Mathematical model

We modelize the interaction between CTL and tumor nodule over time in two dimensions, to simplify and to have a better understanding of the model.

An idealized model of a solid tumor is a disk of center $(0, 0)$ and radius $R_t$ (at time $t$), consisting of several concentric shells [Kan+00], [MDP06]. The innert core, the dark grey region in Figure 2.1, is composed of necrotic cells (death cells). It is a disk of radius $N_t$, it is also characterized by its distance to the nodule border, $E_t$ such that $E_t + N_t = R_t$. The next shell, the dark green region in Figure 2.1, is the quiescent part. It contains alive but non-proliferative cells owing to a low nutrient and oxygen concentration. The superficial shell, with thickness $\delta E_t$, $\delta \in ]0, 1]$, the light green part in Figure 2.1, contains enough nutrient and oxygen to maintain active cellular division, is constituted of proliferative cells. The quiescent part and the proliferative part taken together are called "alive part", its tickness is $E_t$.

Define $A_t$ the number of cells in the alive part of the nodule at time $t$. We want to know the conditions leading to have $A_t = 0$. Indeed, if $A_t = 0$, only necrotic cells stay in the nodule, it means that all alive cells are eradicated. This number depends on the number of CTL on the edge of the nodule, which kill alive cells.

Let us assume the number of CTL, $n_0 \geq 2$, fixed over the time. All CTL are

distributed independently, following a probability distribution $\mu_0^i$. The domain of $\mu_0^i$ is $\mathbb{R}^2 \backslash \overline{\mathcal{B}(0, R_0)}$. Let $Z_t^i$ be the position of the CTL number $i$ at time $t$. Note that some CTL are not in contact with the tumor. Those located on the frontier of the nodule are called "scout CTL". We denote $\mathcal{L}_t$ and $L_t$ respectively, the set of scout CTL and the number of scout CTL, at time $t$. Notations are recalled in Table 3.1.

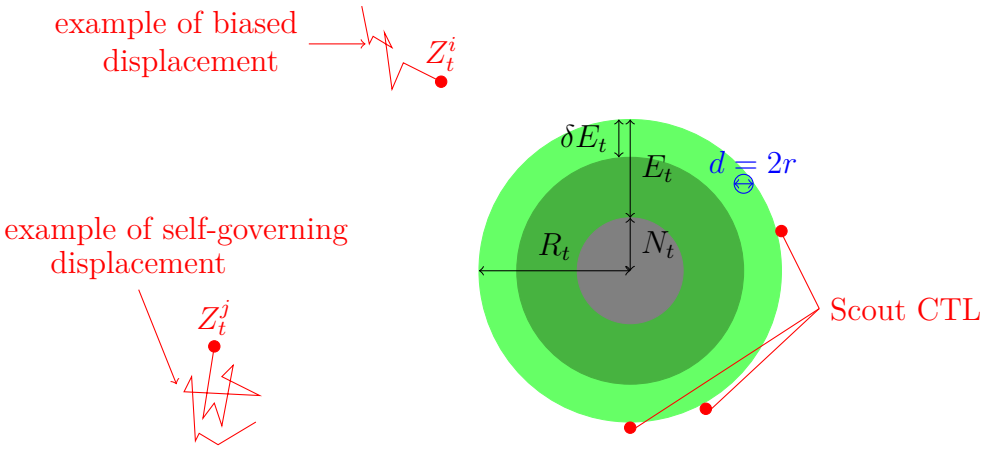



Figure 2.1: **Representation of the interaction between CTL and tumor nodule.** The *dark gray* region is the necrotic part, the *light gray* region is the quiescent part (alive but non-proliferative cells), and the *green shell* is the proliferative part. The *red* dots are CTL, and red lines an example of possible trajectories before the time $t$.

### 2.2.1 Number of alive cells in the nodule

Note that, the alive cell number is tightly dependent on the number of scout CTL. The CTL cytolytic activity is not well characterized and under study, see for instance [Gad+14] or Chapter 4. For the sake of simplicity, let us assume on CTL activity.

First, we assume that a CTL can kill one target cell at a time. In this model, we work in continuous time, that allow us to consider simultaneous killing as very close sequential killing. This is in accordance with biological observations [Rot+78; PB82; Wie+06]. Secondly, we suppose that a CTL can kill an infinite number of tumor cells. In Subsection 3.4, we discuss this assumption. Thirdly, we assume that if a CTL is a scout CTL (touching the nodule), it is a scout CTL the remaining time. Indeed, when a scout CTL kill one tumor cell, the CTL is close to the nodule, then it has a large probability to touch an other target cell, and to remain scout CTL.

Under these conditions the cell number in the alive part (recall alive part contains proliferative and quiescent cells), at time $t + h$, can be described by the following equation:

$$A_{t+h} = A_t + h\lambda \left(\delta A_t - L_t\right) \mathbb{1}_{\delta A_t - L_t \geq 0} - h\mu L_t + o(h).$$

As there is many cells, we consider a deterministic evolution (LLN). Here, we detail more this equation. The alive cell number at time $t+h$ ($A_{t+h}$), is equal to the alive cell number at time $t$ ($A_t$), more tumor cells created, and we substract tumor cells killed by CTL, over the time $[t, t+h[$. As a CTL kills one target cell at once, over a time $h$, the killed cell number is equal to the number of scout CTL $L_t$, multiplied by the killing rate $\mu$. The tumor cell that can duplicate (create new cells) are cells in the proliferative part less cell touched by a CTL. Indeed, we assume that a tumor cell touched by a CTL cannot duplicate. As a CTL kills one cell at a time, the tumor cell number in this case is $L_t$. In conclusion the cell number in the alive part that can duplicate is $\delta A_t - L_t$. And multiplying this number by the division rate of one tumor cell $\lambda$, we obtain the created cell number over a time $h$.

Assuming that $L_t$ tumor cells are touched by a CTL is a simplification. Indeed, we neglect the fact that one tumor cell can be touched by several immune cells. We also consider the fact that only tumor cell close to the nodule border can be touched by a CTL.

Assuming that $\delta A_t$ is the proliferative tumor cell number, is also a simplification. Indeed, the alive part is define by its thickness $E_t$. Then, adding new cell in the alive part can push another cell in the necrotic part. We neglect this aspect, even if one can consider $\tilde{\delta} A_t$, where $\tilde{\delta} \in ]0, \delta[$.

Then in continuous time, we propose the following model

$$A_t' = [\lambda(\delta A_t - L_t)\mathbb{1}_{\delta A_t - L_t > 0} - \mu L_t]\, \mathbb{1}_{A_t > 0}. \tag{2.1}$$

To obtain $A_t$, we have to determine the number of scout CTL: $L_t$.

### 2.2.2 Number of CTL on the border of the nodule

Recall that scout CTL are CTL on the border of the nodule. It means that each scout CTL at time $t$, is at distance the nodule radius ($R_t$) from the nodule center. Then the number of scout CTL, for $|.|$ a norm in $\mathbb{R}^2$, satisfies the equation

$$L_t = \sum_{i=1}^{n_0} \mathbb{1}_{|Z_t^i| \leq R_t}, \tag{2.2}$$

where, we recall that $Z_t^i$ is the position of the CTL $i$, and $n_0$ is the number of CTL. Note that $L_t$ depends on the nodule radius, but, we have not information on $R_t$. Therefore, before to study more precisely $L_t$, from the alive cell number $A_t$, given by Equation (2.1), we develop an equation on the radius of the nodule.

| Variable | Description |
|---|---|
| $R_t$ | nodule radius at time $t$ |
| $E_t$ | alive part thickness of the nodule at time $t$ |
| $N_t$ | necrotic part radius of the nodule at time $t$ |
| $A_t$ | number of alive cells (quiescent cells and proliferative cells) at time $t$ |
| $\mu_0^i$ | initial distribution of the CTL $i$ |
| $\mu_t^i = \mu_t$ | distribution of the CTL $i$ at time $t$ |
| $\mathcal{L}_t$ | set of scout CTL (CTL touching the nodule) at time $t$ |
| $L_t$ | number of scout CTL at time $t$ |
| $Z_t^i$ | trajectory of the CTL $i$ at time $t$ |
| **Parameter** | **Description** |
| $n_0$ | CTL number |
| $\delta$ | proliferative part proportion in the alive part |
| $\mu$ | killing rate |
| $\lambda$ | division rate of one tumor cell |
| $d$ | tumor cell diameter |
| $r$ | tumor cell radius |

Table 2.1: Variables and parameters used in the model

**Equation on the radius of the nodule at time $t$**

Here, we aim to describe the evolution of the radius of the tumor mass that has $A_t$ alive cells.

Considering cell as a disk, we are confronted to a packing problem. There exist some studies on packing circles into a bigger circle [Gra+98]. But, they do not look satisfying in our framework mainly because of tumor cells in nodule are not exactly a disk. They adapt their shape in order to be in their whole, as compact as possible. Neglecting the geometrical aspect, we can solve this problem using the tumor cell area and the tumor mass area. Indeed, alive part area can be approximate by the nodule area less the necrotic part area. By using a rough approximation of the tumor cell shape as a disk of radius $r$, we have

$$\pi r^2 A_t = \pi R_t^2 - \pi N_t^2. \tag{2.3}$$

Until now, no information is available on the behavior of the necrotic part radius $N_t$. However, two phases describe the nodule development: an increase phase and a decrease phase.

Over decreasing phase of the nodule, one can suppose the necrotic part constant $N_t = N$. Since, dead cells in the necrotic part, fortunately, do not come back to life. And alive cells, which are closer to the nodule border, remain alive cells. Then, the thickness of the alive part $E_t$ decreases.

Over increasing phase of the nodule, it is the reverse. Assuming the nodule is large enough that nutrient and oxygen cannot go deeper in the nodule, then $E_t = E$ is

constant. Alive cells furthest from the nodule border, become necrotic cells and $N_t$ increases.

This two behaviors are described in Figure 2.2.

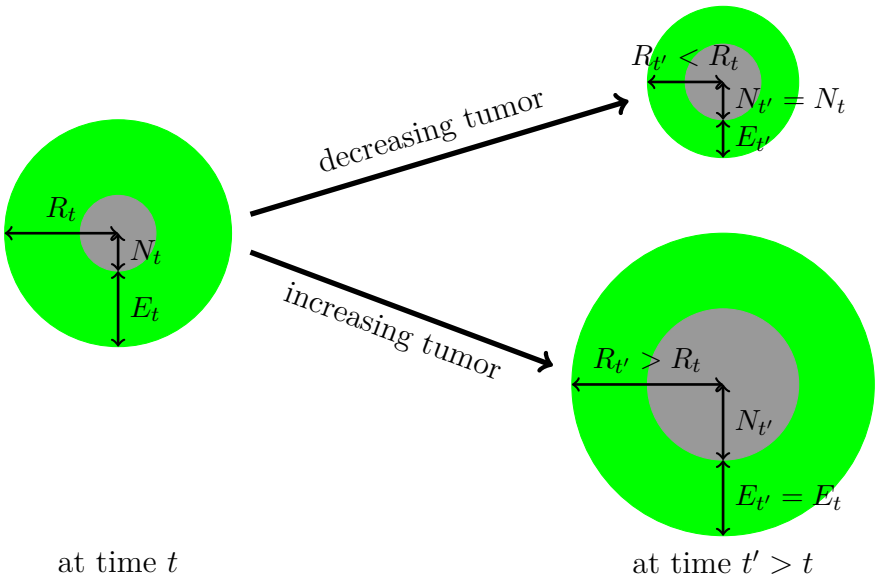



Figure 2.2: **The two dynamics of the nodule**. The green region is the alive part, and the gray region is the necrotic part of the nodule.

Then Relation (2.3) becomes:

– During decreasing tumor phase: $\pi R_t^2 - \pi N^2 = \pi r^2 A_t$, then $R_t^2 = r^2 A_t + N^2$ if and only if $R_t = \sqrt{r^2 A_t + N^2}$. One obtains the differential equation on the nodule size $R_t$,

$$R_t' = \left[\left(\frac{\lambda\delta R_t}{2} - \frac{\lambda\delta N^2}{2R_t} - \frac{r^2\lambda L_t}{2R_t}\right)\mathbb{1}_{\delta(R_t^2-N^2)-r^2L_t>0} - \frac{r^2\mu L_t}{2R_t}\right]\mathbb{1}_{R_t'<0,\ R_t>N}. \quad (2.4)$$

– During increasing tumor phase: $\pi R_t^2 - \pi(R_t - E)^2 = \pi r^2 A_t$, $R_t = \frac{r^2 A_t + E^2}{2E}$ and

$$R_t' = \left[\left(\lambda\delta R_t - \frac{\lambda\delta E}{2} - \frac{r^2\lambda L_t}{2E}\right)\mathbb{1}_{\delta(2R_tE-E^2)-r^2L_t>0} - \frac{r^2\mu L_t}{2E}\right]\mathbb{1}_{R_t'>0}. \quad (2.5)$$

### Equation on the number of scout CTL at time $t$

To compute $L_t$ from Expression (2.2), we have to evaluate the hitting time by $Z_t$ of radius $R_t$. The difficulty comes from both variables $R_t$ and $Z_t^i$ vary with time. To have a better understanding, we study the derivative of $L_t$.

For $h > 0$, the number of scout CTL at time $t + h$ is equal to the number of scout CTL at time $t$ (recall that a scout CTL remains a scout CTL), more number of non-scout CTL that reach the nodule over the time $h$. Formaly

$$\text{card}\left\{i, \quad Z_t^i \notin \mathcal{L}_t \text{ and } \exists\ 0 < s \leq h,\ |Z_{t+s}^i| \leq R_{t+s}\right\}.$$

However, for $h$ small, over $]t, t+h[$, the nodule radius shall not vary much (tumor cell division rate is small). Then, we assume a constant radius $R_t$. One can get the equation following

$$\begin{aligned} L_{t+h} &= L_t + \sum_{i|Z_t^i \notin \mathcal{L}_t} \mathbb{1}_{|Z_{t+h}^i| \leq R_t} \\ &= L_t + \sum_{i=1}^{n_0} \mathbb{1}_{\left\{Z_{t+h}^i \in \mathcal{L}_t\right\} \cap \left\{Z_t^i \notin \mathcal{L}_t\right\}}. \end{aligned}$$

Whereas, the position of each CTL is unknown over time, their distribution at time $t$ can be computed, using the distribution at time 0. Let $i \leq n_0$,

$$A \subset \mathbb{R}^2 \backslash \overline{\mathcal{B}(0, R_t)}, \quad \mu_t^i(A) = \mathbb{P}_{\mu_0^i}(Z_t^i \in A || Z_t^i| > R_t),$$

where $\mu_t^i$ denotes the distribution of the CTL $i$ at time $t$ conditioned not to be in $\mathcal{L}_t$. It is natural to suppose that all CTL are identically distributed, call $\mu_t = \mu_t^i$ for all $i \in \{0, ..., n_0\}$. Instead of $L_{t+h}$ we compute its expectation,

$$\begin{aligned} \mathbb{E}_{\mu_t}(L_{t+h}) &= \mathbb{E}_{\mu_t}(L_t) + \mathbb{E}_{\mu_t}\left(\sum_{i=1}^{n_0} \mathbb{1}_{\left\{Z_{t+h}^i \in \mathcal{L}_t\right\} \cap \left\{Z_t^i \notin \mathcal{L}_t\right\}}\right) \\ &= \mathbb{E}_{\mu_t}(L_t) + n_0 \mathbb{P}_{\mu_t}\left(\left\{Z_{t+h}^1 \in \mathcal{L}_t\right\} \cap \left\{Z_t^1 \notin \mathcal{L}_t\right\}\right), \end{aligned}$$

as the trajectories are assumed identically distributed.

$$\begin{aligned} \mathbb{E}_{\mu_t}(L_{t+h}) &= \mathbb{E}_{\mu_t}(L_t) + n_0 \mathbb{P}_{\mu_t}\left(Z_{t+h}^1 \in \mathcal{L}_t | Z_t^1 \notin \mathcal{L}_t\right) \mathbb{P}_{\mu_t}(Z_t^1 \notin \mathcal{L}_t) \\ &= \mathbb{E}_{\mu_t}(L_t) + n_0 \left(1 - \mathbb{P}_{\mu_t}(Z_t^1 \in \mathcal{L}_t)\right) \mathbb{P}_{\mu_t}\left(T_{R_t}^1 \leq h\right), \end{aligned}$$

where $T_{R_t}^1 = \inf\left\{s > 0, \ \ |Z_{t+s}^1| \leq R_t\right\}$ is the hitting time of $R_t$ by $Z_t^i$. Note that

$$\mathbb{P}_{\mu_t}(Z_t^1 \in \mathcal{L}_t) = \frac{1}{n_0} \sum_{i=1}^{n_0} \mathbb{E}_{\mu_t}\left(\mathbb{1}_{Z_t^1 \in \mathcal{L}_t}\right) = \frac{1}{n_0} \mathbb{E}_{\mu_t}(L_t).$$

Hence,

$$\mathbb{E}_{\mu_t}(L_{t+h}) = \mathbb{E}_{\mu_t}(L_t) + (n_0 - \mathbb{E}_{\mu_t}(L_t)) \mathbb{P}_{\mu_t}\left(T_{R_t}^1 \leq h\right). \tag{2.6}$$

Then in a continuous time, we propose the following model:

$$\partial_t \mathbb{E}(L_t) = (n_0 - \mathbb{E}_{\mu_t}(L_t)) \lim_{h \to 0} \frac{\mathbb{P}_{\mu_t}(T^1_{R_t} \leq h)}{h}. \tag{2.7}$$

### 2.2.3 CTL dynamics

Note that the derivative of the expected number of scout CTL strongly depends on the choice of CTL dynamic, and it governs their hitting time. In the next section, we compute the probability of hitting time, under two dynamics described here.

First, a self-governing CTL displacement is studied. Indeed, it is not known if immune cells are attracted by the nodule. In addition, interactions between CTL are not established. In [Har+12] suggest independent generalized Lévy walk for CTL motility in brain. For the sake of simplicity, we assume independent Brownian motions for CTL dynamic.

Secondly, CTL displacement under simulated immunotherapy is developed. As suggested in Chapter 1, this therapy consists in the generation of CTL that can release chemo-attractants when they are scout CTL. That guides other CTL toward the nodule. In keeping with the Brownian motion, we assume an Ornstein-Uhlenbeck (O.U.) dynamic for CTL motility. Intuitively, an O.U. process is a Brownian motion with higher probability to go to 0.

## 2.3 Hitting time and number of scout CTL

In this part, for $t$ fixed, we obtain an exact expression of $\mathbb{E}(L_t)$, using an asymptotic as $h$ tends to 0. In what follows, all expectations are evaluate under the distribution $\mu_t$.

### 2.3.1 Brownian displacement

In this subsection, let us assume that the dynamics of non-scout CTL are independent Brownian motions. This suggests that CTL does not preferentially direct toward a developing tumor. Let $Z_{t+s}$ be the position of a non-scout CTL at time $t+s$. For $0 \leq s$, it satisfies

$$\mathrm{d}Z_{t+s} = \mathrm{d}B_s.$$

**Bessel process** Recall that $T_{R_t}$ is the first time for non-scout CTL to hit the nodule. It equals the hitting time of $R_t$ by the norm of the process, namely $\rho_s = ||Z_{t+s}||_2$. There exist some results on process $\rho$, the radial process of the Brownian motion also called

Bessel process, see for instance [RY99; BS02; HM13; BMR13]. And $\rho$ is the solution of the following system:

$$\begin{cases} \mathrm{d}\rho_s = \mathrm{d}B_s + \frac{1}{2\rho_s}\mathrm{d}s \\ R_t \leq \rho_s \\ \rho_0 = ||Z_t||_2 > R_t \end{cases}.$$

**Theorem 2.3.1.** *Let $R_t > \frac{1}{2}$ and $\mu_t$ be the probability distribution of one CTL conditioned to not be a scout CTL. Assume $\mu_t$ absolutely continuous with respect to the Lebesgue measure, such as $\mathrm{d}\mu_t(x) = f_t(x)\,\mathrm{d}x$ and $f_t'(R_t) \geq 0$. Then*

$$\partial_t \mathbb{E}(L_t) = (n_0 - \mathbb{E}(L_t)) \lim_{h\to 0} \frac{\mathbb{P}_{\mu_t}(T_{R_t} < h)}{h} = (n_0 - \mathbb{E}(L_t))\frac{f_t'(R_t)}{2}.$$

**Remark 2.3.2.** *The hypothesis of $R_t > \frac{1}{2}$ in the previous theorem, is just technical. This hypothesis has not a real role in biological context, we have to choose a good scale for data. For instance, if we choose a $\mu m$ scale, the theorem fails for $R_t < 0.5\mu m$. As a cell radius is higher than $0.5\mu m$, if $R_t = 0.5\mu m$ the nodule is eradicated. Then, to choose a $\mu m$ scale is a good choice for data scale, but a mm scale would not be a good option.*

**Sketch of proof** The difficulty comes from the computation of $\mathbb{P}_{\mu_t}(T_{R_t} < h)$ where $t$ is fixed. For $x \leq R_t$, the starting point of $\rho$, this probability is well known [BS02], but for $x > R_t$ this calculus is under study. [HM13] recently an exact expression, but it is intricate. They also develop an upper tail behavior of this probability, however in this study we are interested in the lower tail. In the proof of Theorem 2.3.1, we propose an approximation of this probability.

We use the comparison Theorem for one-dimensional Itô process, see for instance [IW81, p 438]. As for all $s \in \mathbb{R}_+$, $\rho_s \in ]R_t, \infty[$, let $B_s$ be a one dimensional Brownian motion, then with probability 1

$$\mathrm{d}\bar{\rho}_s = \mathrm{d}B_s \leq \mathrm{d}\rho_s = \mathrm{d}B_s + \frac{\mathrm{d}s}{\rho_s} \leq \mathrm{d}B_s + \frac{1}{2R_t}\mathrm{d}s = \mathrm{d}\hat{\rho}_s. \tag{2.8}$$

And the hitting time of $R_t$ by this 3 processes are also ordered, such as their probability.

The density probability of the hitting time by a Brownian motion ($\bar{\rho}$) or a Brownian motion with a positive constant drift ($\hat{\rho}$) is known. Their expression are straightforward, see for example [BS02]. Thereby, it is possible to calculate a lower and upper bound of the probability of the hitting time of $R_t$ by the process $\rho$.

Afterwards, noting that, the closer from $R_t$ the particle $\rho_0$ is, the larger the probability to reach $R_t$ before the time $h$ is. It remains to compute separately $\int_{R_t}^{R_t+\varepsilon(h)} \mathbb{P}_x(T_{R_t} \leq h)\,\mathrm{d}\mu_t(x)$, which is expected to be large, and $\int_{R_t+\varepsilon(h)}^{\infty} \mathbb{P}_x(T_{R_t} \leq h)\,\mathrm{d}\mu_t(x)$, which is forecasted to be small, where $\varepsilon(h) \in ]R_t, \infty[$.

A detailed proof of this property is done in the following section. This proof will be a guideline to prove Lemma 2.3.4 below.

**Remark 2.3.3.** *Let us mentioning parallel work [BMR13] that publish similar results. They give a sharp estimation of the probability of the hitting time of reach 1 leaving from $a > 1$ by a Bessel process. Their estimation is up to a multiplicative constant, however constant is important in our study.*

*Yet, using their density we obtain the same result, up to a multiplicative constant. See Section 2.4.3 for more details.*

### 2.3.2 Ornstein-Uhlenbeck displacement

In this subsection, we consider a simulated immunotherapy, such that CTL are attracted by other CTL that are on the nodule border (those called scout CTL). Formally, CTL displacements are represented by independent Ornstein-Uhlenbeck process (O.U.). They only depend on number of scout CTL. Let $Z_{t+s} \notin \mathcal{L}_t$, and $s \in [0, h]$,

$$\mathrm{d}Z_{t+s} = \mathrm{d}B_s - \nu_t Z_{t+s}\,\mathrm{d}s, \tag{2.9}$$

where $\nu_t$ is the weight to direct $Z_t$ toward the nodule. Note that in continuous time, CTL arrive one at a time on the edge of the nodule. Classically, we assume that over the time $[t, t+h[$, for a small $h$, the number of scout CTL does not increase. Then, we suppose $\nu_t = \mathbb{E}(L_t)$, it means that more scout CTL are on the nodule, higher the weight to attract CTL is.

As in the brownian displacement we study the radial process of the dynamic: $\rho_s = ||Z_{t+s}||_2$. Define $T_{R_t}$ the first time for a non-scout CTL to reach $R_t$. Then

$$\begin{cases} \mathrm{d}\rho_s = \mathrm{d}B_s + \left(\frac{1}{2\rho_s} - \mathbb{E}(L_t)\rho_s\right)\mathrm{d}s \\ R_t \leq \rho_s \\ \rho_0 = ||Z_t||_2 > R_t \end{cases}.$$

As in the brownian displacement we aim to compute $\mathbb{P}_{\mu_t}\left(T_{R_t} \leq h\right)$, where $\mu_t$ is the distribution of each non scout-CTL at time $t$. Up to now, there does not exist explicit formula for the radial Ornstein-Uhlenbeck process. As in the Bessel process, using the comparison Theorem for one-dimensional Itô process, we get the following inequalities, for $B_s$ a Brownian motion, a.s.

$$\mathrm{d}\bar{\rho}_s = \mathrm{d}B_s - \mathbb{E}(L_t)\rho_s\,\mathrm{d}s \leq \mathrm{d}\rho_s \leq \mathrm{d}B_s + \left(\frac{1}{2R_t} - \mathbb{E}(L_t)\rho_s\right)\mathrm{d}s = \mathrm{d}\hat{\rho}_s. \tag{2.10}$$

We begin by developing an upper bound of $\mathbb{P}_{\mu_t}(T_{R_t} \leq h)$ in the next lemma. Let us call $\bar{T}_{R_t}$ the hitting time of $R_t$ by the process $\bar{\rho}_s$.

**Lemma 2.3.4.** *Let $R_t > \frac{1}{\sqrt{2}}$ and $\mu_t$ be a probability distribution of a CTL conditioned to not be a souct CTL. It is assumed absolutely continuous with respect to the Lebesgue measure, such as $\mathrm{d}\mu_t(x) = f_t(x, \mathbb{E}(L_t))\,\mathrm{d}x$ and $\partial_x f_t(R_t, \mathbb{E}(L_t)) \geq 0$. Then*

$$\lim_{h\to 0} \frac{\mathbb{P}_{\mu_t}(\bar{T}_{R_t} \leq h)}{h} = \frac{\partial_x f_t(R_t, \mathbb{E}(L_t))}{2}.$$

**Sketch of proof** The density of the hitting time by an O.U. process without constant drift (as the process $\bar{\rho}_s$) is not so explicit, see [APP05] for details. By Girsanov's Theorem, the proof of this lemma is leading back to the proof of Theorem 2.3.1. Details of the proof are in following section.

**Theorem 2.3.5.** *Let $R_t > \frac{1}{\sqrt{2}}$ and $\mu_t$ a probability distribution with the same assumption than in the previous lemma. Then*

$$(n_0 - \mathbb{E}(L_t))\frac{\partial_x f_t\left(R_t - \frac{1}{2\mathbb{E}(L_t)R_t}, \mathbb{E}(L_t)\right)}{2} \leq \partial_t \mathbb{E}(L_t) \leq (n_0 - \mathbb{E}(L_t))\frac{\partial_x f_t(R_t, \mathbb{E}(L_t))}{2}.$$

*Proof.* Using Lemma 2.3.6 below, $\mathbb{P}_x\left(\hat{T}_{R_t} \leq h\right) = \mathbb{P}_{x-\frac{1}{2\mathbb{E}(L_t)R_t}}\left(\bar{T}_{R_t-\frac{1}{2\mathbb{E}(L_t)R_t}} \leq h\right)$ where $\bar{T}_{R_t-\frac{1}{2\mathbb{E}(L_t)R_t}}$ is the hitting time of $R_t - \frac{1}{2\mathbb{E}(L_t)R_t}$ by the translated process $\hat{\rho}_s$. Then, from the variable $y = x - \frac{1}{2\mathbb{E}(L_t)R_t}$

$$\lim_{h\to 0} \frac{\mathbb{P}_{\mu_t}\left(\hat{T}_{R_t} \leq h\right)}{h} = \frac{\partial_x f_t\left(R_t - \frac{1}{2R_t\mathbb{E}(L_t)}, \mathbb{E}(L_t)\right)}{2}.$$

We conclude by (2.7).

□

**Lemma 2.3.6.** *Let $Y_t$ an O.U process, such as $\mathrm{d}Y_t = \mathrm{d}B_t + (a - bY_t)\,\mathrm{d}t$. Then*

$$Y_t = X_t + \alpha,$$

*where $\mathrm{d}X_t = \mathrm{d}B_t - bX_t\,\mathrm{d}t$ and $\alpha = \frac{a}{b}$.*

*Proof.* Let $X_t = Y_t - \frac{a}{b}$, then

$$\begin{aligned}
\mathrm{d}X_t &= \mathrm{d}Y_t \\
&= \mathrm{d}B_t + (a - bY_t)\,\mathrm{d}t \\
&= \mathrm{d}B_t + \left(a - b\left(X_t - \frac{a}{b}\right)\right)\mathrm{d}t \\
&= \mathrm{d}B_t - bX_t\,\mathrm{d}t.
\end{aligned}$$

This concludes the proof.

□

To study the radius of the nodule over time, one has to replace $L_t$ by its expectation in Equations (2.4) and (2.5) in first. Then, one has to solve the differential equation on $\mathbb{E}(L_t)$, given in Theorems (2.3.1) and (2.3.4), and equations on the nodule radius. The following chapter is devoted to numerical studies of this model.

**Remark 2.3.7.** *In this part, we assume that the attraction strength in equation* (2.9)*, $\nu_t$, is equal to $\mathbb{E}(L_t)$. But, any linear function of $\mathbb{E}(L_t)$ with positive parameters can be chosen for $\nu_t$, namely, $\nu_t = \alpha\mathbb{E}(L_t) + \beta$, where $\alpha, \beta \geq 0$.*

## 2.4 Additional proofs

### 2.4.1 Proof of Theorem 2.3.1

*Proof.* Using Inequality (2.8), we develop a lower and an upper bounds of the probability to reach $R_t$ starting with the probability $\mu_t$. Step 1: to maximize the probability, we consider a brownian motion for the CTL dynamic $(\bar{\rho}_s)$ and step 2: to minimize the probability we assume a brownian motion with a positive drift $(\hat{\rho}_s)$. The hitting time of these processes are respectively called $\bar{T}_{R_t}$ and $\hat{T}_{R_t}$.

• Step 1: Define an upper bound of the probability to reach $R_t$.

Let $x \in ]R_t, \infty[$ and $\bar{\rho}_0 = x$, the probability density function of $\bar{T}_{R_t}$, beginning at $x$, is known [BS02; APP05],

$$\bar{p}_{x\to R_t}(s) = \frac{|x - R_t|}{\sqrt{2\pi s^3}} \exp\left(-\frac{(R_t - x)^2}{2s}\right). \tag{2.11}$$

Thus by the variable $u = \frac{(R_t-x)^2}{2s}$,

$$\begin{aligned}\mathbb{P}_x\left(\bar{T}_{R_t} \leq h\right) &= \frac{|x - R_t|}{\sqrt{2\pi}} \int_0^h \frac{1}{s^{3/2}} \exp\left(-\frac{(R_t - x)^2}{2s}\right) \mathrm{d}s \\ &= \int_{\frac{(R_t-x)^2}{2h}}^{\infty} \frac{e^{-u}}{\sqrt{\pi u}} \mathrm{d}u. \end{aligned} \tag{2.12}$$

Now, $x$ belongs to $]R_t, \infty[$ and follows the distribution $\mu_t$. As it is suggested in Sketch of proof, we cut the integral in two pieces, splitting CTL that are close to the nodule, such as $x \in ]R_t, R_t + \varepsilon(h)]$ where $\varepsilon(h) = \sqrt{2h\ln(1/h)}$, and those that are far $(x > R_t + \varepsilon(h))$.

$$\mathbb{P}_{\mu_t}\left(\bar{T}_{R_t} \leq h\right) = \int_{R_t}^{R_t+\varepsilon(h)} \mathbb{P}_x\left(\bar{T}_{R_t} \leq h\right) \mathrm{d}\mu_t(x) + \int_{R_t+\varepsilon(h)}^{\infty} \mathbb{P}_x\left(\bar{T}_{R_t} \leq h\right) \mathrm{d}\mu_t(x). \tag{2.13}$$

For the first integral, from (2.12), Fubini-Tonelli's Theorem and $\mathrm{d}\mu_t(x) = f_t(x)\,\mathrm{d}x$, one can obtain

$$\begin{aligned}\int_{R_t}^{R_t+\varepsilon(h)} \mathbb{P}_x\left(\bar{T}_{R_t} \le h\right)\,\mathrm{d}\mu_t(x) &= \int_{R_t}^{R_t+\varepsilon(h)} \int_{\frac{(R_t-x)^2}{2h}}^{\infty} \frac{e^{-u}}{\sqrt{\pi u}}\,\mathrm{d}u\,\mathrm{d}\mu_t(x)\\ &= \int_0^{\infty} \frac{e^{-u}}{\sqrt{\pi u}} \int_{R_t}^{R_t+\left(\sqrt{2hu}\wedge\varepsilon(h)\right)} f_t(x)\,\mathrm{d}x\,\mathrm{d}u.\end{aligned}$$

Since $\mu_t$ is a probability distribution of the non-scout CTL at time $t$, the density of $\mu_t$ satisfies $f_t(R_t) = 0$. Then with the variable $x = y + R_t$, and Taylor formula, $f_t(y + R_t) = f_t'(R_t)y + o(y)$, we can separate the integral as follows

$$\begin{aligned}\int_{R_t}^{R_t+\varepsilon(h)} \mathbb{P}_x\left(\bar{T}_{R_t} \le h\right)\,\mathrm{d}\mu(x) &= \int_0^{\ln(1/h)} \frac{e^{-u}}{\sqrt{\pi u}} \int_0^{\sqrt{2hu}} f_t(y+R_t)\,\mathrm{d}y\,\mathrm{d}u\\ &\quad + \int_{\ln(1/h)}^{\infty} \frac{e^{-u}}{\sqrt{\pi u}} \int_0^{\varepsilon(h)} f_t(y+R_t)\,\mathrm{d}y\,\mathrm{d}u\\ &= \frac{f_t'(R_t)}{\sqrt{\pi}} \int_0^{\ln(1/h)} \frac{e^{-u}}{\sqrt{u}} \left[\frac{y^2}{2}\right]_0^{\sqrt{2hu}} \mathrm{d}u + o(h) \qquad (2.14)\\ &\quad + \frac{f_t'(R_t)}{\sqrt{\pi}} \int_{\ln(1/h)}^{\infty} \frac{e^{-u}}{\sqrt{u}} \left[\frac{y^2}{2}\right]_0^{\varepsilon(h)} \mathrm{d}u\\ &\quad + o(\varepsilon(h)) \int_{\ln(1/h)}^{\infty} \frac{e^{-u}}{\sqrt{u}}\,\mathrm{d}u. \qquad (2.15)\end{aligned}$$

Equality (2.14) comes from the fact that $\int_0^{\sqrt{2hu}} o(y)\,\mathrm{d}y = o(h)$ and $\int_0^{\infty} \frac{e^{-u}}{\sqrt{u}}\,\mathrm{d}u$ is finite. Equality (2.15) is given by $\int_0^{\varepsilon(h)} o(y)\,\mathrm{d}y = o(\varepsilon(h))$. Then

$$\begin{aligned}\int_{R_t}^{R_t+\varepsilon(h)} \mathbb{P}_x\left(\bar{T}_{R_t} \le h\right)\,\mathrm{d}\mu_t(x) &= \frac{f_t'(R_t)}{\sqrt{\pi}} \int_0^{\ln(1/h)} \frac{e^{-u}}{\sqrt{u}} hu\,\mathrm{d}u + o(h)\\ &\quad + \left(\frac{f_t'(R_t)\varepsilon(h)^2}{2\sqrt{\pi}} + o(\varepsilon(h))\right) \int_{\ln(1/h)}^{\infty} \frac{e^{-u}}{\sqrt{u}}\,\mathrm{d}u\\ &= h\frac{f_t'(R_t)}{\sqrt{\pi}} \int_0^{\ln(1/h)} \sqrt{u}e^{-u}\,\mathrm{d}u.\\ &\quad + o(\varepsilon(h)) \int_{\ln(1/h)}^{\infty} \frac{e^{-u}}{\sqrt{u}}\,\mathrm{d}u + o(h). \qquad (2.16)\end{aligned}$$

As $\int_{R_t+\varepsilon(h)}^{\infty} \mathbb{P}_x\left(\bar{T}_{R_t} \le h\right)\,\mathrm{d}\mu_t(x) = o(h)$ thanks to Lemma 2.4.1, one can conclude

$$\begin{aligned}\mathbb{P}_{\mu_t}\left(\bar{T}_{R_t} \le h\right) &= h\frac{f_t'(R_t)}{\sqrt{\pi}} \int_0^{\ln(1/h)} \sqrt{u}e^{-u}\,\mathrm{d}u + o(\varepsilon(h)) \int_{\ln(1/h)}^{\infty} \frac{e^{-u}}{\sqrt{u}}\,\mathrm{d}u + o(h)\\ &\ge \mathbb{P}_{\mu_t}(T_{R_t} \le h). \qquad (2.17)\end{aligned}$$

• Step 2: Define a lower bound of the probability to reach $R_t$.

Let us assume $\mathrm{d}\hat{\rho}_s = \mathrm{d}B_s + \frac{1}{2R_t}\,\mathrm{d}s$, and $\hat{\rho}_0 = x$. The probability density function of $\hat{T}_{R_t}$, beginning at $x$, is known [BS02] and given by,

$$\hat{p}_{x\to R_t}(s) = \exp\left(\frac{|R_t - x|}{2R_t} - \frac{s}{8R_t^2}\right)\frac{|x - R_t|}{\sqrt{2\pi s^3}}\exp -\frac{(R_t - x)^2}{2s}.$$

Then

$$\mathbb{P}_x(\hat{T}_{R_t} \le h) = e^{\frac{x-R_t}{2R_t}}\frac{x - R_t}{\sqrt{2\pi}}\int_0^h \frac{e^{-\frac{s}{8R_t^2} - \frac{(R_t-x)^2}{2s}}}{s^{3/2}}\,\mathrm{d}s.$$

First cut the probability starting to the measure $\mu_t$ in two, as in (2.13).

Since $e^{-\frac{s}{8R_t^2}} \ge e^{-\frac{h}{8R_t}}$ and $e^{\frac{x-R_t}{2R_t}} \ge 1$, the probability starting from $x \in [R_t, R_t + \varepsilon(h)[$ verifies

$$\int_{R_t}^{R_t+\varepsilon(h)} \mathbb{P}_x(\hat{T}_{R_t} \le h)\,\mathrm{d}\mu_s(x) \ge e^{-\frac{h}{8R_t}}\int_{R_t}^{R_t+\varepsilon(h)} \frac{x - R_t}{\sqrt{2\pi}}\int_0^h \frac{e^{-\frac{(R_t-x)^2}{2s}}}{s^{3/2}}\,\mathrm{d}s\,\mathrm{d}\mu_t(x).$$

Here, one recognizes $\int_{R_t}^{R_t+\varepsilon(h)} \mathbb{P}_x\left(\bar{T}_{R_t} \le h\right)\,\mathrm{d}\mu_t(x)$, computed in the step 1 and Equation (2.16) gives the following result

$$\begin{aligned}\int_{R_t}^{R_t+\varepsilon(h)} \mathbb{P}_x(\hat{T}_{R_t} \le h)\,\mathrm{d}\mu_t(x) \ge h\frac{e^{-\frac{h}{8R_t}} f_t'(R_t)}{\sqrt{\pi}}\int_0^{\ln(1/h)} \sqrt{u}e^{-u}\,\mathrm{d}u \\ + o(\varepsilon(h))\int_{\ln(1/h)}^{\infty}\frac{e^{-u}}{\sqrt{u}}\,\mathrm{d}u + o(h).\end{aligned}$$

For $h$ small enough, Lemma 2.4.2 informs that this probability starting from $x \in [R_t + \varepsilon(h), \infty[$ vanishes, and one concludes

$$\begin{aligned}\mathbb{P}_{\mu_t}(T_{R_t} \le h) &\ge \mathbb{P}_{\mu_t}\left(\hat{T}_{R_t} \le h\right) \\ &\ge h e^{-\frac{h}{8R_t}}\frac{f_t'(R_t)}{\sqrt{\pi}}\int_0^{\ln(1/h)} \sqrt{u}e^{-u}\,\mathrm{d}u + o(\varepsilon(h))\int_{\ln(1/h)}^{\infty}\frac{e^{-u}}{\sqrt{u}}\,\mathrm{d}u + o(h).\end{aligned} \tag{2.18}$$

• Step 3: Computation of $\partial_t\mathbb{E}(L_t)$.

From (2.17),

$$\lim_{h\to 0}\frac{\mathbb{P}_{\mu_t}(T_{R_t} < h)}{h} \le \lim_{h\to 0}\frac{f_t'(R_t)}{\sqrt{\pi}}\int_0^{\ln(1/h)} \sqrt{u}e^{-u}\,\mathrm{d}u + \frac{o(\varepsilon(h))}{h}\int_{\ln(1/h)}^{\infty}\frac{e^{-u}}{\sqrt{u}}\,\mathrm{d}u + o(1).$$

Recognizing Gamma function,

$$\lim_{h\to 0}\int_0^{\ln(1/h)} \sqrt{u}e^{-u}\,\mathrm{d}u = \Gamma(3/2) = \frac{\sqrt{\pi}}{2}.$$

Note that $\lim_{h\to 0} \frac{o(\varepsilon(h))}{h} \int_{\ln(1/h)}^{\infty} \frac{e^{-u}}{\sqrt{u}}\,\mathrm{d}u = 0$. Indeed $0 \leq \lim_{h\to 0} \frac{o(\varepsilon(h))}{h} \int_{\ln(1/h)}^{\infty} \frac{e^{-u}}{\sqrt{u}}\,\mathrm{d}u$ and

$$\lim_{h\to 0} \frac{o(\varepsilon(h)}{h\sqrt{\ln(1/h)}} \int_{\ln(1/h)}^{\infty} e^{-u}\,\mathrm{d}s = \lim_{h\to 0} \frac{o(\varepsilon(h)}{h\sqrt{\ln(1/h)}} e^{-\ln(1/h)} = 0. \tag{2.19}$$

Hence

$$\mathbb{E}(L_t') \leq (n_0 - \mathbb{E}(L_t))\frac{f_t'(R_t)}{2}.$$

From (2.18) and using Gamma function properties and (2.19),

$$\begin{aligned}\lim_{h\to 0} \frac{\mathbb{P}_{\mu_t}(T_{R_t} < h)}{h} &\geq \lim_{h\to 0} \frac{e^{\frac{\varepsilon(h)}{2R_t}} f_t'(R_t)}{\sqrt{\pi}} \int_0^{\ln(1/h)} \sqrt{u}e^{-u}\,\mathrm{d}u + \frac{o(\varepsilon(h))}{h} \int_{\ln(1/h)}^{\infty} \frac{e^{-u}}{\sqrt{u}}\,\mathrm{d}u + o(1) \\ &\geq (n_0 - \mathbb{E}(L_t))\frac{f_t'(R_t)}{2}.\end{aligned}$$

In conclusion

$$\partial_t \mathbb{E}(L_t) = (n_0 - \mathbb{E}(L_t))\frac{f_t'(R_t)}{2}.$$

□

**Lemma 2.4.1.** *Let $\varepsilon(h) = \sqrt{2h\ln(1/h)}$, then*

$$\int_{R_t+\varepsilon(h)}^{\infty} \mathbb{P}_x\left(\bar{T}_{R_t} \leq h\right)\,\mathrm{d}\mu_t(x) = o(h).$$

*Proof.* Using Equation (2.12) and $R_t + \varepsilon(h) \leq x$, one gets

$$\begin{aligned}0 \leq \mathbb{P}_x\left(\bar{T}_{R_t} \leq h\right) &\leq \int_{\frac{(R_t-x)^2}{2h}}^{\infty} \frac{e^{-u}}{\sqrt{u}}\,\mathrm{d}u \\ &\leq \frac{\sqrt{2h}}{(x-R_t)} \int_{\frac{(R_t-x)^2}{2h}}^{\infty} e^{-u}\,\mathrm{d}u \\ &\leq \frac{\sqrt{2h}}{\varepsilon(h)} e^{-\frac{\varepsilon(h)^2}{2h}}.\end{aligned}$$

Note that $\frac{\varepsilon(h)^2}{2h} = \ln(1/h)$,

$$0 \leq \int_{\frac{(R_t-x)^2}{2h}}^{\infty} \frac{e^{-u}}{\sqrt{u}}\,\mathrm{d}u \leq \frac{h}{\sqrt{\ln(1/h)}}.$$

And $\lim_{h\to 0} \frac{\frac{h}{\sqrt{\ln(1/h)}}}{h} = 0$. Then $\int_{\frac{(R_t-x)^2}{2h}}^{\infty} \frac{e^{-u}}{\sqrt{u}}\,\mathrm{d}u = o(h)$, for $x \in [R_t + \varepsilon(h), \infty[$.
As $\mu_t$ is a probability $\int_{R_t+\varepsilon(h)}^{\infty} \mathrm{d}\mu_t(x)$ is finite. This concludes the proof.

□

**Lemma 2.4.2.** *Let $R_t > \frac{1}{2}$, $h$ such that $\frac{1}{2R_t} - \frac{1}{8h} \leq \ln h$ and $\varepsilon(h) = \sqrt{2h\ln(1/h)}$. Then*

$$\int_{R_t+\varepsilon(h)}^{\infty} \mathbb{P}_x(\hat{T}_{R_t} \leq h)\,\mathrm{d}\mu_t(x) = o(h).$$

*Proof.* Using the variable $u = \frac{(R_t-x)^2}{2s}$ and $e^{-\frac{s}{8R_t^2}} \leq 1$,

$$\int_{R_t+\varepsilon(h)}^{\infty} \mathbb{P}_x(\hat{T}_{R_t} \leq h)\,\mathrm{d}\mu_t(x) = \int_{R_t+\varepsilon(h)}^{\infty} \frac{e^{\frac{x-R_t}{2R_t}}}{\sqrt{\pi}} \int_{\frac{(x-R_t)^2}{2h}}^{\infty} \frac{e^u}{\sqrt{u}}\,\mathrm{d}u\,\mathrm{d}\mu_t(x).$$

Unlike to the proof of Lemma 2.4.1, maximize $\frac{1}{\sqrt{u}}$ is insufficient to conclude because of the term $e^{x/2R_t}$ explode as $x$ tends to infinity. We shall control this number with $\int_{\frac{(x-R_t)^2}{2h}}^{\infty} \frac{e^u}{\sqrt{u}}\,\mathrm{d}u$.

If $h$ is such that $R_t+\varepsilon(h) \leq 2R_t$, we aim to show $\int_{R_t+\varepsilon(h)}^{2R_t} \mathbb{P}_x(\hat{T}_{R_t} < h)\,\mathrm{d}\mu_t(x) \leq o(h)$ and $\int_{2R_t}^{\infty} \mathbb{P}_x(\hat{T}_{R_t} < h)\,\mathrm{d}\mu_t(x) \leq o(h)$, either the proof of the second integrale is sufficient.

- As $x > 2R_t$, $\frac{(x-R_t)^2}{2h} > \frac{x}{8h}$,

$$\begin{aligned}\int_{2R_t}^{\infty} \frac{e^{\frac{x-R_t}{2R_t}}}{\sqrt{\pi}} \int_{\frac{(x-R_t)^2}{2h}}^{\infty} \frac{e^u}{\sqrt{u}}\,\mathrm{d}u\,\mathrm{d}\mu_t(x) &\leq e^{-1/2} \int_{2R_t}^{\infty} \frac{\sqrt{2h}e^{x/2R_t}}{\sqrt{\pi}(x-R_t)} \int_{x/8h}^{\infty} e^{-u}\,\mathrm{d}u\,\mathrm{d}\mu_t(x) \\ &\leq \frac{\sqrt{2h}e^{-1/2}}{R_t\sqrt{\pi}} \int_{2R_t}^{\infty} e^{x\left(\frac{1}{2R_t}-\frac{1}{8h}\right)}\,\mathrm{d}\mu_t(x).\end{aligned}$$

Consider $h$ such that $\frac{1}{2R_t} - \frac{1}{8h} \leq \ln h$, then

$$\begin{aligned}\int_{2R_t}^{\infty} \frac{e^{\frac{x-R_t}{2R_t}}}{\sqrt{\pi}} \int_{\frac{(x-R_t)^2}{2h}}^{\infty} \frac{e^u}{\sqrt{u}}\,\mathrm{d}u\,\mathrm{d}\mu_t(x) &\leq \frac{\sqrt{2h}e^{-1/2}}{R_t\sqrt{\pi}} \int_{2R_t}^{\infty} e^{-x\ln(1/h)}\,\mathrm{d}\mu_t(x) \\ &\leq \frac{\sqrt{2h}e^{-1/2}}{R_t\sqrt{\pi}} \int_{2R_t}^{\infty} e^{R_t\ln h}\,\mathrm{d}\mu_t(x).\end{aligned}$$

As $\mu_t$ is a probability distribution, $\frac{h^{R_t}\sqrt{2h}e^{-1/2}}{R_t\sqrt{\pi}} \int_{2R_t}^{\infty} \mathrm{d}\mu_t(x) = o(h)$, as soon as $R_t > \frac{1}{2}$.

- As $R_t + \varepsilon(h) \leq x \leq 2R_t$, and using Lemma 2.4.1

$$\begin{aligned}\int_{R_t+\varepsilon(h)}^{2R_t} \frac{e^{\frac{x-R_t}{2R_t}}}{\sqrt{\pi}} \int_{\frac{(x-R_t)^2}{2h}}^{\infty} \frac{e^u}{\sqrt{u}}\,\mathrm{d}u\,\mathrm{d}\mu_t(x) &\leq \frac{e^{1/2}}{\sqrt{\pi}} \int_{R_t+\varepsilon(h)}^{2R_t} \int_{\frac{(x-R_t)^2}{2h}}^{\infty} \frac{e^u}{\sqrt{u}}\,\mathrm{d}u\,\mathrm{d}\mu_t(x) \\ &\leq \frac{e^{1/2}}{\sqrt{\pi}} \int_{R_t+\varepsilon(h)}^{\infty} \int_{\frac{(x-R_t)^2}{2h}}^{\infty} \frac{e^u}{\sqrt{u}}\,\mathrm{d}u\,\mathrm{d}\mu_t(x) = o(h).\end{aligned}$$

In addition $\int_{R_t+\varepsilon(h)}^{\infty} \mathbb{P}_x(\hat{T}_{R_t} < h)\,\mathrm{d}\mu_t(x) \geq 0$. That concludes the proof.

□

### 2.4.2 Proof of Lemma 2.3.4

*Proof.* To demonstrate this lemma we use the same way than in Theorem 2.3.1. As suggested in Sketch of proof, using Girsanov formula we get an upper and lower bounds of $\mathbb{P}_{\mu_t}(\bar{T}_{R_t} < h)$, depending on the probability of first hitting time in case of Brownian motion given in Equations (2.17) and (2.18).

Let us introduce some notations, $\mathbb{P}^P$, $\mathbb{E}^P$ denote respectively, the probability and the expectation with respect to the law of $\bar{\rho}_s$ where $\bar{\rho}_s$ is an O.U. process defined in (2.10). And $\mathbb{P}^Q$, $\mathbb{E}^Q$ denote, respectively, probability and expectation, with respect to a Brownian motion law. Thanks to Girsanov formula, one gets

$$\begin{aligned}\mathbb{P}_x^P\left(\bar{T}_{R_t} \le h\right) &= \mathbb{E}_x^Q\left(\mathbb{1}_{\bar{T}_{R_t}<h}\frac{\mathrm{d}P_h}{\mathrm{d}Q_h}\right)\\ &= \mathbb{E}_x^Q\left(\mathbb{1}_{\bar{T}_{R_t}<h}\exp\left(-\mathbb{E}(L_t)\int_0^h \bar{\rho}_s\,\mathrm{d}\bar{\rho}_s - \frac{\mathbb{E}(L_t)^2}{2}\int_0^h (\bar{\rho}_s)^2\,\mathrm{d}s\right)\right).\end{aligned}$$

Note that $\mathrm{d}\left(\bar{\rho}_s\right)^2 = 2\bar{\rho}_s\,\mathrm{d}\bar{\rho}_s + \mathrm{d}s$ and if $h > T_{R_t}$, then $\bar{\rho}_h = R_t$, thus

$$\begin{aligned}\mathbb{P}_x^P\left(\bar{T}_{R_t} \le h\right) &= \mathbb{E}_x^Q\left(\mathbb{1}_{\bar{T}_{R_t}<h}\exp\left(-\frac{\mathbb{E}(L_t)\left((\bar{\rho}_h)^2 - x^2 - h\right)}{2} - \frac{\mathbb{E}(L_t)^2}{2}\int_0^h (\bar{\rho}_s)^2\,\mathrm{d}s\right)\right)\\ &= e^{\frac{\mathbb{E}(L_t)(x^2+h-R_t^2)}{2}}\mathbb{E}_x^Q\left(\mathbb{1}_{\bar{T}_{R_t}<h}\exp\left(-\frac{\mathbb{E}(L_t)^2}{2}\int_0^h (\bar{\rho}_s)^2\,\mathrm{d}s\right)\right). \qquad (2.20)\end{aligned}$$

• Step 1: Define an upper bound of the probability to reach $R_t$.

As $(\bar{\rho}_s)^2 > 0$,

$$\mathbb{P}_x^P(\bar{T}_{R_t} \le h) \le \exp\left(\frac{\mathbb{E}(L_t)(x^2+h-R_t^2)}{2}\right)\mathbb{P}_x^Q\left(\bar{T}_{R_t} \le h\right).$$

Next, cutting the probability starting from the measure $\mu_t$ in two, denote $M = \int_{R_t}^{R_t+\varepsilon(h)} \mathbb{P}_x^P(\bar{T}_{R_t} \le h)\,\mathrm{d}\mu_t(x)$, hence

$$M \le \exp\left(\frac{\mathbb{E}(L_t)(\varepsilon(h)^2 + h + 2\varepsilon(h)R_t)}{2}\right)\int_{R_t}^{R_t+\varepsilon(h)} \mathbb{P}_x^Q(\bar{T}_{R_t} \le h)\,\mathrm{d}\mu_t(x).$$

Using (2.16),

$$\begin{aligned}M \le &\exp\left(\frac{\mathbb{E}(L_t)(\varepsilon(h)^2 + h + 2\varepsilon(h)R_t)}{2}\right)\frac{\partial_x f_t(R_t, \mathbb{E}(L_t))h}{\sqrt{\pi}}\int_0^{\ln(1/h)}\sqrt{s}e^{-s}\,\mathrm{d}s\\ &+ \exp\left(\frac{\mathbb{E}(L_t)(\varepsilon(h)^2 + h + 2\varepsilon(h)R_t)}{2}\right)\int_{\ln(1/h)}^{\infty}\frac{o(\varepsilon(h))e^{-s}}{\sqrt{s}}\,\mathrm{d}s + o(h).\end{aligned}$$

By Lemma 2.4.3, one finally obtains

$$\mathbb{P}^P_{\mu_t}(\bar{T}_{R_t} \leq h) \leq h\frac{\exp\left(\frac{\mathbb{E}(L_t)(\varepsilon(h)^2+h+2\varepsilon(h)R_t)}{2}\right)\partial_x f_t(R_t,\mathbb{E}(L_t))}{\sqrt{\pi}}\int_0^{\ln(1/h)}\sqrt{s}e^{-s}\,\mathrm{d}s + \int_{\ln(1/h)}^{\infty}\frac{o(\varepsilon(h))e^{-s}}{\sqrt{s}}\,\mathrm{d}s + o(h). \tag{2.21}$$

• Step 2: Define a lower bound of the probability to reach $R_t$.

In (2.20), let $Y = \exp\left(-\frac{\mathbb{E}(L_t)^2}{2}\int_0^h(\bar{\rho}_s)^2\,\mathrm{d}s\right) \in [0,1]$,

$$\begin{aligned}\mathbb{E}^Q_x\left(\mathbb{1}_{\bar{T}_{R_t}<h}Y\right) &\geq \mathbb{E}^Q_x\left(\mathbb{1}_{\bar{T}_{R_t}<h}(Y-1)\right) + \mathbb{E}^Q_x\left(\mathbb{1}_{\bar{T}_{R_t}<h}\right)\\ &\geq \mathbb{P}^Q_x\left(\bar{T}_{R_t}\leq h\right) - \mathbb{E}^Q_x\left(\mathbb{1}_{\bar{T}_{R_t}<h}(1-Y)\right).\end{aligned}$$

To derive a lower bound of $\int_{R_t}^{\infty}\exp\left(\frac{\mathbb{E}(L_t)(x^2+h-R_t^2)}{2}\right)\mathbb{P}^Q_x(T_{R_t}\leq h)\,\mathrm{d}\mu_t(x)$, note that $\exp\left(\frac{\mathbb{E}(L_t)(x^2+h-R_t^2)}{2}\right)\geq 1$. Then, using (2.17), one obtains

$$\int_{R_t}^{\infty}e^{\frac{\mathbb{E}(L_t)(x^2+h-R^2)}{2}}\mathbb{P}^Q_x(\bar{T}_{R_t}\leq h)\,\mathrm{d}\mu_t(x) \geq \frac{\partial_x f_t(R_t,\mathbb{E}(L_t))h}{\sqrt{\pi}}\int_0^{\ln(1/h)}\sqrt{s}e^{-s}\,\mathrm{d}s + \int_{\ln(1/h)}^{\infty}\frac{o(\varepsilon(h))e^{-s}}{\sqrt{s}}\,\mathrm{d}s + o(h).$$

By Lemma 2.4.4, one concludes

$$\mathbb{P}^P_{\mu_t}\left(\bar{T}_{R_t}\leq h\right) \geq \frac{\partial_x f_t(R_t,\mathbb{E}(L_t))h}{\sqrt{\pi}}\int_0^{\ln(1/h)}\sqrt{s}e^{-s}\,\mathrm{d}s + \int_{\ln(1/h)}^{\infty}\frac{o(\varepsilon(h))e^{-s}}{\sqrt{s}}\,\mathrm{d}s + o(h). \tag{2.22}$$

• Step 3: Computation of $\partial_t\mathbb{E}(L_t)$.

For the lower bound, using (2.22)

$$\lim_{h\to 0}\frac{\mathbb{P}_{\mu_t}(\bar{T}_{R_t}<h)}{h} \geq \lim_{h\to 0}\left[\frac{\partial_x f_t(R_t,\mathbb{E}(L_t))}{\sqrt{\pi}}\int_0^{\ln(1/h)}\sqrt{s}e^{-s}\,\mathrm{d}s + \int_{\ln(1/h)}^{\infty}\frac{o(\varepsilon(h))e^{-s}}{h\sqrt{s}}\,\mathrm{d}s + o(1)\right].$$

By Gamma function properties and (2.19),

$$\lim_{h\to 0}\frac{\mathbb{P}_{\mu_t}(\bar{T}_{R_t}<h)}{h} \geq \frac{\partial_x f_t(R_t,\mathbb{E}(L_t))}{2}.$$

For the upper bounds, from (2.21), by Gamma function properties and (2.19),

$$\lim_{h\to 0}\frac{\mathbb{P}_{\mu_t}(\bar{T}_{R_t}<h)}{h}\leq \lim_{h\to 0}\left[\frac{\partial_x f_t(R_t,\mathbb{E}(L_t))h}{\sqrt{\pi}}\int_0^{\ln(1/h)}\sqrt{s}e^{-s}\,\mathrm{d}s+\int_{\ln(1/h)}^{\infty}\frac{o(\varepsilon(h))e^{-s}}{\sqrt{s}}\,\mathrm{d}s+o(h)\right]$$
$$=\frac{\partial_x f_t(R_t,\mathbb{E}(L_t))}{2}.$$

That proves the lemma. □

**Lemma 2.4.3.** *Let $R_t>\frac{1}{\sqrt{2}}$, $h$ such that $\frac{4h\mathbb{E}(L_t)-1}{8h}\leq \ln h$ and $\varepsilon(h)=\sqrt{2h\ln(1/h)}$. Then*

$$\int_{R_t+\varepsilon(h)}^{\infty}\mathbb{P}_x^P(\bar{T}_{R_t}\leq h)\,\mathrm{d}\mu_t(x)=o(h).$$

*Proof.* As under $Q$, $\bar{\rho}$ is a brownian motion, using the variable $u=\frac{(R_t-x)^2}{2s}$, we aim to demonstrate

$$\int_{R+\varepsilon(h)}^{\infty}\exp\left(\frac{\mathbb{E}(L_t)(x^2+h-R_t^2)}{2}\right)\int_{\frac{(R_t-x)^2}{2h}}^{\infty}\frac{e^u}{\sqrt{\pi u}}\,\mathrm{d}u\,\mathrm{d}\mu_t(x)=o(h).$$

Here, as in Lemma 2.4.2, the term $e^{\frac{\mathbb{E}(L_t)x^2}{2}}$ explodes as $x$ tends to infinity. If $h$ is such that $2R_t>R_t+\varepsilon(h)$, we cut the integral in two, either we compute only step 1.

- Step 1: Suppose $x>2R_t$, $\frac{(x-R_t)^2}{2h}>\frac{x^2}{8h}$, hence

$$\int_{2R_t}^{\infty}e^{\mathbb{E}(L_t)(x^2+h-R_t^2)/2}\int_{\frac{(R_t-x)^2}{2h}}^{\infty}\frac{e^u}{\sqrt{\pi u}}\,\mathrm{d}u\,\mathrm{d}\mu_t(x)\leq\int_{2R_t}^{\infty}\frac{e^{\mathbb{E}(L_t)(x^2+h-R_t^2)/2}\sqrt{2h}}{\sqrt{\pi}(x-R_t)}\int_{\frac{x^2}{8h}}^{\infty}e^{-u}\,\mathrm{d}u\,\mathrm{d}\mu_t(x)$$
$$=\frac{\sqrt{2h}e^{\mathbb{E}(L_t)(h-R_t^2)/2}}{R\sqrt{\pi}}\int_{2R_t}^{\infty}e^{x^2\left(\frac{4h\mathbb{E}(L_t)-1}{8h}\right)}\,\mathrm{d}\mu_t(x).$$

Assume $h$ such that $\frac{4h\mathbb{E}(L_t)-1}{8h}\leq\ln h$, then

$$\int_{2R_t}^{\infty}e^{\mathbb{E}(L_t)(x^2+h-R_t^2)/2}\int_{\frac{(R_t-x)^2}{2h}}^{\infty}\frac{e^u}{\sqrt{\pi u}}\,\mathrm{d}u\,\mathrm{d}\mu_t(x)\leq\frac{\sqrt{2h}e^{\mathbb{E}(L_t)(h-R_t^2)/2}}{R_t\sqrt{\pi}}\int_{2R_t}^{\infty}e^{-R_t^2\ln(1/h)}\,\mathrm{d}\mu_t(x).$$

As $\mu_t$ is a probability distribution, $\int_{2R_t}^{\infty}\mathrm{d}\mu_t(x)<\infty$, $\frac{h^{R_t^2}\sqrt{2h}\exp\left(\mathbb{E}(L_t)(h-R_t^2)/2\right)}{R_t\sqrt{\pi}}\int_{2R_t}^{\infty}\mathrm{d}\mu_t(x)$ vanishes as $h$ tends to 0 and $R_t>\frac{1}{\sqrt{2}}$.

Step 2: Note that $0\leq e^{\mathbb{E}(L_t)(x^2+h-R_t^2)/2}\leq e^{\mathbb{E}(L_t)(R_t^2+h)/2}$. Using Lemma 2.4.1 and noting that $\int_{R_t+\varepsilon(h)}^{\infty}\mathbb{P}_x^P(\bar{T}_{R_t}\leq h)\,\mathrm{d}\mu_t(x)\geq 0$. This concludes the proof.

□

**Lemma 2.4.4.** *Let $R_t > \frac{1}{\sqrt{2}}$ and $h$ such that $\frac{8h(\mathbb{E}(L_t)+2)-1}{8h} \leq \ln h$. Then*

$$\int_{R_t}^{\infty} \exp\left(\frac{\mathbb{E}(L_t)(x^2+h-R_t^2)}{2}\right) \mathbb{E}_x^Q\left(\mathbb{1}_{\bar{T}_{R_t}<h}(1-Y)\right) \mathrm{d}\mu_t(x) = o(h).$$

*Proof.* One can use Young Inequality to maximise $\mathbb{E}_x^Q\left(\mathbb{1}_{T_{R_t}<h}(1-Y)\right)$, let $\eta = \eta(x,h) > 0$,

$$\mathbb{E}_x^Q\left(\mathbb{1}_{T_{R_t}<h}(1-Y)\right) \leq \frac{\mathbb{P}_x^Q\left(\bar{T}_{R_t}<h\right)}{2\eta} + \frac{\eta\mathbb{E}_x^Q\left((1-Y)^2\right)}{2}.$$

- Step 1: considere $\alpha \in ]0,1[$ and $\eta = \frac{1}{x^4 e^{\mathbb{E}(L_t)x^2/2} h^\alpha}$, we shall prove

$$\int_{2R_t}^{\infty} \exp\left(\frac{\mathbb{E}(L_t)(x^2+h-R_t^2)}{2}\right) \frac{\mathbb{P}_x^Q\left(\bar{T}_{R_t}\leq h\right)}{2\eta} \mathrm{d}\mu_t(x) = o(h).$$

Replace $\eta$ by its value,

$$\int_{R_t}^{\infty} e^{\frac{\mathbb{E}(L_t)(x^2+h-R_t^2)}{2}} \frac{\mathbb{P}_x^Q\left(\bar{T}_{R_t}\leq h\right)}{2\eta} \mathrm{d}\mu_t(x) = h^\alpha \frac{e^{\frac{\mathbb{E}(L_t)(h-R_t^2)}{2}}}{2} \int_{R_t}^{\infty} x^4 e^{\mathbb{E}(L_t)x^2} \mathbb{P}_x^Q\left(\bar{T}_{R_t}\leq h\right) \mathrm{d}\mu_t(x).$$

In the same way as in the previous lemma, if $h$ is enough small, such that $R_t+\varepsilon(h) \leq 2R_t$, we cut the integral in two, either the second calculus is enough. Using (2.17)

$$\begin{aligned}\int_{R_t}^{2R_t} e^{\frac{\mathbb{E}(L_t)(x^2+h-R_t^2)}{2}} \frac{\mathbb{P}_x^Q\left(\bar{T}_{R_t}\leq h\right)}{2\eta} \mathrm{d}\mu_t(x) &\leq h^\alpha 8R_t^4 e^{\frac{\mathbb{E}(L_t)(h+7R_t^2)}{2}} \int_{R_t}^{2R_t} \mathbb{P}_x^Q\left(\bar{T}_{R_t}\leq h\right) \mathrm{d}\mu_t(x) \\ &= o(h).\end{aligned}$$

As $h > 0$, note that

$$\begin{aligned}\int_{2R_t}^{\infty} e^{\frac{\mathbb{E}(L_t)(x^2+h-R_t^2)}{2}} \frac{\mathbb{P}_x^Q\left(\bar{T}_{R_t}\leq h\right)}{2\eta} \mathrm{d}\mu_t(x) &\leq h^\alpha \int_{2R_t}^{\infty} e^{(\mathbb{E}(L_t)+2)x^2+\frac{\mathbb{E}(L_t)(h-R_t^2)}{2}} \mathbb{P}_x^Q\left(\bar{T}_{R_t}\leq h\right) \mathrm{d}\mu_t(x) \\ &\leq e^{\mathbb{E}(L_t)\frac{h-R_t^2}{2}} h^\alpha \int_{2R_t}^{\infty} e^{(\mathbb{E}(L_t)+2)x^2} \int_{\frac{x}{8h}}^{\infty} e^{-u}\,\mathrm{d}u\,\mathrm{d}\mu_t(x).\end{aligned}$$

Moreover, as $h$ such that $\frac{8h(\mathbb{E}(L_t)+2)-1}{8h} \leq \ln h$, and using the same methode to demonstrate Lemma 2.4.3, $\int_{2R_t}^{\infty} e^{\frac{\mathbb{E}(L_t)(x^2+h-R_t^2)}{2}} \frac{\mathbb{P}_x^Q\left(\bar{T}_{R_t}\leq h\right)}{2\eta} \mathrm{d}\mu_t(x) \leq o(h)$.

In addition $\int_{R_t}^{\infty} e^{\frac{\mathbb{E}(L_t)(x^2+h-R_t^2)}{2}} \frac{\mathbb{P}_x^Q\left(\bar{T}_{R_t}\leq h\right)}{2\eta} \mathrm{d}\mu_t(x) \geq 0$. This integrale vanishes as $h$ tends to 0.

• Step 2: for $\eta = \frac{1}{x^4 e^{\mathbb{E}(L_t)x^2/2} h^\alpha}$, where $\alpha \in ]0,1[$, we shall prove

$$\int_{R_t}^{\infty} \frac{\eta \mathbb{E}_x^Q \left((1-Y)^2\right)}{2} \, \mathrm{d}\mu_t(x) = o(h).$$

First, prove that $\mathbb{E}^Q \left((1-Y)^2\right) = \frac{\mathbb{E}(L_t)^4 x^4}{4} h^2 + o(h^2, x^4)$.

Note that $Y = e^X$, where $X = \frac{\mathbb{E}(L_t)^2}{2} \int_0^h (\rho_s)^2 \, \mathrm{d}s$, then $1 - Y \leq X$. Hence $\mathbb{E}_x^Q \left((1-Y)^2\right) \leq \mathbb{E}_x^Q \left( \frac{L_t^4}{4} \left( \int_0^h (\rho_s)^2 \, \mathrm{d}s \right)^2 \right)$. As $\bar{\rho}_s$ is a brownian motion starting at $x$ under the probability $Q$, denote $\bar{\rho}_s = B_s + x$. Developing the square

$$\begin{aligned}
\mathbb{E}_x^Q \left((1-Y)^2\right) &= \frac{\mathbb{E}(L_t)^4}{4} \left[ x^4 h^2 + 4x^2 \mathbb{E}\left[ \left( \int_0^h B_s \, \mathrm{d}s \right)^2 \right] + \mathbb{E}\left[ \left( \int_0^h B_s^2 \, \mathrm{d}s \right)^2 \right] \right. \\
&\quad \left. + 4x^3 h \mathbb{E}\left[ \int_0^h B_s \, \mathrm{d}s \right] + 2x^2 h \mathbb{E}\left[ \int_0^h B_s^2 \, \mathrm{d}s \right] + 4x \mathbb{E}\left[ \int_0^h B_s \, \mathrm{d}s \int_0^h B_s^2 \, \mathrm{d}s \right] \right]. \\
&\leq \frac{\mathbb{E}(L_t)^4}{4} \left( x^4 h^2 + 4x^2 \frac{h^3}{3} + x^2 h^3 + h^4 + 4x \frac{h^2}{\sqrt{3}} \left(h^3\right)^{1/2} \right).
\end{aligned}$$

Indeed

1. Note that $\int_0^h B_s \, \mathrm{d}s \sim \mathcal{N}\left(0, \frac{h^3}{3}\right)$, then $4x^2 \mathbb{E}\left( \left( \int_0^h B_s \, \mathrm{d}s \right)^2 \right) = 4x^2 \frac{h^3}{3}$
2. Thanks to Cauchy Schwarz Inequality and the fact that the foorth moment of a gausian distribution with 0 mean and $s$ variance is $s^3$, $\mathbb{E}\left( \left( \int_0^h B_s^2 \, \mathrm{d}s \right)^2 \right) \leq h^4$;
3. $4x^3 h \mathbb{E}\left( \int_0^h B_s \, \mathrm{d}s \right) = 0$;
4. $2x^2 h \mathbb{E}\left( \int_0^h B_s^2 \, \mathrm{d}s \right) = x^2 h^3$;
5. From Cauchy Schwarz Inequality, 1. and 2., $4x \mathbb{E}\left( \int_0^h B_s \, \mathrm{d}s \int_0^h B_s^2 \, \mathrm{d}s \right) \leq 4x \frac{h^2}{\sqrt{3}} \left(h^3\right)^{1/2}$.

Secondly, as $\mu_t$ is a distribution of probability,

$$\int_{R_t}^{\infty} e^{\frac{\mathbb{E}(L_t)(x^2+h-R_t^2)}{2}} \frac{\eta \mathbb{E}(L_t)^4 x^4 h^2}{8} \, \mathrm{d}\mu_t(x) = \frac{\mathbb{E}(L_t)^4 e^{\mathbb{E}(L_t)(h-R_t^2)/2} h^{2-\alpha}}{8} \int_R^{\infty} \mathrm{d}\mu_t(x) = o(h).$$

This ends the proof. □

### 2.4.3 More details on Remark 2.3.3

Y. Hamana and H Matsumoto, in [BMR13], state, for $x > 1$

$$q_{x \to 1}(s) \approx \frac{x-1}{x} e^{-\frac{(x-1)^2}{2s}} \frac{\sqrt{x+s}}{s^{3/2}} \frac{1 + \ln(x)}{(1 + \ln(1 + s/x))(1 + \ln(s+x))},$$

where $q_{x \to 1}(s)$ is the density of the fist time for a Bessel process to reach 1 starting from $x$. Here $f \approx g$, means that there exist strictly positive constants $c_1$ and $c_2$, such

that $c_1 \leq \frac{f}{g} \leq c_2$. Note that, for a small $s$, our result reveals the dependance of the constants in the absorbing point, $R_t$.

For the sake of clarity, in what follows, we demonstrate that using their density of the hitting time of $R_t$ by a Bessel process (such as $\bar{\rho}$), we obtain, up to a multiplicative constant, the same result than using Girsanov's Theorem.

From scaling Property of the Bessel process, one gets for $x > R_t > 0$

$$\mathbb{P}_x\left(T_{R_t} \leq h\right) = \mathbb{P}_{x/R_t}(R_t^2 T_1 \leq h) = \mathbb{P}_{x/R_t}\left(T_1 \leq \frac{h}{R_t^2}\right).$$

In this study $s \in [0, h]$, and $x > R_t$,

$$\begin{aligned}
q_{x\to R_t}(s) &= q_{x/R_t\to 1}\left(s\right) \\
&\approx \frac{x - R_t}{\sqrt{2\pi}\left(sR_t^2\right)^{3/2}} \exp\left(-\frac{(x-R_t)^2}{2sR_t^2}\right) \\
&\quad \times \frac{R_t^3 \frac{\sqrt{\frac{2\pi x}{R_t}}\sqrt{1+sR_t/x}}{x}\left(1+\ln\left(\frac{x}{R_t}\right)\right)}{\left(1+\ln\left(1+\frac{sR_t}{x}\right)\right)\left(1+\ln\left(\frac{sR_t}{x}+1\right)+\ln\left(\frac{x}{R_t}\right)\right)} \\
&= \frac{x - R_t}{\sqrt{2\pi}\left(sR_t^2\right)^{3/2}} \exp\left(-\frac{(x-R_t)^2}{2sR_t^2}\right) \\
&\quad \times \frac{R_t^{5/2}\sqrt{2\pi}}{\sqrt{x}} \frac{\sqrt{1+\frac{sR_t}{x}}\left(1+\ln\left(\frac{x}{R_t}\right)\right)}{\left(1+\ln\left(1+\frac{sR_t}{x}\right)\right)\left(1+\ln\left(\frac{sR_t}{x}+1\right)+\ln\left(\frac{x}{R_t}\right)\right)} \\
&= p_{x\to R_t}(sR_t^2)\frac{R_t^{5/2}\sqrt{2\pi}}{\sqrt{x}} C(s, x, R_t),
\end{aligned}$$

where $p_{x\to R_t}(sR_t^2)$ is the density of the hitting time by a brownian motion define in (2.11). Using a taylor formula for ln and $\sqrt{.}$,

$$\begin{aligned}
C(s, x, R_t) &= \frac{\left(1+\frac{sR_t}{2x}+o\left(\frac{sR_t}{x}\right)\right)\left(1+\ln\left(\frac{x}{R_t}\right)\right)}{\left(1+\frac{sR_t}{x}+o\left(\frac{sR_t}{x}\right)\right)\left(1+\frac{sR_t}{x}+o\left(\frac{sR_t}{x}\right)+\ln\left(\frac{x}{R_t}\right)\right)} \\
&= 1 + \frac{3sR_t}{x2(1+\ln(x/R_t))} + o\left(\frac{sR_t}{x}\left(1+\ln\left(\frac{x}{R_t}\right)\right)\right).
\end{aligned}$$

Since we have not the same expression of the density, we develop $\mathbb{P}_x\left(T_{R_t} \leq h\right)$ in order to state the equality between the results.

For the upper bound, using the variable $u = sR_t^2$

$$\mathbb{P}_x\left(T_{R_t} \leq h\right) \approx \int_0^h p_{x\to R_t}(u)\frac{\sqrt{2\pi R_t}}{\sqrt{x}} C(u/R_t^2, x, R_t)\,\mathrm{d}u.$$

As $x \geq R_t$, one can obtain

$$\begin{aligned}
0 &\leq \mathbb{P}_x(T_{R_t} \leq h) \\
&\lesssim \int_0^h p_{x\to R_t}(u) \left(1 + \frac{3u}{R_t^2 2(1+\ln(R_t/R_t))} + o\left(u\frac{1+\ln\left(\frac{x}{R_t}\right)}{xR_t}\right)\right) \mathrm{d}u \\
&\leq \int_0^h p_{x\to R_t}(u)\,(1+o\,(1))\,\mathrm{d}u \\
&\leq \int_0^h p_{x\to R_t}(u)\,\mathrm{d}u + o(h).
\end{aligned}$$

Hence we obtain the same upper bound of $\mathbb{P}_{\mu_t}(T_{R_t} \leq h)$ up to a multiplicative constant.

For the lower bound, as $C \geq 1$,

$$\mathbb{P}_x(T_{R_t} \leq h) \gtrsim \int_0^h p_{x\to R_t}(u)\frac{\sqrt{R_t}}{\sqrt{x}}\,\mathrm{d}u.$$

Note that,

$$\begin{aligned}
\mathbb{P}_{\mu_t}(T_{R_t} \leq h) &= \int_{R_t}^{\infty} \mathbb{P}_x(T_{R_t} \leq h)\,\mathrm{d}\mu_t(x) \\
&\gtrsim \sqrt{\frac{R_t}{R_t+\varepsilon(h)}} \int_{R_t}^{R_t+\varepsilon(h)} \int_0^h p_{x\to R_t}(u)\,\mathrm{d}u\,\mathrm{d}\mu_t(x) \\
&+ \int_{R_t+\varepsilon(h)}^{\infty} \int_0^h p_{x\to R_t}(u)\sqrt{\frac{R_t}{x}}\,\mathrm{d}u\,\mathrm{d}\mu_t(x) \\
&\gtrsim \int_{R_t}^{R_t+\varepsilon(h)} \int_0^h p_{x\to R_t}(u)\,\mathrm{d}u\,\mathrm{d}\mu_t(x).
\end{aligned}$$

One obtains an equivalent lower bound of $\mathbb{P}_{\mu_t}(T_{R_t} \leq h)$ up to a multiplicative constant.

# Chapter 3

# Numerical studies

This chapter is devoted to numerical simulations concerning the model, describing the interaction between CTL and tumor nodule, studied in Chapter 2. Firstly, we illustrate solutions of the systems proposed in the previous chapter. Secondly, we investigate qualitative properties like phase transitions in case of success in nodule eradication.

The present model lets unknown the distribution $\mu_t$ of non scout CTL, those located are not on the nodule frontier, denoted $\mu_t$ at time $t$. To get numerical results, alternatively a Gamma distribution or a Quasi-Stationary Distribution are assumed for the CTL distribution.

## 3.1 Introduction

To decipher the importance of the attraction in CTL response against tumor nodule, a system of EDO was proposed in the previous chapter. This model postulates two different dynamics for CTL displacement: one without attraction modelized through a Brownian motion, and one with attraction due to an immunotherapy through an Ornstein-Uhlenbeck (O.U.) process.

Equations of the nodule (2.4) and (2.5) take into account the scout CTL number. This number, depending on hitting time for the process, is random. Hence, an equation on its expectation is developed, alternatively for a self-governing CTL displacements (Theorem 2.3.1) and for drifted CTL displacements (Theorem 2.3.5).

Then, the goal of this chapter is to determine the minimal CTL number leading to nodule eradication, with a nodule radius, at the beginning, equals to $R_0$. And determine if this number is lower under a simulated immunotherapy. To that end, a classical Euler method is used to find numerical approximations of the solutions of these tow systems: for all $t > 0$

1. for self-governing CTL displacements:

$$\begin{cases} R_t' = \left[\left(\lambda\delta R_t - \frac{\lambda\delta E}{2} - \frac{r^2\lambda\mathbb{E}(L_t)}{2E}\right)\mathbb{1}_{\delta(2R_tE-E^2)-r^2\mathbb{E}(L_t)>0} - \frac{r^2\mu\mathbb{E}(L_t)}{2E}\right]\mathbb{1}_{R_t'\geq 0} \\ \qquad + \left[\left(\frac{\lambda\delta R_t}{2} - \frac{\lambda\delta N^2}{2R_t} - \frac{r^2\lambda\mathbb{E}(L_t)}{2R_t}\right)\mathbb{1}_{\delta(R_t^2-N^2)-r^2\mathbb{E}(L_t)>0} - \frac{r^2\mu\mathbb{E}(L_t))}{2R_t}\right]\mathbb{1}_{R_t'<0,\ R_t>N} \quad ; \\ \partial_t\mathbb{E}(L_t) = (n_0+\mathbb{E}(L_t))\frac{f_t'(R_t)}{2} \end{cases} \tag{3.1}$$

2. for biased CTL displacements:

$$\begin{cases} R_t' = \left[\left(\lambda\delta R_t - \frac{\lambda\delta E}{2} - \frac{r^2\lambda\mathbb{E}(L_t)}{2E}\right)\mathbb{1}_{\delta(2R_tE-E^2)-r^2\mathbb{E}(L_t)>0} - \frac{r^2\mu\mathbb{E}(L_t)}{2E}\right]\mathbb{1}_{R_t'\geq 0} \\ \qquad + \left[\left(\frac{\lambda\delta R_t}{2} - \frac{\lambda\delta N^2}{2R_t} - \frac{r^2\lambda\mathbb{E}(L_t)}{2R_t}\right)\mathbb{1}_{\delta(R_t^2-N^2)-r^2\mathbb{E}(L_t)>0} - \frac{r^2\mu\mathbb{E}(L_t))}{2R_t}\right]\mathbb{1}_{R_t'<0,\ R_t>N} \quad . \\ (n_0+\mathbb{E}(L_t))\frac{f_t'\left(R_t-\frac{1}{2\mathbb{E}(L_t)R_t},\mathbb{E}(L_t)\right)}{2} \leq \partial_t\mathbb{E}(L_t) \leq (n_0+\mathbb{E}(L_t))\frac{f_t'(R_t,\mathbb{E}(L_t))}{2} \end{cases} \tag{3.2}$$

A recall of parameter definitions and variable definitions are in Table 3.1.

| Variables | Description | |
|---|---|---|
| $R_t$ | nodule radius at $t$ | |
| $E_t$ | alive part thickness of the nodule at $t$ | |
| $N_t$ | necrotic part radius of the nodule at $t$ | |
| $A_t$ | number of alive cells at $t$ | |
| | (quiescent cells and proliferative cells) | |
| $L_t$ | scout CTL number at $t$ | |
| **Parameters** | **Description** | **Value** |
| $n_0$ | CTL number | variyng |
| $\mu$ | killing rate | 0.0379 |
| $\lambda$ | tumor cell division rate | 0.001 |
| $d$ | tumor cell diameter | 12.5 $\mu m$ |
| $r$ | tumor cell radius | 6.25 $\mu m$ |
| $\boldsymbol{\delta}$ | proliferative part proportion in alive part | 0.14 |
| $\boldsymbol{R_{max}}$ | maximal radius of the tumor mass at $t=0$ | $100\cdot 10^2$ $\mu m$ |
| $\boldsymbol{R_{min}}$ | minimal radius of the tumor mass at $t=0$ | $1\cdot 10^2$ $\mu m$ |
| $R_0$ | tumor mas radius et $t=0$ | $R_0 \in [R_{min}, R_{max}]$ |
| $\boldsymbol{E_0}$ | alive part thickness at $t=0$ | $\min(219, R_0)$ $\mu m$ |
| $N_0$ | necrotic part radius at $t=0$ | $(R_0 - E_0)$ $\mu m$ |
| $\mathbb{E}(\boldsymbol{L_0})$ | mean number of scout CTL at $t=0$ | 1 |
| $\boldsymbol{\nu_{max}}$ | maximal attraction strength | $\frac{1}{100}$ |

Table 3.1: **Variables and parameter values.** Parameters $\mu$, $\lambda$, $d$, $r$, are fixed using experimental measurements and statistical studies developed in Section 1.3 of Chapter 1. Bold parameter values are not revealed in Chapter 1 and need more analysis. They are done in "Choice of parameter values" paragraph.

**Choice of parameter values** The success or not of CTL response against tumor nodule depends on the nodule radius at time $t=0$, namely $R_0 \in [R_{min}, R_{max}]$. Measurements of Figure 1.3 are obtained for a mean diameter nodule at the beginning

equals to $300\mu m$. Thus, we assume $R_{min} = 100\ \mu m$ and $R_{max} = 1\ cm$. There exist nodules with diameter higher than $2cm$, but $R_0$ varying between $100\ \mu m$ to $1\ cm$ gives an overview of the behavior of the interaction between CTL and tumor nodule.

As noted in Remark 2.3.2, the scale of the parameters have to be carefully determined. Note that, a $100\ \mu m$ scale is a good choice. Indeed, this allows numerical analyses with low numbers, it is furthermore in accordance with hypotheses of Theorem 2.3.1 and 2.3.5. Let $\boldsymbol{R_{min}} = 1 \cdot 100\ \mu m$, and $\boldsymbol{R_{max}} = 100 \cdot 100\ \mu m$.

The value of the parameters $\boldsymbol{E_0}$ and $\boldsymbol{\delta}$ are not yet available. Then, they are tuned such that the nodule diameter equation fits experimental results of Figure 1.2 (C).

The parameter $\mathbb{E}(L_0)$ is the mean scout CTL number at time $t = 0$. Formally, recall that $Z_0^i$ denotes the $i^{th}$ CTL position at time $t = 0$, and CTL trajectories are assumed identically distributed and independent. Then,

$$\mathbb{E}(L_0) = n_0 \mathbb{P}_{\mu_0}(Z_t^1 \in [R_0, R_0 + r]).$$

But, for the sake of simplicity we assumed that $\mathbb{E}(\boldsymbol{L_0}) = 1$. In the attraction case, this allows us to see directly the impact of the attraction.

In drifted CTL displacement case, a maximal attraction strength parameter has to be calibrate. Indeed, let $\rho_s$ be the distance between a CTL and the center of the nodule at time $s$, recall that

$$\forall s > 0, \qquad \mathrm{d}\rho_s = \mathrm{d}B_s + \left(\frac{1}{2\rho_s} - \nu_t \rho_s\right) \mathrm{d}s, \tag{3.3}$$

where $\nu_t$ is the attraction strength, depending on $\mathbb{E}(L_t)$, and $\rho_s \geq R_t$. Biological measurements (see Table 1.1) give a mean CTL velocity equals to $8.66\mu m/min$. To counterbalance $\rho_s$, in the drift term, which is higher than $100\ \mu m$, we assume $\boldsymbol{\nu_{max}} = \frac{1}{100}$. The attraction strength then vary between 0 and $\nu_{max}$.

**Choice of non scout CTL distribution** We do not have informations on the non-scout CTL distribution over the time. First, among classical probability distributions, a Gamma distribution is assumed. The advantages of this distribution are that it satisfies Theorem 2.3.1 and 2.3.5 conditions, and the attraction strength can be converted by varying the scale parameter of such distribution.

Using stochastic trajectories for CTL displacements, one of the most natural distribution is a quasi-stationary distribution (QSD). This distribution depends on the trajectories chosen. Then the attraction of CTL, characterized by an Ornstein Uhlenbeck dynamics, is directly converted with this distribution.

**Increasing and decreasing tumor nodule** Note that, at the beginning, the tumor nodule is growing. Since, the time $t = 0$ is assumed to be the first time of the

CTL/nodule interaction. Then, to begin, the first line of equation of $R_t$ has to be considered.

But, if at time $T$ the nodule enters in a decreasing phase, the nodule remains in a decreasing phase. This behavior comes from the hypothesis: a scout CTL remains a scout CTL, which implies the increase of the mean scout CTL number, $\mathbb{E}(L_t)$. And, if the nodule is decreasing, it means that there are already enough CTL on the nodule overcome to control its growth. Then, the necrotic part radius is fixed to $N_T$ the remaining time.

There is two different behaviors, a decreasing tumor: leading to alive tumor cell number equals to 0, or an equilibrium between new tumor cells obtained by division and tumor cells killed by CTL. In both cases, equilibrium or decreasing tumor mass, it is a success of the CTL response against tumor.

The behaviors of tumor mass radius, alive cell number and mean scout CTL number are studied along the time under different conditions. The success of the CTL responses is computed depending on CTL number and on tumor mass radius at time $t = 0$. These analysis are done, alternatively, in self-governing CTL displacements or in drifted CTL displacements. Results are presented in two different sections according to the choice of the non scout CTL distribution. A last section is devoted to a back on the model proposed in this part.

## 3.2 Gamma distribution

In this section we are interested in numerical results on the CTL response against tumor nodule according to a Gamma distribution for non scout CTL at time $t$. Recall that CTL at time $t$ are distributed on $]R_t, \infty[$.

**Gamma distribution** The Gamma distribution is defined on $\mathbb{R}_+$ and characterized by a shape parameter $\alpha > 0$ and a scale parameter $\beta > 0$. Since, $\rho_t > R_t$, we consider a translation of $-R_t$ of this density. Then

$$g_t(x; \alpha, \beta_t) = \frac{(x - R_t)^{\alpha-1}}{\Gamma(\alpha)\beta_t^\alpha} e^{-\frac{(x-R_t)}{\beta_t}} \mathbb{1}_{x \geq R_t}.$$

Note that, if $\alpha > 1$, there exists $\varepsilon > 0$ such that $g_t(., \alpha, \beta)$ is increasing on $[0, \varepsilon]$. This distribution then satisfies hypotheses of Theorems 2.3.1 and 2.3.5 and we can compute $\mathbb{E}(L_t)$. Let us fix $\alpha = 2$.

The scale parameter $\beta$ describes the spreading out of a distribution, *i.e.* larger $\beta$ is, more spread out the distribution is. In self-governing, there is no reason to have a lot of CTL close to the nodule, and the parameter $\beta$ is then large and fixed at 10000.

In the drifted displacement, the scale parameter translates the attraction effect. Indeed, higher the scout CTL number is, lower the scale parameter $\beta$ has to be, and in this way, more non scout CTL are closed to the nodule. Then $\beta_t$, depending on $\mathbb{E}(L_t)$, is assumed $\beta_t = \max(1.1, 10000 - 100 * \mathbb{E}(L_t))$.

We have an upper and lower bound of $\partial_t \mathbb{E}(L_t)$, then to compute it, note that, for $x \geq R_t$,

$$\begin{aligned} g'(x;\alpha,\beta_t) &= \frac{(\alpha-1)(x-R_t)^{\alpha-2}}{\Gamma(\alpha)\beta_t^\alpha} e^{-\frac{(x-R_t)}{\beta_t}} - \frac{(x-R_t)^{\alpha-1}}{\beta_t\Gamma(\alpha)\beta_t^\alpha} e^{-\frac{(x-R_t)}{\beta_t}} \\ &= \frac{(x-R_t)^{\alpha-2}}{\Gamma(\alpha)\beta_t^\alpha} e^{-\frac{(x-R_t)}{\beta_t}} \left(\alpha - 1 - \frac{x-R_t}{\beta_t}\right) \\ &= \frac{e^{-\frac{(x-R_t)}{\beta_t}}}{\Gamma(\alpha)\beta_t^\alpha} \left(1 - \frac{x-R_t}{\beta_t}\right). \end{aligned}$$

The last inequality is obtained by replacing $\alpha$ by 2. Then $g'(R_t, 2, \beta_t) = \frac{1}{\beta_t^2} > 0$, and $g'\left(R_t - \frac{1}{2\mathbb{E}(L_t)R_t}, 2, \beta_t\right) = \frac{e^{-\frac{1}{2\mathbb{E}(L_T)R_t\beta_t}}}{\beta_t^2}\left(1 + \frac{1}{2\mathbb{E}(L_t)R_t\beta_t}\right) > 0$. As $\frac{1}{2\mathbb{E}(L_T)R_t\beta_t} \leq \frac{1}{2}$, we assume $g'(R_t, 2, \beta_t) = g'\left(R_t - \frac{1}{2\mathbb{E}(L_t)R_t}, 2, \beta_t\right)$, that implies the following equality:

$$\partial_t \mathbb{E}(L_t) = (n_0 + \mathbb{E}(L_t))\frac{1}{2\beta_t^2}.$$

Numerical simulations are performed to decipher the behaviors of solutions of Systems (3.1) and (3.2), varying the nodule radius at the beginning: $R_0 \in [R_{min}, R_{max}]$ and the CTL number $n_0$.

### 3.2.1 Numerical results under self-governing CTL displacements

This subsection is devoted to illustrate solutions of System (3.1).

First of all, we give numerical approximations of nodule radius, alive cells number and scout CTL number along the time (around 70 days), for two fixed radius: $R_0 = 1cm$ and $R_0 = 1mm$, and according to fixed CTL numbers. Graphs obtained inform on the behaviors of these data. Secondly, in order to determine CTL number leading to nodule eradication, we compute the decrease or not of the nodule for $R_0$ varying between $R_{min}$ and $R_{max}$ and for a range of $n_0$.

A quick analyze of Figure 3.1 indicates an exponential growth of the nodule and a linear arrival of CTL on the nodule border. Comparing Figures 3.1 and 3.2, one can remark, higher the nodule radius is, higher the CTL number leading to eradication is.

In Figure 3.3 (A), notice that, if a CTL number is enough to eradicate the tumor mass with radius $R_0$, then, an higher CTL number leads also to nodule eradication. This is a characterization of a phase transition, in CTL number leading to nodule

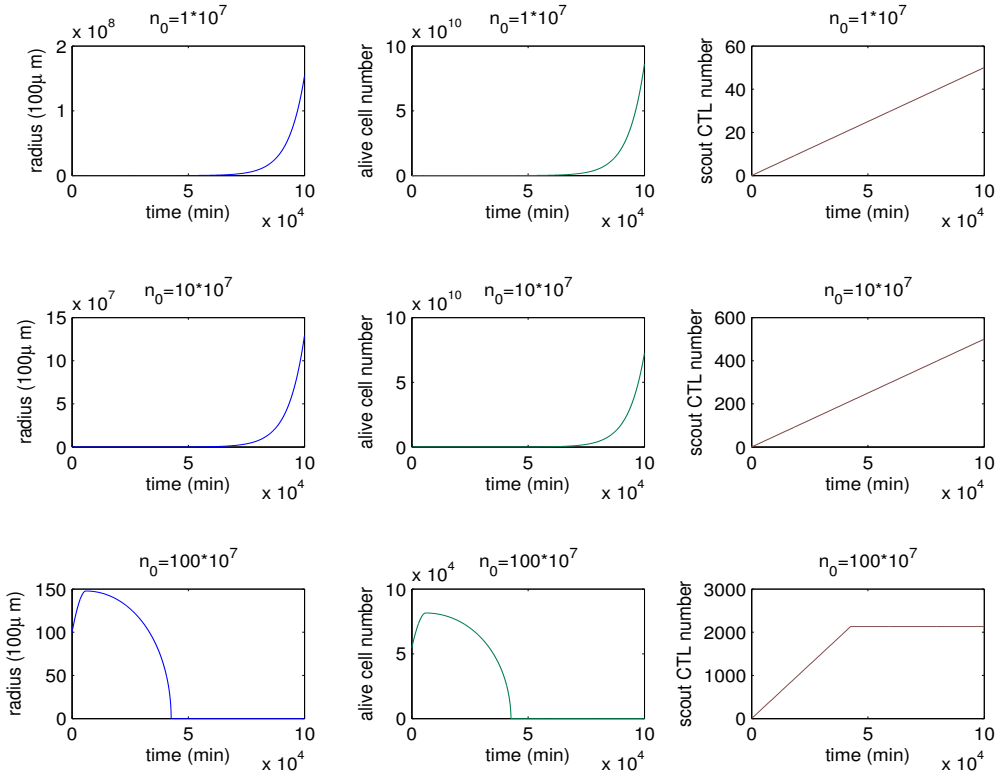



Figure 3.1: **Nodule radius, alive cell number and scout CTL number along the time, beginning at** $R_0 = 1\ cm$**.** There is no attraction of CTL. In first line, there are $n_0 = 10^7$ CTL, around 5 alive tumor cells against 1000 CTL; in second line, there are $n_0 = 10^8$ CTL, around 5 alive tumor cells against 10000 CTL; in the third line, there are $n_0 = 10^9$ CTL, around 5 alive tumor cells against 100000 CTL.

eradication.

Hence, a graph of minimal CTL number leading to eradication of nodule is plotted, see Figure 3.3 (B). One notes the linear dependance between the CTL number $n_0$ and the nodule radius $R_0$. The linear function fitting the empirical data is:

$$y = 5827.3 \cdot 10^3 x - 7190.9 \cdot 10^3. \tag{3.4}$$

Thanks to this equation, a prediction of the minimal CTL number leading to nodule eradication for a tumor mass with a radius higher than $1cm$ can be obtained.

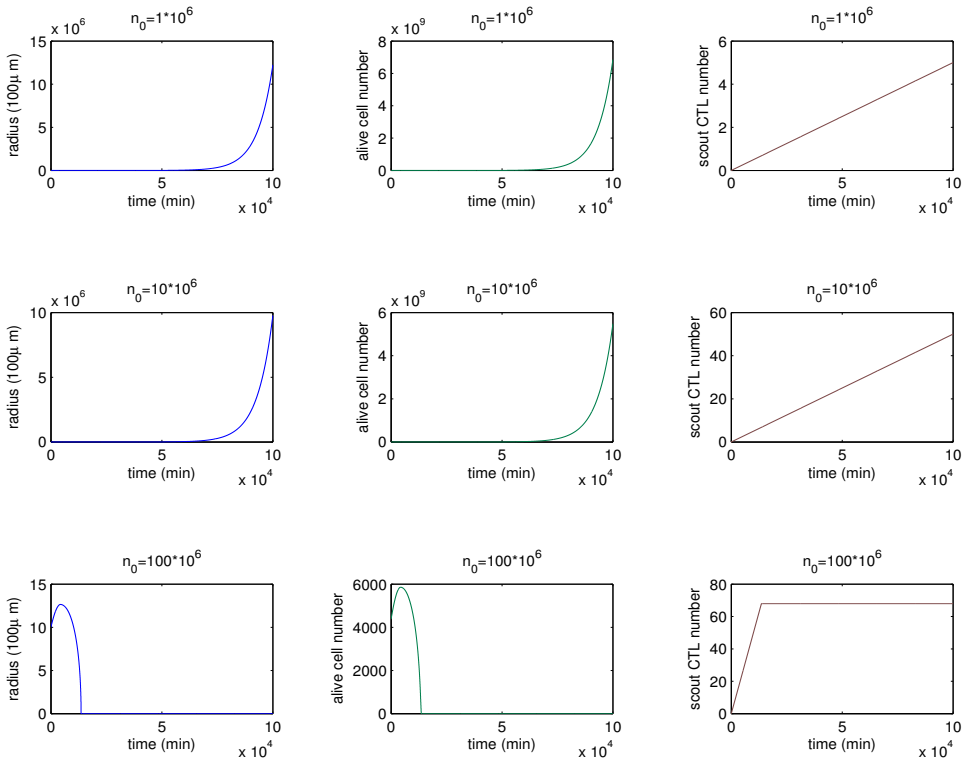
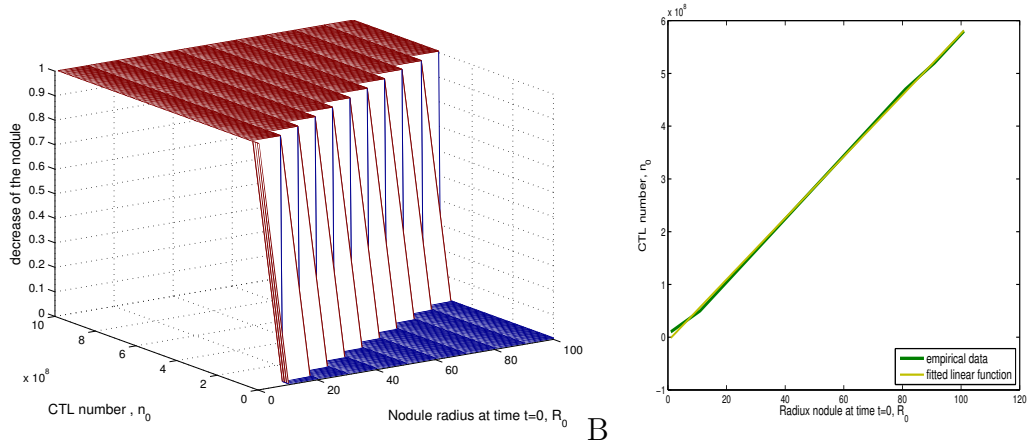



Figure 3.2: **Nodule radius, alive cell number and scout CTL number along the time, begining at $R_0 = 1\ mm$.** There is no attraction of CTL. In first line there is 4.5 tumor cells against 1000 CTL.




A B

Figure 3.3: **Decrease or not of the nodule according to $n_0$ and $R_0$.** There is no attraction of CTL. (A) Decrease (1) or not (0) of the nodule according to the CTL number and the nodule radius at time $t = 0$. (B) Minimal CTL number leading to nodule eradication, according to the nodule radius.

### 3.2.2 Numerical results under biased CTL displacements

We perform the same experimental measurements than in the previous subsection, in order to approximate numerically the solutions of System (3.2). It means that, we consider a simulated immunotherapy consisting in CTL attraction toward the nodule.

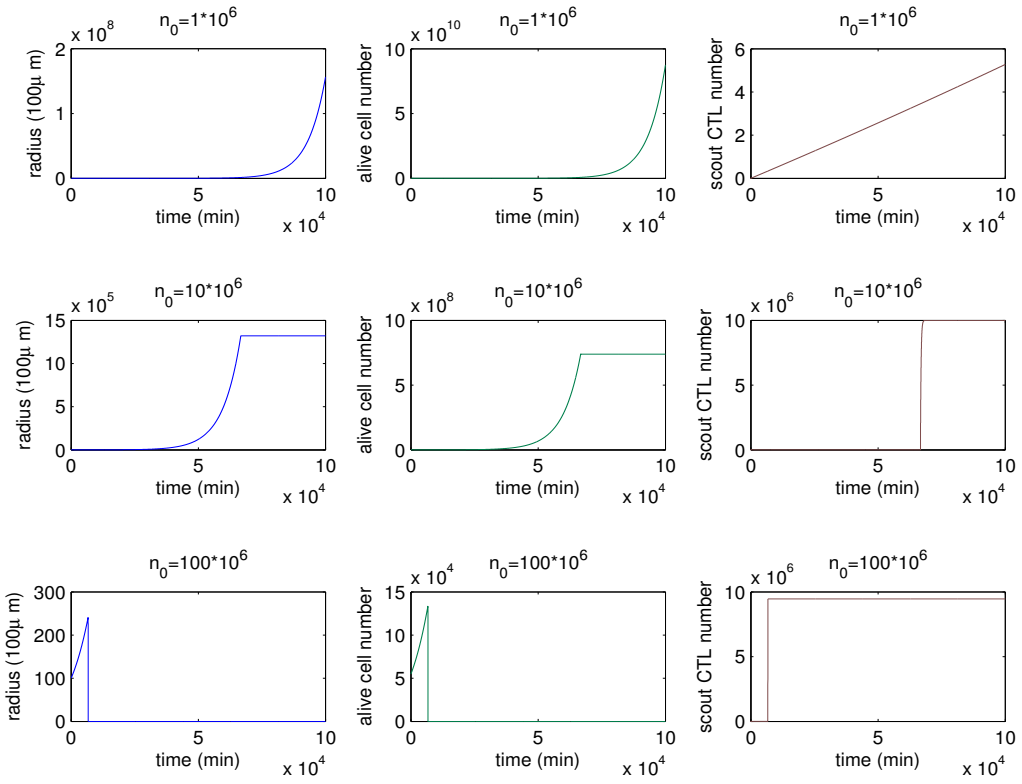



Figure 3.4: **Nodule radius, alive cell number and scout CTL number along the time, beginning at** $R_0 = 1cm$**.** There is attraction of CTL toward the nodule. In first line, there are $10^6$ CTL, around 5 alive tumor cells against 100 CTL; in second line, there are $10^7$ CTL, around 5 alive tumor cells against 1000 CTL; in third line, there are $n_0 = 10^9$, around 5 alive tumor cells against 10000 CTL.

One can note, under a simulated immunotherapy, the CTL number leading to the eradication is smaller than without immunotherapy, see Figure 3.4 and Figure 3.1. Indeed, with CTL attraction, for a radius of $1cm$ at the beginning, $10 \cdot 10^6$ CTL leads to a control of the nodule growth, and $10^7$ CTL leads to nodule eradication. While without attraction, $10^8$ CTL do not lead to decrease of the nodule, neither equilibrium.

Under simulated therapy, notice that an exponential CTL arrival on the nodule border in victory cases; that explains certainly the result. The same behaviors is encountered for a nodule radius of $R_0 = 1mm$, see Figure 3.5 and 3.2.

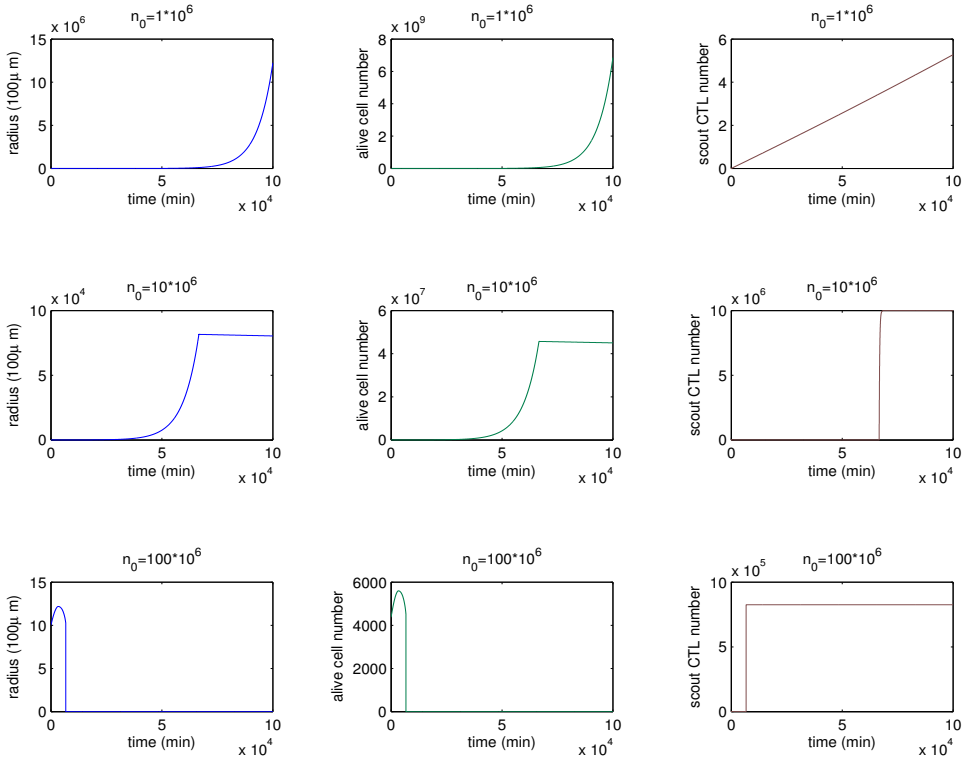



Figure 3.5: **Nodule radius, alive cell number and scout CTL number along the time, beginning at $R_0 = 1mm$.** There is attraction of CTL towards the nodule. In first line there are $10^6$ CTL,around 4.5 tumor cells against 1000 CTL.

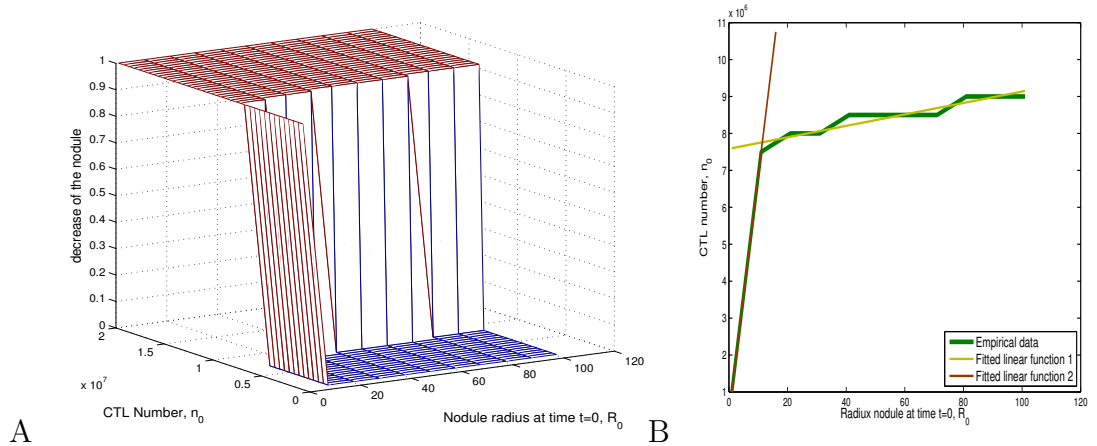



Figure 3.6: **Decrease or not of the nodule according to $n_0$ and $R_0$, under simulated immunotherapy.** (A) Decrease (1) or not (0) of the nodule according to the CTL number and the nodule radius at time $t = 0$. (B) Minimal CTL number leading to nodule eradication, according to the nodule radius.

As without CTL attraction, with attraction of CTL towards the nodule, a phase transition un CTL number leading to nodule eradication can be remarked, see Figure 3.6 (A). But, under simulated immunotherapy, the minimal CTL number leading to nodule eradication, is clearly not linear in the nodule radius $R_0$, see Figure 3.6 (B).

However, two different linear behaviors can be noted. For high nodule radius: $R_0 \geq 1.1mm$, the fitted equation is

$$y_1 = 15.5 \cdot 10^3 x + 7584.5 \cdot 10^3,$$

see yellow line in Figure 3.6 (B). And for small radius: $R_0 \in [100, 1100[\mu m$, the fitted equation is

$$y_2 = 650 \cdot 10^3 * x + 350 \cdot 10^3,$$

see brown line in Figure 3.6 (B).

Hence, we can consider a phase transition in minimal CTL number leading to nodule eradication, according to the size of the tumor. If the nodule radius is higher than $1.1mm$, the linear equation, giving the minimal CTL number leading to eradication, has a slope of order $10^4$. But, for small nodule radius, a linear equation with a slope of order $10^5$ gives this minimal CTL number.

Then using the equation $y_1$, prediction of CTL number leading to nodule eradication with a radius higher than $R_0 = 1cm$ can be done.

### 3.2.3 Conclusions

**Results** A phase transition in CTL number leading to nodule eradication, with or without CTL attraction to the nodule, is highlighted by this study. And phase transition in minimal CTL number leading to nodule eradication, according to nodule size, is also revealed.

When $R_0 \geq 1.1$, one can note different slopes of the linear functions fitting the graph of minimal CTL number leading to nodule eradication, if there is a simulated immunotherapy or not. Namely, the slope of this function is lower in CTL attrition case: of order $10^4$ of $R_0$, than without CTL attraction: it is of order $10^6$ of $R_0$.

Then, an immunotherapy consisting in directing CTL toward the nodule, would be a good treatment. Since most of CTL present in the micro-environment of the nodule are requisitioned to kill tumor cells, and less CTL are needed in the micro-environment of the tumor mass to eradicate it. That could imply a reduction of second effects of a treatment.

**Discussion on Gamma distribution** If CTL are attracted towards the nodule, a more detailed study could give a better estimation of the nodule radius, implying the phase transition in the equations that describe the minimal CTL number leading to

nodule eradication. But we do not develop this study because there is not link between Gamma distribution and the trajectory proposed to built Equations (3.1) and (3.2). That is why, we can choose $\nu_t = \max(1.1, 1000 - 100 * \mathbb{E}(L_t))$, even if, it is contrary to Remark 2.3.7.

However, this study gives a first idea of the behaviors of such interaction.

In the following subsection, we develop numerical results for $\mu_t$ a Quasi-Stationary Distribution associated to the Bessel process and the radial O.U. process.

## 3.3 Quasi-stationary distribution

In the model established in the previous chapter, two stochastic equations describe CTL dynamics. Alternatively, a Brownian motion for self-governing CTL displacements, and Ornstein-Uhlenbeck (O.U.) process for CTL displacements under simulated immunotherapy. Let $\rho$ be the distance between a CTL and the nodule center, namely, according cases, a Bessel process or a radial O.U. process defined in Section 2.3. To compute the number of scout CTL, recall the hitting time to $R_t$ (the nodule radius) for the process $\rho$:

$$T_{R_t} = \inf\{s > t,\ \rho_s \leq R_t\}.$$

The main difficulty of the model proposed in Chapter 2, is the computation of the limit of the probability of $T_{R_t}$:

$$\lim_{h \to 0} \frac{\mathbb{P}_{\mu_t}(T_{R_t} \leq h)}{h}, \tag{3.5}$$

where $\mu_t$ is the distribution of non scout CTL at time $t$. More precisely, $\mu_t$ is the distribution of particles at time $t$ conditioned not to be absorbed at this time.

The distribution $\mu_t$ is unknown. But, considering a sufficiently long time and under existence condition, we propose a Quasi-Stationay Distribution (QSD) for $\mu_t$. For a general view on QSD we refer to the survey [MV12] or the monograph [CMSM13]. Definitions and first properties are recalled in the following subsection.

### 3.3.1 Definition and first properties

Let $\rho$ a Markov process and $T_{R_t}$ its first time to reach $R_t$. We say that the process $\rho$ reaches $R_t$ in finite time, if it satisfies:

$$\forall x \in ]R_t, \infty[, \quad \mathbb{P}_x(T_{R_t} < \infty) = 1. \tag{3.6}$$

We are interested in the distribution of the particles, which trajectory is the process $\rho$, conditioned not to be absorbed at time $t$. Let us define the Quasi-Stationarity

Distribution.

**Definition 3.3.1.** *Let $t > 0$, $R_t > 0$ and $\mu_t$ be a probability distribution on $]R_t, \infty[$. The distribution $\mu_t$ is said to be a* **quasi-stationary distribution (QSD)** *if, for all $s \geq 0$ and any measurable set $A \subset ]R_t, \infty[$,*

$$\mu_t(A) = \mathbb{P}_{\mu_t}(\rho_s \in A | T_{R_t} > s).$$

A first remark is that, starting from a QSD, the process $T_{R_t}$ is exponentially distributed.

**Theorem 3.3.1.** *Let us consider a Markov process $\rho$ with absorbing point $R_t$ satisfying* (3.6)*. If $\mu_t$ is a QSD then there exists a positive real number $\theta_{\mu_t}$ depending on the QSD, such that*

$$\mathbb{P}_{\mu_t}(T_{R_t} > s) = e^{-\theta_{\mu_t} s}.$$

We refer to [CMSM13, theorem 2.2 p19] for a proof of this theorem. From a spectral theory, $\theta_{\mu_t}$ is the top of the spectrum of $-\mathcal{L}$, where the operator $\mathcal{L}$ is the infinitesimal generator associated with the process $\rho$.

Note that, if there exists QSD for a Bessel process or a radial O.U. process, by the previous theorem, one gets

$$\begin{aligned} \lim_{h \to 0} \frac{\mathbb{P}_{\mu_t}(T_{R_t} \leq h)}{h} &= \lim_{h \to 0} \frac{1 - \mathbb{P}_{\mu_t}(T_{R_t} > h)}{h} \\ &= \lim_{h \to 0} \frac{1 - e^{-\theta_{\mu_t} h}}{h} \\ &= \theta_{\mu_t}. \end{aligned} \tag{3.7}$$

In the subsection following, we set conditions for existence of QSD, alternatively, for the Bessel process and a radial O.U. process. To our knowledge, there is no result on the eigenvalue $\theta_{\mu_t}$, for these both processes. Then, in a following subsection, we use a Fleming-Viot type algorithm to compute numerically this value.

### 3.3.2 Existence of the quasi-stationary distribution

For $\rho$ a Markov process, let $-\alpha$ be its drift, $\theta_\alpha$ be the lowest eigenvalue of the infinitesimal generator associated with the process of drift $-\alpha$, and

$$Q(z) = 2 \int_0^z \alpha(x) \, \mathrm{d}x.$$

Using [CMSM13, theorem 6.26], one can state the theorem bellow.

**Theorem 3.3.2.** *Assume $\rho$ satisfies* (3.6) *and*

$$\begin{cases} \int_{R_t}^{\infty} e^{Q(z)} \int_{R_t}^{z} e^{-Q(x)} \,\mathrm{d}x \,\mathrm{d}z = \infty \\ \int_{R_t}^{\infty} e^{-Q(z)} \int_{R_t}^{z} e^{Q(x)} \,\mathrm{d}x \,\mathrm{d}z = \infty \end{cases} . \tag{3.8}$$

*If $\theta_\alpha > 0$, then the process $\rho$ admits a QSD.*

Intuitively a process admits a QSD if it comes back from infinity quickly. Corollary 6.32 of [CMSM13], compare two lowest eigenvalue of infinitesimal generator of two different processes. We recall it here.

**Lemma 3.3.3.** *If $\rho$ and $\tilde{\rho}$ satisfy* (3.6) *and* (3.8)*, and if $\alpha$ and $\tilde{\alpha}$ are $\mathcal{C}^1$ function such that $\alpha \geq \tilde{\alpha}$, then*

$$\theta_\alpha \geq \theta_{\tilde{\alpha}}.$$

*In particular, let $c$ a constant, if $\alpha(x) \geq c \geq 0$, for all $x$, then $\theta_\alpha \geq \frac{c^2}{2}$.*

**Remark 3.3.4.** *In [CMSM13], the study is done for process absorbed at 0, but in our study the process are absorbed at $R_t$. But, considering $Y_s^{(i)} = \rho_s^{(i)} - R_t$, we obtain same results for our case.*

### QSD existence or not for the Bessel process

Let $\rho^{(1)}$ be a Bessel process satisfying the differential Equation (2.8), $\mathcal{L}_1$ the infinitesimal generator of the process $\rho^{(1)}$, and $-\alpha_1$ its drift, namely

$$\alpha_1(x) = -\frac{1}{2x},$$

and

$$Q_1(z) = 2\int_0^z \alpha_1(x) \,\mathrm{d}x = \ln R_t - \ln z.$$

Let $T_{R_t}^{(1)}$ the hitting time to reach $R_t$ for the Bessel process.

Thanks to the recurrence property of the Brownian motion, Bessel process reaches $R_t$ in finite time, then Equation (3.6) is satisfied. Equations (3.8) are also verified. Indeed,

$$\begin{aligned} \int_{R_t}^{\infty} e^{Q_1(z)} \int_{R_t}^{z} e^{-Q_1(x)} \,\mathrm{d}x \,\mathrm{d}z &= \int_{R_t}^{\infty} e^{\ln R_t - \ln z} \int_{R_t}^{z} e^{-\ln R_t + \ln x} \,\mathrm{d}x \,\mathrm{d}z \\ &= \int_{R_t}^{\infty} \frac{1}{z} \int_{R_t}^{z} x \,\mathrm{d}x \,\mathrm{d}z \\ &= \int_{R_t}^{\infty} \frac{1}{z} \frac{z^2 - R_t^2}{2} \,\mathrm{d}z = \infty, \end{aligned}$$

and

$$\begin{aligned}\int_{R_t}^{\infty} e^{-Q_1(z)} \int_{R_t}^{z} e^{Q_1(x)} \,\mathrm{d}x \,\mathrm{d}z &= \int_{R_t}^{\infty} e^{\ln z} \int_{R_t}^{z} e^{-\ln x} \,\mathrm{d}x \,\mathrm{d}z \\ &= \int_{R_t}^{\infty} z \int_{R_t}^{z} \frac{1}{x} \,\mathrm{d}x \,\mathrm{d}z \\ &= \int_{R_t}^{\infty} \frac{\ln z - \ln R_t}{z} \,\mathrm{d}z \\ &= \int_{R_t}^{eR_t} \frac{\ln(z/R_t)}{z} \,\mathrm{d}z + \int_{eR_t} \frac{1}{z} \,\mathrm{d}z = \infty.\end{aligned}$$

As $0 \geq \alpha_1(x)$, using Lemma 3.3.3, $\theta_{\alpha_1} \leq 0$. By Theorem 3.3.2, the Bessel process on $]R_t, \infty[$ does not admit a QSD.

To bypass this problem, the idea is to suppose the process defined in bounded set. Notice $\bar{R} \in \mathbb{R}$ such that $\bar{R} >> R_0$. Assume $\rho$ a Bessel process with normal reflection at $\bar{R}$, and absorbed at $R_t$, hence, for all $s$, $\rho_s \in ]R_t, \bar{R}]$. This process, viewed as a Markov process, is defined by its generator $\mathcal{L}_1$. Let $\varphi \in D(\mathcal{L}_1)$, where $D$ is the domain of the generator $\mathcal{L}_1$, satisfies the differential equation following, for all $x \in ]R_t, \bar{R}]$,

$$\begin{cases} \mathcal{L}_1\varphi(x) = \frac{1}{2}\Delta\varphi(x) + \frac{1}{2x}\partial\varphi(x) \\ \varphi(R_t) = 0 \\ \partial\varphi(\bar{R}) = 0 \end{cases},$$

with Dirichlet boundary condition on $R_t$, the absorbing point, and Neumann boundary condition on $\bar{R}$, the reflecting point.

We think that existence property can be shown using the same way to prove [CMSM13, theorem 6.4], which give QSD existence for killed process at both $R_t$ and $\bar{R}$, or adapting [Cat+09] to our simpler setting.

**Remark 3.3.5.** *This model is consistent with biology, indeed, we can assume that a particle leaving the nodule microenvironement is replaced by a new entering particle. In addition, it is also consistent with our numerical simulations, since they are performed on compact domaine.*

### Existence of the QSD for the radial Ornstein-Uhlenbeck process?

Let $\rho^{(2)}$ be a radial O.U. process satisfying the differential Equation (3.3). Let $\mathcal{L}_2$ be the infinitesimal generator of the process $\rho^{(2)}$, $-\alpha_2$ its drift, namely

$$\alpha_2(x) = -\frac{1}{2x} + \nu_t x.$$

And

$$Q_2(z) = 2\int_0^z \alpha_2(x) \,\mathrm{d}x = \ln R_t - \ln z + \nu_t z^2 - \nu_t R_t^2.$$

We recall that $\nu_t$ and $R_t$ are constant, here we consider the process $(\rho_s)_{s>0}$. Let $T^{(2)}_{R_t}$ be the hitting time to reach $R_t$ for the radial O.U. process.

Using the coupling with the same Brownian motion between an O.U. process and a Brownian motion, the hitting time to $R_t$ for the O.U. process is almost surely lower than one of the Brownian motion. Then, the hitting time for an O.U. process satisfies Equation (3.6). Then, this equation is also satisfied for $T^{(2)}_{R_t}$. And $\rho^{(2)}$ satisfies also Equations(3.8). Indeed,

$$\begin{aligned}\int_{R_t}^{\infty} e^{Q_2(z)} \int_{R_t}^{z} e^{-Q_2(x)}\,\mathrm{d}x\,\mathrm{d}z &= \int_{R_t}^{\infty} e^{-\ln z+\nu_t z^2} \int_{R_t}^{z} e^{\ln x-\nu_t x^2}\,\mathrm{d}x\,\mathrm{d}z \\ &= \int_{R_t}^{\infty} \frac{e^{\nu_t z^2}}{z} \int_{R_t}^{z} x e^{-\nu_t x^2}\,\mathrm{d}x\,\mathrm{d}z \\ &= \frac{1}{2\nu_t} \int_{R_t}^{\infty} \frac{e^{\nu_t (z^2-R_t^2)}-1}{z}\,\mathrm{d}z \geq \infty,\end{aligned}$$

and

$$\begin{aligned}\int_{R_t}^{\infty} e^{-Q_2(z)} \int_{R_t}^{z} e^{Q_2(x)}\,\mathrm{d}x\,\mathrm{d}z &= \int_{R_t}^{\infty} e^{\ln z-\nu_t z^2} \int_{R_t}^{z} e^{-\ln x+\nu_t x^2}\,\mathrm{d}x\,\mathrm{d}z \\ &= \int_{R_t}^{\infty} \frac{e^{\nu_t x^2}}{x} \int_{x}^{\infty} z e^{-\nu_t z^2}\,\mathrm{d}z\,\mathrm{d}x \\ &= \frac{1}{2L} \int_{R_t}^{\infty} \frac{e^{\nu_t x^2}}{x} e^{-\nu_t x^2}\,\mathrm{d}x \\ &= \frac{1}{2L} \int_{R_t}^{\infty} \frac{1}{x}\,\mathrm{d}x = \infty.\end{aligned}$$

Note that $\alpha_2(x) \geq \frac{-1}{2R_t} + \nu_t R_t$, and $\alpha(x) > 0$ is equivalent to $R_t > \frac{1}{\sqrt{2\nu_t}}$. For $R_t > \frac{1}{\sqrt{2\nu_t}}$, using Lemma 3.3.3, $\theta_{\alpha_2} > 0$ and by Theorem 3.3.2, there exists QSD for the process. But, if $R_t \leq \frac{1}{\sqrt{2\nu_t}}$, we cannot assure the result. Then, as for the Bessel process, we consider the Radial O.U. process absorbed in $R_t$, and reflected in $\bar{R} >> R_t$. That assure the existence of the QSD for the Radial O.U. process.

### 3.3.3 Fleming-Viot type algorithm and QSD

Under QSD existence conditions the two processes: the Bessel process and the radial O.U. process, this section is devoted to the computation of $\theta_{\mu_t}$, the lowest eigenvalue of these two processes, where $\mu_t$ denote the QSD. Until now, there are no theoretical results on these eigenvalues. Then, we will approximate this value using numerical simulations. To this end, a Fleming-Viot type particle system is frequently used.

**Fleming-Viot algorithm** The main difficulty in approximating the QSD is that the probability of the event "the process remains not killed a time $s$" vanish when $s$ goes to infinity. Indeed, this is a rare event simulation problem and, in particular, naive

Monte-Carlo methods are not well-suited. Then to overcome this issue a Fleming-Viot algorithm, introduced by [Bur+96], with rebirth is used. It consists in a modification of a Monte-Carlo algorithm, where killed particles are reintroduced through a rebirth in $F$, where $F$ is the state space of the particle that are not killed. This leads to keep a constant population size of particles.

[Vil13] gives only one assumption leading to convergence of the Fleming-Viot type particle system to the QSD. In [Vil13, theorem 1] postulates a non explosion of the number of rebirths in finite time almost surely, and a survival probability of the original process remains positive in finite time. A speed of convergence and criterion for non explosion are also available in this article.

**Non explosion of reflected Process** In our setting, the Bessel process and radial O.U. process evolve in bounded space, reflected on $\bar{R}$ and absorbed on $R_t$. Thanks to [BBF12, Theorem 5.4 ], there is no explosion of the number of rebirths.

**Algorithm** Let $N \in \mathbb{N}^*$, $N > 2$ be the number of particles, for $i \in \{1, ..., N\}$, $Z_t^i$ be the position of the particle at time $t$, this particle follows a radial O.U. dynamics or a Bessel Process. Let $\mu_0$ be the distribution of the particle at time $t$.

1. The $N$ particles are positioned according to a distribution $\mu_0$.
2. Each particle evolves independently, following the dynamic chosen. Until one of them is absorbed. Let $\tau_1$ be the time of absorption time.
3. The absorbed particle jumps instantaneously onto the position of the $N-1$ remaining particles.
4. Go to step 2., and denote $\tau_i$ the $i$ th absorbing time.

**Numerical results for the first eigenvalue** Now, knowing QSD existence conditions and convergence of the Fleming-Viot type particles to the QSD, for the two processes of our interest, our aim is to compute Limit (3.5). Note that, Event $\{T_{R_t} \leq h\}$ is a rare event. However, the complementary of this event is not rare. Using (3.7), with $N$, a hight number of particle and by LLN, (3.5) can be approximated by

$$\hat{\theta}_{\mu_t} = -\frac{1}{N} \sum_{i=1}^{N} \frac{\ln(N_{alive})}{N},$$

where $N_{alive}$ is the number of non absorbed particles in our simulations.

The value of the limit in (3.5) is estimated for different range of absorbing point and attraction strength, and presented in Figure 3.7.

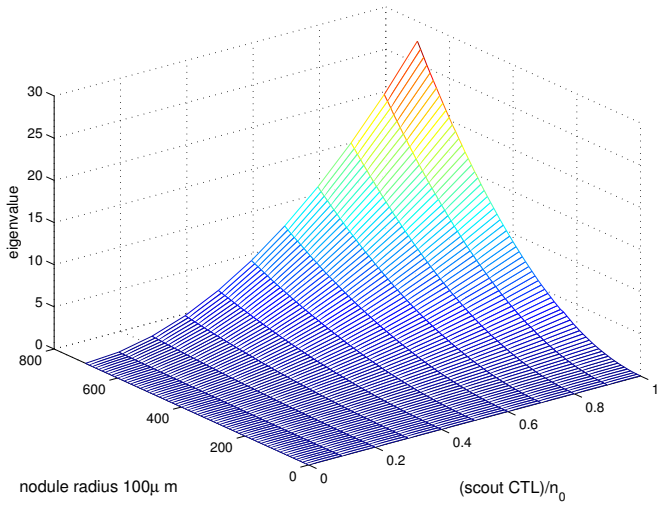



Figure 3.7: **Lowest eigenvalue associated to the radial O.U. process and the Bessel process.** The absorbing point is varying between 1 to 701, and the attraction strength is varying between 0 to 1.

### 3.3.4 Numerical results under self-governing CTL displacements

As in the previous section, we perform numerical simulations to decipher the behaviors of solution of Equations (3.1).

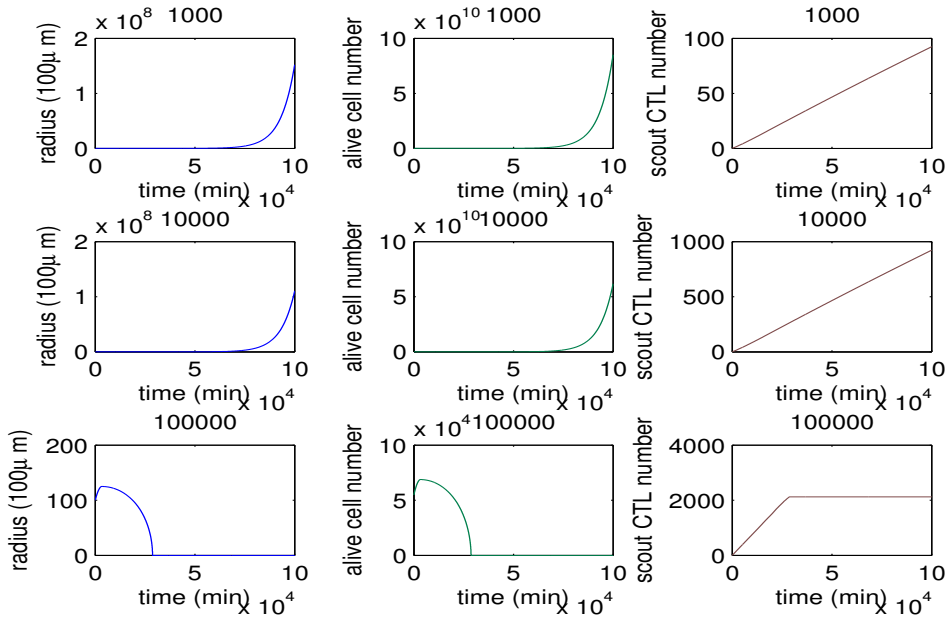



Figure 3.8: **Nodule radius, alive cell number and scout CTL number along the time, beginning at** $R_0 = 1\ cm$**.** There is no attraction of CTL. In first line, there are around 4.5 alive tumor cells against 1 CTL.

First, the behaviors of the solutions of 3.1 are performed for $R_0 = 1cm$ and $R_0 = 1mm$, and for different CTL numbers, $n_0$. Secondly the decrease or not of the nodule is computed for $R_0 \in [R_{min}, R_{max}]$ and different values of $n_0$.

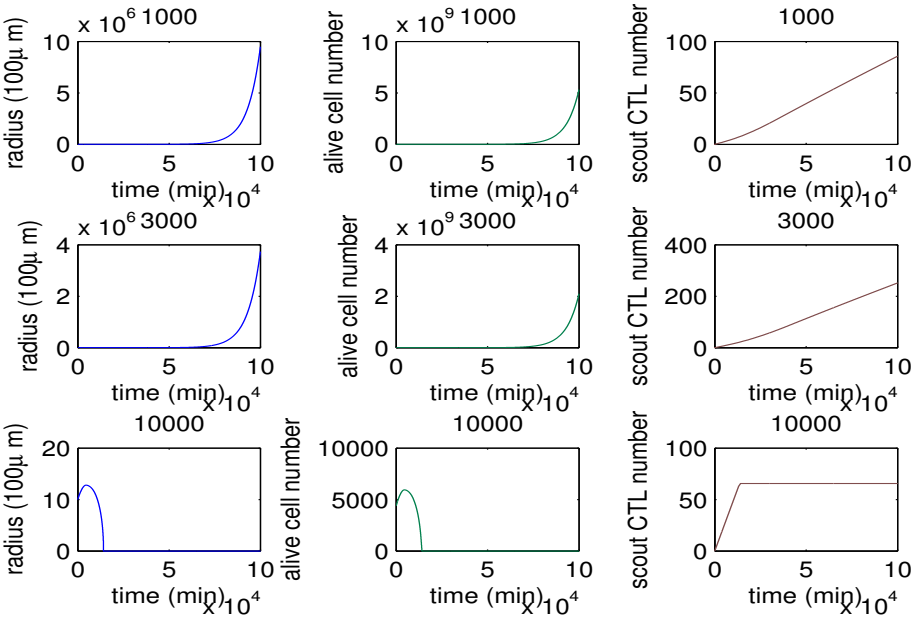
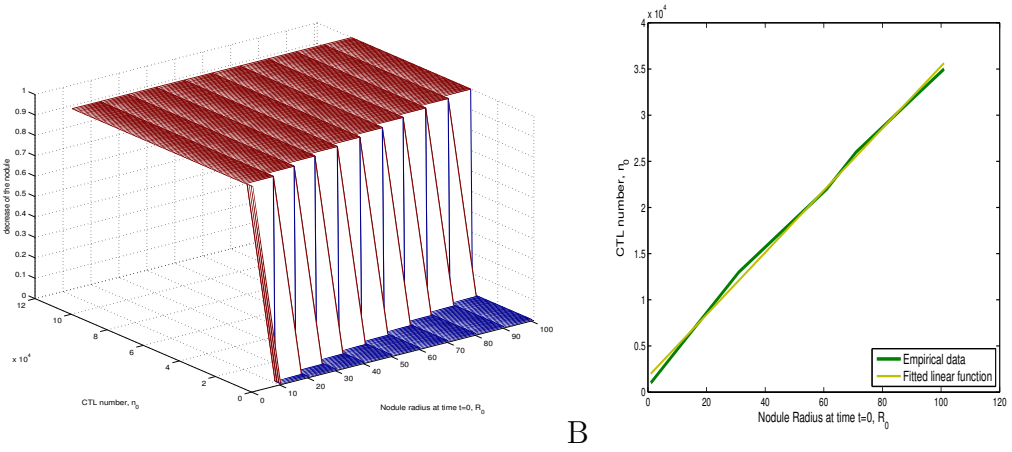



Figure 3.9: **Nodule radius, alive cell number and scout CTL number along the time, beginning at** $R_0 = 1\ mm$**.** There is no attraction of CTL. In first line, there are 4.5 tumor cells against 1 CTL.

As for a Gamma distribution, an exponential growth of the nodule, and a linear arrival of CTL on the nodule border is observed for $R_0 = 1cm$ as for $R_0 = 1mm$, see Figure 3.8 and 3.9; and also, higher the nodule radius is, higher the CTL number leading to tumor eradication is.

In Figure 3.10 (A), we observe, also, a phase transition in CTL number leading to tumor eradication; and a linear behaviors of the minimal CTL number leading to nodule eradication according to radius nodule is noted too (B). The fitted equation of the graphs of this number is the following:

$$y = 336.4x + 1663.6.$$




A B

Figure 3.10: **Decrease or not of the nodule according to $n_0$ and $R_0$.** There is no attraction of CTL. (A) Decrease (1) or not (0) of the nodule according to the CTL number and the nodule radius at time $t = 0$. (B) Minimal CTL number leading to nodule eradication, according to the nodule radius.

### 3.3.5 Numerical results under biased CTL displacements

In this subsection, solutions of System (3.2) are illustrated, it means that, we assume that CTL are attracted toward the nodule.

In Equations (3.2), only an upper and lower bound of $\partial_t \mathbb{E}(L_t)$ is done. But, for a QSD distribution, unsing Equation (3.7) we obtain $\frac{f_t'(R_t,\mathbb{E}(L_t))}{2} = \theta_{\mu_t}$, where $\theta_{\mu_t}$ is the lowest eigenvalue of the generator associated with the radial O.U. process with a drift equal to the proportion of scout CTL number over $n_0$ and absorbed in $R_0$. Since, $\frac{1}{2R_t\mathbb{E}(L_t)} \leq \frac{1}{2}$, we assume

$$\frac{f_t'(R_t,\mathbb{E}(L_t))}{2} = \frac{f_t'\left(R_t - \frac{1}{2R_t\mathbb{E}(L_t)},\mathbb{E}(L_t)\right)}{2},$$

and $\partial_t \mathbb{E}(L_t) = (n_0 - \mathbb{E}(L_t)\theta_{\mu_t}$.

As in Gamma distribution case, we note a reduction of CTL number leading to nodule eradication under simulated immunotherapy, see Figure 3.11 and 3.8. And, one notes the exponential arrival of CTL on the nodule border, in CTL response success.

As for a Gamma distribution, for a QSD distribution, a phase transition is noted in CTL number leading to nodule eradication. And also, a phase transition is observed in the minimal CTL number leading to nodule eradication, according to the nodule radius at the beginning: $R_0$. Indeed two linear equations describe this minimal CTL number: for $R_0 \geq 1.1mm$

$$y_1 = 3.0303x + 370.3030,$$

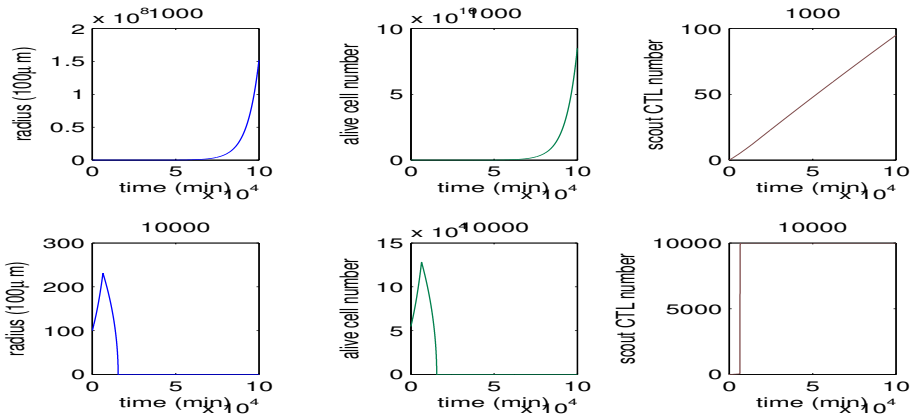
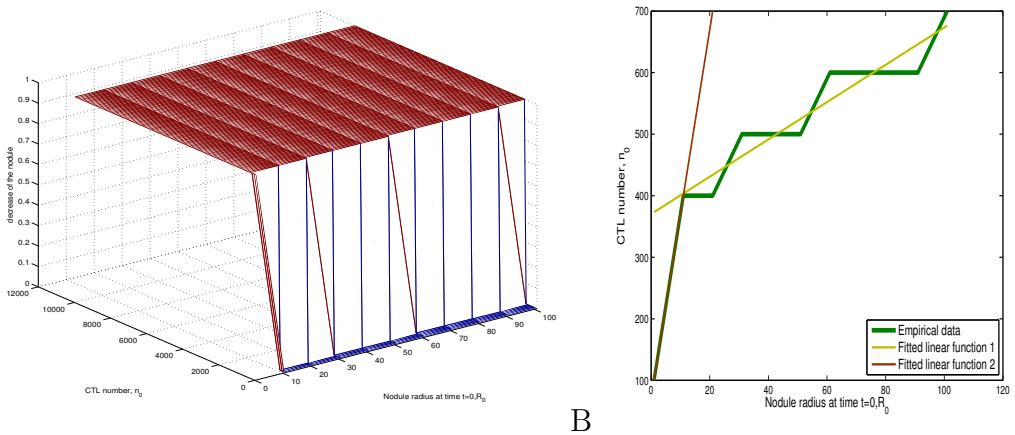



Figure 3.11: **Nodule radius, alive cell number and scout CTL number along the time, beginning at $R_0 = 1cm$.** There is attraction of CTL toward the nodule. In first line, there are around 5 alive tumor cells against 1 CTL.




A B

Figure 3.12: **Decrease or not of the nodule according to $n_0$ and $R_0$, under simulated immunotherapy.** (A) Decrease (1) or not (0) of the nodule according to the CTL number and the nodule radius at time $t = 0$. (B) Minimal CTL number leading to nodule eradication, according to the nodule radius.

and for $R_0 \in [100, 1100]\mu m$

$$y_2 = 30 * x + 70.$$

### 3.3.6 Conclusion et perspectives

As similar results are obtained for a Gamma distribution and for a QSD distribution of non scout CTL distribution at time $t$, same conclusion are done. Then, we refer to Results in Subsection 3.2.3.

**QSD calibration** One notes that the CTL number leading to nodule eradication are small. Then, a better calibration of parameters of the QSD has to be done.

In simulated immunotherapy, if we compute the decrease or not of the nodule, for a range of value of $R_0$ and $n_0$ more thin, we expect a better estimation of:

- the fitted equation describing the minimal CTL number leading to nodule eradication;
- the nodule radius that substitutes one equation to the another equation.

**Lowest eigenvalue of the radial O.U. process on compact** This value is always unknown. Here, we give numerical approximation for absorbing point in $\{1, 11, ..., 101\}$ and a drift in $\{0, 0.1, ..., 1\}$. Note that, if the drift is equal to 0, this is the lowest eigenvalue of the Bessel process on compact.

A first view of Figure 3.7, an expected function fitting the first eigenvalue associated to the radial O.U process, could be of the form

$$f(R, d) = \exp(\alpha R + \beta d) - 1,$$

where $d$ is the drift, $R$ the absorbing point, and $\alpha, \beta \in \mathbb{R}_+^2$. But, a logarithmic transformation of the value obtained, does not give a plane. Then, more researches have to be done to estimate a function $f$ description the lowest eigenvalue associated to the radial O.U. process.

## 3.4 Return to the model

**Phase transitions** Here as in the previous part, a phase transition in the minimal CTL number leading to eradicate the nodule is highlighted. In addition we obtain a linear equation describing this number, allowing to predict for nodule radius higher than $1cm$.

**Hypothesis on CTL activities** We assume that a CTL can kill an infinite number of target cell. This hypothesis is quite strong. But, we can consider that a scout CTL, which cannot kill anymore tumor cells, is instantaneously replaced by a new scout CTL. The CTL, which cannot kill anymore, is said "exhausted CTL", and it is instantaneously replaced by a new non scout CTL. That maintains a constant CTL number.

The previous assumptions are quite raw, a better manner to modelize the fact that a CTL can kill only a finite number of tumor cell, is to change the equation describing the expected scout CTL number. Adding in Equation (2.6) a proportion $p_{ehaust}$ of scout CTL leaving the nodule border, because they are exhausted, we obtain:

$$\mathbb{E}_{\mu_t}(L_{t+h}) = \left[\mathbb{E}_{\mu_t}(L_t) + (n_0 - \mathbb{E}_{\mu_t}(L_t))\mathbb{P}_{\mu_t}\left(T^1_{R_t} \leq h\right) - p_{ehaust}\mathbb{E}(L_t)\right] \mathbb{1}_{\mathbb{E}(L_t)>0}.$$

Eliminate scout CTL, which cannot kill anymore, implies to eliminate CTL in the total

CTL population. That leads to have a new equation on expected CTL number:

$$\mathbb{E}(n_0(t+h)) = [\mathbb{E}(n_0(t)) - p_{exhaust}\mathbb{E}(L_t)]\,\mathbb{1}_{\mathbb{E}(n_0(t))>0}.$$

Now, the difficulty is to estimate the parameter $p_{exhaust}$.

**Hypotheses on immunoediting** This model does not take into account the immunoediting process. For the part of the exhaustion of the CTL we refer to the suggestion just bellow. The resistance of tumor cell can be described by a decrease of the parameter $\mu$: the killing rate. But, for the invisibility of the tumor cells, it remains a complex question.

# Troisième partie

# Modélisation de l'activité cytolytique des CTLs par un mélange de lois de Poisson

# Chapitre 4

# Algorithme $EM$ pour des mélanges de lois de Poisson

Des études menées par des biologistes ont montré que la population des précurseurs des Lymphocytes T cytologiques (CTL) est hétérogène. Ainsi, la question de l'hétérogénéité de la population de CTL aux travers de leur fonction cytolytique se pose. En effet, cela peut être la clef de nouvelles stratégies d'immunothérapies anti-tumorales.

Des expériences biologiques soutenues par une étude statistique, conduisent à privilégier l'hypothèse que la population de CTL est divisée en 2 sous-populations. L'algorithme *Expectation Maximisation* ($EM$) mis en place pour un mélange fini de lois de Poisson, permet d'estimer les caractéristiques de ces 2 sous-populations. Une sous-population de CTL, représentant 34% des CTL, détruit en moyenne 6,4 cibles en 12 heures, la seconde sous-population élimine en moyenne 2,8 cibles.

## 4.1 Introduction

### 4.1.1 Modèle biologique

Le mécanisme amenant un CTL à détruire des cellules cibles est étudié par les biologistes depuis de nombreuses années. Ceux ci ont établi qu'ils éliminaient aussi bien des cellules infectées par un virus que des cellules tumorales grâce à la sécrétion de molécules lytiques stockées dans des granules intracellulaires. Cette sécrétion s'opère dès que le CTL a reconnu l'antigène spécifique à la surface de la cellule cible, elle advient au niveau de la synapse immunologique [1] [DL10]. Ce coup létal peut se produire très rapidement (quelques minutes) après la rencontre entre un CTL et une cellule cible [PTSV87]. Des recherches supplémentaires ont montré que les CTL sont capables d'éliminer plusieurs cellules cibles, soit simultanément, soit en série par rebondissement d'une cellule cible à une autre [Rot+78 ; PB82 ; Wie+06].

1. Aire de contact très organisée entre le CTL et sa cible.

Récemment, des études montrent que les précurseurs [2] des CTL réagissent de manière très différente à une même stimulation. Elles vont avoir par exemple des taux de prolifération différents ou des différentiations différentes [Beu+10 ; BGB12].

La question de l'hétérogénéité de la population de CTL se pose, notamment au travers de leur fonction de lyse des cellules cibles. Une telle étude pourrait inspirer de nouvelles stratégies immunothérapeutiques. En effet dans le Chapitre 1 nous avons démontré qu'augmenter le nombre de cellules tumorales éliminées par un seul CTL permet d'augmenter la probabilité d'éradication d'un nodule (voir la Figure 1.7 (A)). Cette augmentation semble bien plus intéressante que l'augmentation de la vitesse d'élimination d'une cible.

Pour observer l'activité cytolytique individuelle des CTL, les biologistes ont confronté un CTL à un grand nombre de cellules cibles (une dizaine) Figure 4.1 (A). Le nombre de cellules mortes par puits a été relevé toutes les 2 heures pendant 12 heures, Figure 4.1 (B) [3]. Il est important de noter que tous les CTL utilisés sont issus du même clone [4]. Ils sont alors a priori identiques, un comportement homogène est donc attendu.

Plus le temps augmente, plus la variabilité dans le nombre de cellules cibles détruites d'un puits à l'autre augmente. L'histogramme (C) de la Figure 4.1 est la répartition du nombre de cellules mortes sur l'ensemble des puits au bout de 12 heures. Ainsi, un CTL peut détruire de 0 à 12 cellules cibles.

Il est donc légitime de se demander si cette variabilité provient de l'activité cytolytique d'une population homogène de CTL ou au contraire si elle provient d'une population hétérogène.

### 4.1.2 Modèle statistique

Pour cette étude, nous disposons du nombre de cibles mortes par puits au cours de 12 heures (Figure 4.1 (B)). Il est traditionnellement admis que les temps de mort arrivent à des instants aléatoires, modélisés par une loi exponentielle.

**Test préliminaire** Nous avons mis en oeuvre un test statistique simple afin d'étudier l'homogénéité de la population.

Dans un premier temps, nous avons supposé que la population de CTL est homogène, c'est-à-dire que tous les CTL ont statistiquement la même capacité d'élimination

2. Les précurseurs sont des cellules qui se développent dans le thymus pour donner naissance à des Lymphocytes T. Pendant ce développement, par des stimulations, les lymphocytes immatures subissent un certain nombre de modifications phénotypiques. Le but final est de produire des CTL capables de reconnaître une cellule cible.

3. Expériences réalisées par Zilton Vasconcelos et Loïc Dupré, membres de l'équipe de Salvatore Valitutti.

4. CTL obtenus par des divisions successives d'un même CTL

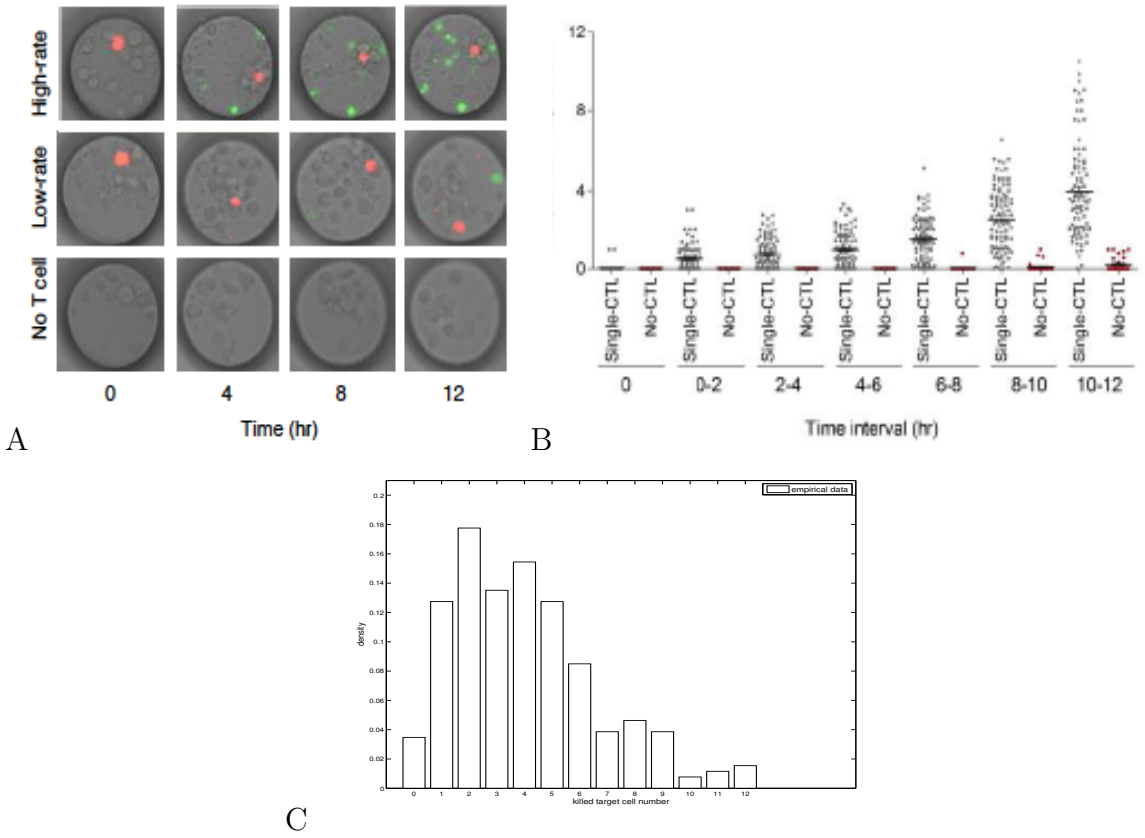



FIGURE 4.1 – **Observations biologiques (*in vitro*).** (A) : Photographie de puits à différentes heures. Les CTL sont marqués par les points rouges. Les cellules cibles mortes sont marquées par les points verts. (B) : Quantification aux temps indiqués du nombre de cellules cibles mortes sur 88 puits contenant un seul CTL et environ 10 cellules cibles (points gris). Même quantification mais sur 33 puits contenant que des cellules cibles, servant à contrôler le faible taux de mort spontanée des cibles (points marrons). Les barres indiquent le nombre moyen de cellules mortes. (C) : Proportion du nombre de cellules mortes par puits.

des cellules cibles. Le nombre de cellules mortes au cours du temps a donc été modélisé par un processus de Poisson. Soient $x = (x_1, ..., x_n)$, où $n = 259$, les observations à 12 heures du nombre de cibles éliminées par un CTL (la Figure 4.1 (C)). Sous notre hypothèse d'homogénéité, cet échantillon proviendrait de $n$ réalisations i.i.d. de lois de Poisson.

Pour tester l'homogénéité de la population, nous avons utilisé un test d'adéquation à une loi. Plus formellement, nous souhaitions tester l'hypothèse

$H_0$ : Les observations sont issues d'une loi de Poisson de paramètre $\lambda$.

Il n'y avait pas d'a priori sur $\lambda$, et nous avons donné l'estimateur obtenu sur les données :

$$\lambda_n = 4,06,$$

la moyennes des observations. Pour tout $i \in \{0, ..., 13\}$, soit $f_i$ la fréquence d'élimination de $i$ cellules cibles. Pour tout $i \in \{1, ..., 12\}$, soit $p_i = \mathbb{P}(X_\lambda = i)$, et $p_{13} = \mathbb{P}(X_\lambda \geq 13)$, où $X_\lambda \sim \mathcal{P}(\lambda)$. Ainsi la statistique de test a été définie par

$$T = \sum_{i=0}^{13} \frac{n \cdot (f_i - p_i)^2}{p_i}.$$

Sous l'hypothèse $H_0$, la statistique de test devrait être distribuée comme un $\chi_2$ a 11 degrés de libertés ($(13-1)-1$ car 1 paramètre a été estimé). Soit $\alpha = 0,001$, nous avons obtenu $T > 130$, la statistique de test $T$ est beaucoup plus grande que le quantile d'ordre $1-\alpha$ d'une distribution de $\chi_2$ à 11 degrés de liberté, qui vaut 31,26.

Nous sommes donc amenés à rejeter l'hypothèse $H_0$ d'homogénéité de la population de CTL avec une erreur de première espèce (probabilité de rejeter $H_0$ alors que $H_0$ est vraie) de 0,001.

Nous supposons alors que la population de CTL est hétérogène et qu'elle est divisée en plusieurs sous-populations. Chacune de ces sous-populations est caractérisée par le nombre moyen de cibles éliminées par un CTL et par sa proportion dans la population totale. Il est raisonnable de modéliser l'hétérogénéité des CTL par un mélange de lois de Poisson.

### 4.1.3 Modèles de Mélanges de lois

Les modèles de mélanges finis de lois sont étudiés depuis de nombreuses années (voir par exemple les monographies [TSM85 ; MP00]) en raison de leurs applications aussi bien en informatique, en biologie ou en finance. Ces modèles traduisent le fait qu'une population est divisée en plusieurs sous-populations avec leurs propres caractéristiques définies par une distribution de probabilité.

Généralement, les lois de probabilités caractérisant les sous-populations sont supposées appartenir à une famille de lois paramétriques. La valeur des paramètres des lois ainsi que les proportions des sous-groupes sont inconnues. Une problématique statistique de tels modèles consiste à estimer de tels paramètre. A cette fin, l'algorithme *Expectation Maximisation*, introduit par [DLR77], est couramment utilisé.

Plus particulièrement, les modèles de mélanges gaussiens, aussi bien univariés que multivariés, ont été étudiés [MRM13 ; GV01], notamment à cause du rôle central que joue la loi normale dans la théorie des probabilités et leurs applications ainsi que la possibilité de faire des calculs plus ou moins explicites. Des logiciels ont été développés afin de réaliser des études statistiques pour des classes particulières de modèles gaussiens, par exemple les logiciels Capushe par [BMM12], Emmix par [McL+99] ou Mixmod par [Bie+06].

La loi de Poisson trouve elle aussi de nombreuses applications, notamment en télécommunication, en finance, ou en biologie comme dans notre cas. Cependant, le mélange de lois de Poisson est moins étudié. Remarquons que les logiciels cités ne sont pas développés pour de tels mélanges, alors que l'algorithme $EM$ peut s'appliquer pour ces modèles.

Dans ce chapitre nous développons un algorithme $EM$ pour de tels mélanges afin d'estimer leurs paramètres.

Dans un premier temps, nous définissons plus formellement un mélange de lois de Poisson et nous nous assurons de l'indentifiabilité du modèle. Nous n'avons pas d'a priori sur le nombre de sous-groupes qui constitue la population totale des CTL, nous supposons que le mélange est formé de $\kappa$ lois de Poisson, où $\kappa \in \mathbb{N}^*$. Dans un deuxième temps, nous mettons en place l'algorithme *Expectation Maximisation* ($EM$) pour un mélange de $\kappa$ lois de Poisson, afin d'estimer ses paramètres. Pour finir nous appliquons nos résultats à des données simulées ainsi qu'aux observations biologiques.

## 4.2 Identifiabilité du modèle de mélange poissoniens

Pour inférer les paramètres d'un modèle, il faut au préalable s'assurer de l'identifiabilité de ce modèle. C'est-à-dire que deux valeurs distinctes dans l'espace des paramètres possibles du modèle donnent lieux à deux lois différentes. Cette unicité assure l'identifiabilité du modèle.

### 4.2.1 Mélanges de lois de Poisson

Nous rappelons la définition d'un mélange de lois de Poisson par sa densité par rapport à la mesure de comptage. Elle sera utilisée dans la section suivante pour estimer les paramètres de la densité de ce mélange grâce à l'algorithme $EM$.

**Définition 4.2.1.** *Soit l'ensemble de toutes les densités de loi de Poisson :*

$$\mathcal{F} = \left\{ f_\lambda : \mathbb{N} \to \mathbb{R}_+ \ : \ k \mapsto e^{-\lambda}\frac{\lambda^k}{k!}, \quad \lambda \in \mathbb{R}_+ \right\}. \tag{4.1}$$

*Considérons $\mathcal{G}$ l'ensemble des probabilités sur $\mathbb{R}_+$ et notons $G$ un élément de $\mathcal{G}$. Nous appelons un $G-$mélange de $\mathcal{F}$, ou un* **mélange de lois de Poisson**, *une loi qui a pour densité*

$$h_G(k) = \int_{\mathbb{R}_+} f_\lambda(k)\,\mathrm{d}G(\lambda), \quad \textit{pour tout } k \in \mathbb{N}.$$

*L'ensemble des densités de mélanges de lois de Poisson est noté $\mathcal{H}$.*

Par exemple, si nous considérons un mélange, par une loi de Bernoulli de paramètre $p \in ]0,1[$, de 2 lois de Poisson de paramètre $\lambda$ et $\mu \in \mathbb{R}_+$. Alors la loi de mélange, pour $\alpha \in \mathbb{R}_+$ est $G(\alpha) = p\delta_\lambda(\alpha) + (1-p)\delta_\mu(\alpha)$, et la densité du mélange est

$$h_G(k) = pe^{-\lambda}\frac{\lambda^k}{k!} + (1-p)e^{-\mu}\frac{\mu^k}{k!}, \text{ pour tout } k \in \mathbb{N}.$$

### 4.2.2 Identifiabilité

**Définition 4.2.2.** *On dit que $\mathcal{H}$ est identifiable si la correspondance*

$$\begin{aligned} h : \mathcal{G} &\to \mathcal{H} \\ G &\mapsto h_G(k) = \int_{\mathbb{R}_+} f_\lambda(k)\,\mathrm{d}G(\lambda), \text{ pour tout } k \in \mathbb{N}, \end{aligned}$$

*est injective de $\mathcal{G}$ sur $\mathcal{H}$.*

**Proposition 4.2.1.** *Le modèle de mélange de lois de Poisson est identifiable.*

*Démonstration.* Soient $G_1, G_2 \in \mathcal{G}$, pour montrer l'injectivité de la fonction $h$, il faut montrer

$$\forall\ k \in \mathbb{N},\quad h_{G_1}(k) = h_{G_2}(k) \ \Rightarrow\ G_1 = G_2.$$

Pour cela nous allons montrer que si deux mélanges (obtenus par les lois mélangeantes $G_1$ et $G_2$) ont même loi, alors les fonctions caractéristiques des lois $G_1$ et $G_2$ sont égales.

Soit $\mathcal{X}_1$, respectivement $\mathcal{X}_2$, une variable aléatoire qui suit la loi du $G_1-$mélange, respectivement $G_2-$mélange, alors par hypothèse, pour tout $t \in \mathbb{R}$,

$$\mathbb{E}\left(e^{it\mathcal{X}_1}\right) = \sum_{k=0}^{\infty} e^{itk} h_{G_1}(k) = \sum_{k=0}^{\infty} e^{itk} h_{G_2}(k) = \mathbb{E}\left(e^{it\mathcal{X}_2}\right). \tag{4.2}$$

Nous noterons $X_\lambda$ une variable aléatoire qui suit une loi de Poisson de paramètre $\lambda$.

Remarquons que

$$\begin{aligned}
\mathbb{E}\left(e^{it\mathcal{X}_1}\right) &= \sum_{k=0}^{\infty} e^{itk} h_{G_1}(k) \\
&= \sum_{k=0}^{\infty} e^{itk} \int_{\mathbb{R}_+} f_\lambda(k)\, \mathrm{d}G_1(\lambda) \\
&= \int_{\mathbb{R}_+} \sum_{k=0}^{\infty} e^{itk} f_\lambda(k)\, \mathrm{d}G_1(\lambda) \\
&= \int_{\mathbb{R}_+} \mathbb{E}(e^{itX_\lambda})\, \mathrm{d}G_1(\lambda) \\
&= \int_{\mathbb{R}_+} e^{\lambda(e^{it}-1)}\, \mathrm{d}G_1(\lambda).
\end{aligned}$$

Nous obtenons de même $\mathbb{E}\left(e^{it\mathcal{X}_2}\right) = \int_{\mathbb{R}_+^*} e^{\lambda(e^{it}-1)}\, \mathrm{d}G_2(\lambda)$.

Notons, pour $i = 1, 2$, $\varphi_i(z) = \int_{\mathbb{R}_+} e^{\lambda(z-1)}\, \mathrm{d}G_i(\lambda)$. La fonction $\varphi_i$ est définie pour tout $z \in \varpi$, où $\varpi = \{z = a + ib \in \mathbb{C},\ a, b \in \mathbb{R} \text{ et } a < 1\}$ (voir la Figure 4.2). En effet, notons, pour tout $\lambda \in \mathbb{R}_+$, $\phi_\lambda : z \to e^{\lambda(z-1)}$. La fonction $\phi_\lambda$ est holomorphe sur $\mathbb{C}$, en revanche $\varphi_i$ est holomorphe où $\phi_\lambda$ est intégrable, c'est-à-dire sur $\varpi$. De plus, par (4.2), pour tout $z \in \mathcal{S}^1$ (la sphère unité), $\varphi_1(z) = \varphi_2(z)$. Donc par prolongement analytique $\varphi_1 \equiv \varphi_2$ sur tout $\varpi$.

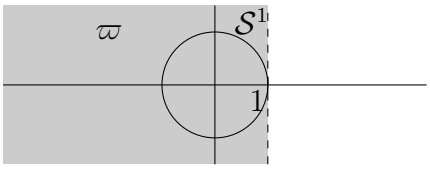



FIGURE 4.2 – Représentation du domaine $\varpi$.

Pour conclure, montrons que $\varphi_1 \equiv \varphi_2$ sur la fermeture de $\varpi$.
On définit, pour $b \in \mathbb{R}$ et $\lambda \in \mathbb{R}_+^*$, la suite $z_n^{(\lambda)} = 1 - \frac{1}{n\lambda} + ib$, et pour $\lambda = 0$, $z_n^{(0)} = 1 + ib$. Remarquons que, quelque soit $\lambda \in \mathbb{R}_+$, cette suite converge vers $z = 1 + ib$. En utilisant le théorème de convergence dominée nous montrons que $\lim_{n\to\infty} \varphi_1(z_n) = \varphi_1(z)$.

En effet,

$$\begin{aligned}\int_{\mathbb{R}_+} \left|e^{\lambda(z_n-1)} - e^{\lambda ib}\right| \mathrm{d}G_1(\lambda) &= \int_{\mathbb{R}_+} \left|e^{\lambda\left(-\frac{1}{n\lambda}+ib\right)} - e^{\lambda ib}\right| \mathrm{d}G_1(\lambda) \\ &= \int_{\mathbb{R}_+} \left|e^{\lambda ib}\left(e^{-\frac{\lambda}{n\lambda}} - 1\right)\right| \mathrm{d}G_1(\lambda) \\ &= \int_{\mathbb{R}_+} \left|e^{-\frac{1}{n}} - 1\right| \mathrm{d}G_1(\lambda) \\ &= \int_{\mathbb{R}_+} \left(\frac{1}{n} + o\left(\frac{1}{n}\right)\right) \mathrm{d}G_1(\lambda) \\ &= \frac{1+o(1)}{n}.\end{aligned}$$

Donc quand $n$ tend vers $\infty$, $\int_{\mathbb{R}_+} \left|e^{\lambda(z_n-1)} - e^{\lambda ib}\right| \mathrm{d}G_1(\lambda)$ tend vers 0.

De plus, vu que $\varphi_1 \equiv \varphi_2$ sur $\varpi$ et que, pour tout $n$ $z_n \in \varpi$, il vient $\varphi_1(z_n) = \varphi_2(z_n)$ pour tout $n$, et donc $\varphi_1(z) = \varphi_2(z)$. Or cette égalité est vraie pour tout $z = 1 + ib$ avec $b \in \mathbb{R}$, *i.e.* $\varphi_1 \equiv \varphi_2$ sur tout $\bar{\varpi}$.

Ainsi, pour tout $t \in \mathbb{R}$, $\mathbb{E}(e^{itZ_1}) = \varphi_1(1+it) = \varphi_2(1+it) = \mathbb{E}(e^{itZ_2})$, où $Z_i$ suit la loi de mélange $G_i$. Nous concluons que les variables aléatoires $Z_1$ et $Z_2$ ont la même fonction caractéristique, elles suivent donc la même loi.

□

**Remarque 4.2.2.** *Cette démonstration est adaptée des articles [Tei54 ; Tei61] qui montrent plus généralement l'identifiabilité de mélange de lois additivement fermées.*

### 4.2.3 Mélange fini de lois de Poisson

**Définition 4.2.3.** *Soient $\kappa \in \mathbb{N}$, $\kappa \geq 2$. Notons*

$$\breve{\Theta}_\kappa = \{\theta = (\underbrace{p_1, ..., p_\kappa}_{p}, \underbrace{\lambda_1, ..., \lambda_\kappa}_{\lambda});\ p \in \breve{\mathfrak{S}}_{\kappa-1} \ \textit{et} \ \ \lambda \in \mathbb{R}_+^\kappa\}, \tag{4.3}$$

*où $\breve{\mathfrak{S}}_{\kappa-1} = \{p \in ]0,1[^\kappa$ tel que $\sum_{k=0}^{\kappa} p_k = 1\}$ est le simplexe ouvert de dimension $\kappa - 1$.*

*Soit $\theta \in \breve{\Theta}_\kappa$. Un* **mélange de $\kappa$ lois de Poisson** *de vecteur des paramètres $\theta$ est défini par*

$$h(k;\theta) = \sum_{j=1}^{\kappa} p_j e^{-\lambda_j} \frac{\lambda_j^k}{k!}, \ \textit{pour tout } k \in \mathbb{N}.$$

Autrement dit, $\breve{\Theta}_\kappa$ est l'ensemble des vecteurs de paramètres possibles pour un mélange d'exactement $\kappa$ lois de Poisson, où, pour $i = 1, ..., \kappa$, $(\lambda_i)_i$ sont les paramètres des $\kappa$ lois de Poisson, et $p_i$ est la probabilité de choisir la loi de paramètre $\lambda_i$.

**Remarque 4.2.3.** *Si nous considérons un mélange fini de lois, il n'y a pas identifiabilité. En effet, soient $\theta = (p, \lambda) \in \breve{\Theta}_\kappa$ et $\theta' = (p', \lambda') \in \breve{\Theta}_\kappa$, tel qu'il existe une*

*permutation $s$ de $\{1,..,\kappa\}$ qui vérifie $\theta' = (s(p), s(\lambda))$, alors $h(.;\theta) = h(.;\theta')$.*

*On parle alors d'identifiabilité à permutation près. Imposer un ordre sur les poids ou sur les paramètres des lois permet d'assurer l'identifiabilité.*

## 4.3 Algorithme $EM$ pour un mélange fini de lois de Poisson

### 4.3.1 Mélanges poissoniens et vraisemblance

**Vraisemblance d'un mélange poissonien** Dans un cadre statistique, $n$ réalisations indépendantes sont observées, notées $x = (x_1, ..., x_n)$, de loi de mélange $h(.;\theta)$, de paramètre $\theta$ inconnu. Le but est d'estimer le vecteur des paramètres $\theta$, donnant une densité la plus proche de $h$ au sens de la divergence de Kullback-Leibler. C'est un problème d'estimation de densité en cherchant à estimer le vecteur des paramètres $\hat{\theta}$ maximisant la vraisemblance observée :

$$L(\theta; x) = \prod_{i=1}^{n} h(x_i; \theta).$$

**Vraisemblance complète** A ce stade, il n'est pas possible de calculer le maximum de la vraisemblance. Un vecteur aléatoire $z = (z_1, ..., z_n)$, non observé et à valeurs dans $\{1, ..., \kappa\}$, vient compléter les données. Soient $i, j \in \{1, ..., n\}$, la variable $z_i$ correspond à la loi de Poisson dont est issue l'observation $x_i$, ainsi la loi de $z_i$ est définie par $\mathbb{P}(z_i = k) = p_k$. De plus $z_i$ est indépendante de $z_j$, pour tout $i \neq j$.

Notant $g(z_i|x_i;\theta)$ la densité de $z_i$ sachant $x_i$ et le paramètre $\theta$, la log-vraisemblance complète du modèle est déifinie par

$$\begin{aligned} L(\theta; x, z) &= \prod_{i=1}^{n} g(z|x;\theta) h(x;\theta) \\ &= \prod_{i=1}^{n} \prod_{j=1}^{\kappa} \left( p_j e^{-\lambda_j} \frac{\lambda_j^{x_i}}{x_i!} \right)^{\eta_{j,i}}, \end{aligned} \tag{4.4}$$

où $\eta_{j,i} = \frac{\prod_{\alpha=1|\alpha\neq j}^{\kappa}(\alpha - z_i)}{\prod_{\alpha=1|\alpha\neq j}^{\kappa}(\alpha - j)}$ est le polynôme d'interpolation de Lagrange qui satisfait,

$$\eta_{j,i} = \begin{cases} 1 & \text{si } z_i = j \\ 0 & \text{sinon} \end{cases}.$$

On en déduit $L(\theta; x) = \frac{L(\theta;x,z)}{L(\theta;z)}$. L'estimateur du maximum de vraisemblance devient :

$$\begin{aligned}\hat{\theta} &= \underset{\breve{\Theta}_\kappa}{\operatorname{argmax}} \frac{L(\theta; x, z)}{L(\theta; z)} \\ &= \underset{\breve{\Theta}_\kappa}{\operatorname{argmax}} \log \left[\frac{L(\theta; x, z)}{L(\theta; z)}\right] \\ &= \underset{\breve{\Theta}_\kappa}{\operatorname{argmax}} \left(\log\left[L(\theta; x, z)\right] - \log\left[L(\theta; z)\right]\right).\end{aligned}$$

A présent, pour calculer $\hat{\theta}$, nous utilisons l'algorithme $EM$ proposé par [DLR77], une monographie est donnée dans [MK08]. Le livre [MP00] étudie entre autres les propriétés de l'algorithme $EM$ dans le cadre spécifique des mélanges finis de lois.

### 4.3.2 Algorithme EM

**Définition 4.3.1.** *Un* **algorithme itératif***, d'après [DLR77], est une règle, applicable à partir de n'importe quel point d'un ensemble $\breve{\Theta}_\kappa$, et à valeur dans $\breve{\Theta}_\kappa$. C'est-à-dire, une application $M : \breve{\Theta}_\kappa \to \breve{\Theta}_\kappa : \theta \mapsto M(\theta)$, telle qu'à chaque étape, $\theta^t \to \theta^{t+1} = M(\theta^t)$.*

L'algorithme $EM$, pour $Expectation$, $Maximisation$, est un algorithme itératif qui, après initialisation du vecteur $\theta^1$, alterne les deux étapes suivantes : à l'itération $t$,

1. $Etape - E$ (Expectation) : calcule l'espérance de la log-vraisemblance complète par rapport à la variable aléatoire $Z$, conditionnellement aux observations $x$ et au paramètre courant $\theta^t$,

$$\mathbb{E}\left(\log\left[L(\theta; x, z)\right] - \log\left[L(\theta; z)\right] | x, \theta^t\right);$$

2. $Etape - M$ (Maximisation) : maximise le paramètre inconnu $\theta$ dans l'espérance obtenue à l'étape précédente,

$$\theta^{t+1} = \underset{\theta \in \breve{\Theta}_\kappa}{\operatorname{argmax}} \mathbb{E}\left(\log\left[L(\theta; x, z)\right] - \log\left[L(\theta; z)\right] | x, \theta^t\right).$$

L'estimateur $\theta^{t+1}$ de $\theta$ est le paramètre courant de l'étape $t + 1$.

Notons $Q(\theta, \theta^t) = \mathbb{E}\left(\log\left[L(\theta; x, z)\right] | x, \theta^t\right)$ et $H(\theta, \theta^t) = \mathbb{E}\left(\log\left[L(\theta; z)\right] | x, \theta^t\right)$. La fonction $H(\theta, \theta^t)$ étant inconnue, il n'est pas possible de calculer l'*étape*$-M$. L'inégalité de Jensen permet de montrer que, pour tout $\theta \neq \theta^t$, $H(\theta, \theta^t) \leq H(\theta^t, \theta^t)$. Cela garantit que maximiser $Q(\theta, \theta^t)$ augmente $L(\theta; x, z)$.

Ainsi l'*étape* $-E$ consiste à calculer

$$\begin{aligned} Q(\theta, \theta^t) &= \mathbb{E}\left(\log\left[L(\theta; x, z)\right] | x, \theta^t\right) \\ &= \sum_{i=1}^{n}\sum_{j=1}^{\kappa} \log\left(p_j e^{-\lambda_j}\frac{\lambda_j^{x_i}}{x_i!}\right)\epsilon_{i,j,t}, \end{aligned}$$

où $\epsilon_{i,j,t} = \frac{e^{-\lambda_j^t}(\lambda_j^t)^{x_i} p_j^t}{\sum_{l=1}^{\kappa} p_l^t e^{-\lambda_l^t}(\lambda_l^t)^{x_i}}$. Notons que pout tout $i, j$, $\epsilon_{i,j,t} \in [0,1]$ et que $\sum_{j=1}^{k} \epsilon_{i,j,t} = 1$.

L'*étape* $-M$ consiste à calculer

$$\begin{aligned} \theta^{t+1} &= \underset{\theta\in\breve{\Theta}_\kappa}{\operatorname{argmax}}\, Q(\theta, \theta^t) \\ &= \left(\left(\frac{\sum_{i=1}^{n}\epsilon_{i,j,t}}{n}\right)_{j\in\{1,\ldots,\kappa\}}, \left(\frac{\sum_{i=1}^{n}\epsilon_{i,j,t}x_i}{\sum_{i=1}^{n}\epsilon_{i,j,t}}\right)_{j=1,\ldots,\kappa}\right). \end{aligned}$$

Chaque itération garantit une augmentation de la vraisemblance [DLR77]. Dans la sous-section suivante nous montrons que l'algorithme converge.

Les résultats obtenus pour les étapes $-E$ et $-M$ dans le cas d'un mélange fini de lois de Poisson sont démontrés dans les lemmes suivants.

**Lemme 4.3.1.** *Soit $x = (x_1, ..., x_n)$ un $n-$échantillon d'un mélange de $\kappa$ lois de Poisson de paramètre inconnu $\theta \in \breve{\Theta}_\kappa$. Soit $\theta^t = (p_1^t, ..., p_\kappa^t, \lambda_1^t, ..., \lambda_\kappa^t)$ l'estimateur de $\theta$ obtenu à l'itération $t-1$. A l'itération $t$,*

$$Q(\theta, \theta^t) = \sum_{i=1}^{n}\sum_{j=1}^{\kappa} \log\left(p_j e^{-\lambda_j}\frac{\lambda_j^{x_i}}{x_i!}\right)\epsilon_{i,j,t}.$$

*Démonstration.* En utilisant la vraisemblance complète du modèle (4.4), l'*étape* $-E$ devient

$$\begin{aligned} Q(\theta, \theta^t) &= \mathbb{E}\left[\log\left(\prod_{i=1}^{n}\prod_{j=1}^{\kappa}\left(p_j e^{-\lambda_j}\frac{\lambda_j^{x_i}}{x_i!}\right)^{\eta_{j,i}}\right) | x, \theta^t\right] \\ &= \sum_{i=1}^{n}\sum_{j=1}^{\kappa}\mathbb{E}\left[\log\left(\left(p_j e^{-\lambda_j}\frac{\lambda_j^{x_i}}{x_i!}\right)^{\eta_{j,i}}\right) | x, \theta^t\right] \\ &= \sum_{i=1}^{n}\sum_{j=1}^{\kappa}\mathbb{E}\left[\eta_{j,i}\log\left(p_j e^{-\lambda_j}\frac{\lambda_j^{x_i}}{x_i!}\right) | x, \theta^t\right] \\ &= \sum_{i=1}^{n}\sum_{j=1}^{\kappa}\log\left(p_j e^{-\lambda_j}\frac{\lambda_j^{x_i}}{x_i!}\right)\mathbb{E}\left[\eta_{j,i}|x, \theta^t\right]. \end{aligned}$$

Rappelons que $z_i \in \{1, ..., \kappa\}$ est une variable aléatoire et $\eta_{j,i} \sim \max(1 - |z_i - j|, 0)$,

d'où

$$\begin{aligned}\mathbb{E}\left[\eta_{j,i}|x,\theta^t\right] &= \mathbb{P}(\eta_{j,i}=1|x,\theta^t)\\ &= \mathbb{P}(z_i=j|x,\theta^t)\\ &= \frac{\mathbb{P}\left[X_{\lambda_i}=x_i|z_i=j,\theta^t\right]\mathbb{P}\left[z_i=j|\theta^t\right]}{\mathbb{P}(X_{\lambda_i}=x_i|\theta^t)}\\ &= \frac{e^{-\lambda_j^t}(\lambda_j^t)^{x_i}p_j^t}{\sum_{l=1}^{\kappa}p_l^t e^{-\lambda_l^t}(\lambda_l^t)^{x_i}}.\end{aligned}$$

□

**Lemme 4.3.2.** *Sous les conditions du lemme précédent, soit $\lambda \in \mathbb{R}_+^\kappa$ fixé, l'application*

$$Q_1 : \left\{p \in ]0,1[^\kappa, \quad \sum_{l=1}^{\kappa} p_l = 1\right\} \to \mathbb{R} \ : \ p \mapsto Q\left((p,\lambda),\theta^t\right)$$

*a un unique maximum, atteint en :*

$$\hat{p} = \left(\left(\frac{\sum_{i=1}^n \epsilon_{i,j,t}}{n}\right)_{j\in\{1,\dots,\kappa\}}\right).$$

*Démonstration.* Commençons par localiser les points critiques de $Q_1$. Or $\sum_{l=1}^{\kappa} p_l = 1$, il existe une coordonnée de $p$ non nulle, notons-la $\kappa$. Ainsi $p_\kappa \neq 0$ et $\epsilon_{i,\kappa,t} \neq 0$,

$$\begin{aligned}\frac{\partial Q(\theta,\theta^t)}{\partial p_j} &= \sum_{i=1}^n\sum_{j=1}^{\kappa}\frac{\partial}{\partial p_j}\log\left(p_j e^{-\lambda_j}\frac{\lambda_j^{x_i}}{x_i!}\right)\epsilon_{i,j,t}\\ &= \sum_{i=1}^n \frac{\partial}{\partial p_j}\log(p_j)\,\epsilon_{i,j,t} + \sum_{i=1}^n\frac{\partial}{\partial p_j}\log(p_\kappa)\,\epsilon_{i,\kappa,t}\\ &= \frac{\sum_{i=1}^n\epsilon_{i,j,t}}{p_j} - \frac{\sum_{i=1}^n\epsilon_{i,\kappa,t}}{p_\kappa}.\end{aligned}$$

Ainsi

$$\frac{\partial Q(\theta,\theta^t)}{\partial p_j} = 0 \iff \frac{\sum_{i=1}^n\epsilon_{i,j,t}}{p_j} = \frac{\sum_{i=1}^n\epsilon_{i,\kappa,t}}{p_\kappa}. \tag{4.5}$$

Notons que pour tout $j \in \{1,...,\kappa\}$,

$$\hat{p}_j = \frac{\sum_{i=1}^n\epsilon_{i,j,t}}{n},$$

est une solution de l'équation différentielle (4.5).

De plus, remarquons que $Q(\theta,\theta^t)$ est strictement concave en $\theta$. En effet, soient

$\alpha \in [0,1]$, $\theta, \theta' \in \breve{\Theta}_\kappa$, par (4.4)

$$\begin{aligned} Q(\alpha\theta + (1-\alpha)\theta', \theta^t) &= Q(\alpha(p_1,...,p_\kappa,\lambda_1,...,\lambda_\kappa) + (1-\alpha)(p_1,...,p_\kappa,\lambda_1,...,\lambda_\kappa), \theta^t) \\ &= \sum_{i=1}^{n}\sum_{j=1}^{\kappa} \log\left[(\alpha p_j + (1-\alpha)p'_j)e^{-\alpha\lambda_j-(1-\alpha)\lambda'_j}\frac{(\alpha\lambda_j+(1-\alpha)\lambda'_j)^{x_i}}{x_i!}\right] \\ &\quad \times \mathbb{E}\left[\eta_{j,i}|x,\theta^t\right] \end{aligned}$$

Or, la somme de fonctions concaves est concave et par la propriété de stricte concavité de la fonction logarithme, l'application $(p,\lambda) \to \log\left(pe^{-\lambda}\frac{\lambda^x}{x!}\right)$ est strictement concave en ses deux variables. Ainsi, $p^t = \left(\frac{\sum_{i=1}^n \epsilon_{i,1,t}}{n}, ..., \frac{\sum_{i=1}^n \epsilon_{i,\kappa,t}}{n}\right)$ est le maximum global de $Q$. □

**Lemme 4.3.3.** *Sous les conditions du lemme précédent, soit $p \in ]0,1[^\kappa$ fixé tel que $\sum_{j=1}^\kappa p_j = 1$, l'application*

$$Q_2 : \mathbb{R}_+^\kappa \to \mathbb{R} \ : \ \lambda \mapsto Q\left((p,\lambda),\theta^t\right).$$

*a un unique maximum en :*

$$\hat{\lambda} = \left(\left(\frac{\sum_{i=1}^n \epsilon_{i,j,t}x_i}{\sum_{i=1}^n \epsilon_{i,j,t}}\right)_{j=1,...,\kappa}\right).$$

*Démonstration.* Commençons par calculer les points critiques de l'application de $Q_2$,

$$\begin{aligned} \frac{\partial Q(\theta,\theta^t)}{\partial\lambda_j} &= \sum_{i=1}^{n}\sum_{j=1}^{\kappa}\frac{\partial}{\partial\lambda_j}\log\left(p_j e^{-\lambda_j}\frac{\lambda_j^{x_i}}{x_i!}\right)\epsilon_{i,j,t} \\ &= \sum_{i=1}^{n}\frac{\partial}{\partial\lambda_j}\log\left(p_j e^{-\lambda_j}\frac{\lambda_j^{x_i}}{x_i!}\right)\epsilon_{i,j,t} \\ &= \sum_{i=1}^{n}\frac{\partial}{\partial\lambda_j}[-\lambda_j + x_i\log(\lambda_j)]\epsilon_{i,j,t} \\ &= \sum_{i=1}^{n}\epsilon_{i,j,t}\left(-1+\frac{x_i}{\lambda_j}\right). \end{aligned}$$

Ainsi,

$$\begin{aligned} \frac{\partial Q(\theta,\theta^t)}{\partial\lambda_j} = 0 &\iff \sum_{i=1}^{n}\epsilon_{i,j,t}\left(-1+\frac{x_i}{\lambda_j}\right) = 0 \\ &\iff \frac{\sum_{i=1}^n \epsilon_{i,j,t}x_i}{\lambda_j} = \sum_{i=1}^{n}\epsilon_{i,j,t} \\ &\iff \lambda_j = \frac{\sum_{i=1}^n \epsilon_{i,j,t}x_i}{\sum_{i=1}^n \epsilon_{i,j,t}}. \end{aligned}$$

Dans la démonstration précédente, nous avons vu que $Q$ est strictement concave, alors $\lambda^t$ est l'unique maximum.

□

**Remarque 4.3.4.** *A présent, nous pouvons donner la loi a posteriori de la variable aléatoire $z_1$ sachant les observations. C'est-à-dire pour chaque observation $x_i$, il est possible d'estimer $z_i$, sa probabilité d'appartenir au groupe $j$. L'estimateur $\hat{z}_i$ de cette probabilité s'obtient grâce à l'estimation de $\hat{\theta}$ par l'algorithme EM :*

$$\hat{z}_i = \begin{cases} 1 & si \;\; \hat{p}_k e^{-\hat{\lambda}_k} \frac{\hat{\lambda}_k^{x_i}}{x_i!} > \hat{p}_l e^{-\hat{\lambda}_l} \frac{\hat{\lambda}_l^{x_i}}{x_i!}, \quad pour\ tout\ l \neq k \\ 0 & sinon \end{cases} .$$

*En statistique bayésienne, cette règle de classification est appelée règle du maximum a posteriori.*

### 4.3.3 Quelques propriétés de l'algorithme $EM$

**Convergence de l'algorithme $EM$** La convergence de l'algorithme $EM$ a été étudiée notamment dans les articles [DLR77 ; Wu83].

Dans un cadre plus général, notons $\Theta$ l'ensemble des paramètres, $l(\theta; x, z) = \log L(\theta; x, z)$ la log-vraisemblance complète du modèle et $Q(\theta, \theta') = \mathbb{E}(\log L(\theta'; x, z)|x, \theta')$. Soient les hypothèses

1. $\Theta$ est un sous-ensemble de $\mathbb{R}^r$, où $r \in \mathbb{N}^*$ ;
2. *compacité :* $\Theta_{\theta_0} = \{\theta \in \Theta \;:\; l(\theta; x, z) > l(\theta_0; x, z)\}$ est compact pour tout $\theta_0$ tel que $l(\theta_0) > -\infty$ ;
3. *régularité de $l$ :* l'application $l$ est continue sur $\Theta$ et différentiable à l'intérieur de $\Theta$ ;
4. *régularité de $Q$ :* la fonction $Q(\theta', \theta)$ est continue en ses deux variables $\theta$ et $\theta'$, de plus $\sup_{\theta' \in \Theta} Q(\theta', \theta) > Q(\theta, \theta)$ pour tout point critique mais non extremum local $\theta$ ;

**Theorem 4.3.5** (Jeff Wu (1983))**.** *Si $\Theta$ vérifie les conditions 1, 2, $l$ satisfait 3 et $Q$ satisfait 4, alors toutes limites de suites $(\theta^t)$ d'un algorithme EM sont des maximum locaux de $L$ et $L(\theta^t)$ converge de manière monotone vers $L(\theta^*)$ où $\theta^*$ est un maximum local.*

Nous allons montrer que notre modèle vérifie les hypothèses de ce théorème.

L'ensemble $\breve{\Theta}_\kappa$, défini en (4.3), est un sous-ensemble de $\mathbb{R}^{2\kappa-1}$. La fonction $l(\theta; x, z)$ est continue et différentiable sur $\breve{\Theta}_\kappa$. L'application $Q(\theta; \theta')$ est de classe $\mathcal{C}^\infty$ en les variables $\theta$ et $\theta'$. De plus, comme $Q(\theta, \theta^t)$ est strictement concave en $\theta$, l'hypothèse 4 est vérifiée. Par contre, l'ensemble $\breve{\Theta}_{\kappa,\theta_0} = \{\theta \in \breve{\Theta}_\kappa,\ l(\theta; x) \geq l(\theta_0; x)\}$ ne vérifie pas la propriété 2 de compacité. La fonction $l$ étant continue en $\theta$, l'ensemble $\breve{\Theta}_{\kappa,\theta_0}$ est fermé,

il reste à compactifier $\breve{\Theta}_\kappa$. Remarquons que l'estimateur de l'algorithme $EM$ à l'étape $t-1$, $(p^t, \lambda^t) \in [0,1]^\kappa \times [0, \max_i x_i]^\kappa$. Remplaçons $\breve{\Theta}_\kappa$ par

$$\Theta_\kappa^1 = \{(p,\lambda) = (p_1, ..., p_\kappa, \lambda_1, ..., \lambda_\kappa);\ p \in \mathfrak{S}_{\kappa-1},\ \lambda \in [0, \max_i x_i]^\kappa\},$$

qui est compact.

**Consistance de l'algorithme** $EM$ Cela revient à étudier la consistance de l'estimateur du maximum de vraisemblance (EMV), puisque l'on vient de voir que l'algorithme $EM$ converge vers un maximum local. Notons alors que si la vraisemblance admet un unique point critique, il y a consistance de l'algorithme $EM$.

Dans le cas où l'espace des paramètres du mélange de lois est compact, [Red81] assure la convergence de l'EMV vers les paramètres de la vraie densité de mélange, à permutation près. Dans notre cas, l'ensemble des paramètres de la loi de mélange $\Theta_\kappa$ n'est pas compact. Mais c'est le cas lorsqu'une borne supérieure des paramètres des lois de Poisson entrant dans le mélange est connue. En notant $\Lambda \in \mathbb{R}_+$ cette borne, alors $\Theta_{\kappa,\Lambda} = \{(p,\lambda),\ p \in [0,1]^\kappa,\ \lambda \in [0,\Lambda]^\kappa\}$ est compact.

**Autres avantages** L'algorithme $EM$ est facile à implémenter, ce qui permet de l'amender. En effet, de nombreux algorithmes en découlent, par exemple si l'étape $M$ est difficile à évaluer, l'algorithme $GEM$, entre autres, (proposé en même temps que l'algorithme $EM$ par [DLR77]), consiste à choisir $\theta^{t+1}$ tel que $Q(\theta^{t+1}, \theta^t) \geq Q(\theta^t, \theta^t)$. Dans le cas où il est difficile de calculer l'espérance conditionnelle (étape $E$), l'algorithme $SAEM$, notamment, proposé par [DLM99], remplace l'étape $E$ par une approximation stochastique. De nombreux autres algorithmes existent, certains sont présentés dans [MK08].

**Inconvénients** L'algorithme $EM$ peut converger lentement, par exemple [MK08] propose des méthodes d'accélération de la convergence de l'algorithme $EM$. Cet algorithme peut être sensible au choix du vecteur des paramètres initiaux $\theta^1$, surtout dans le cas multivarié. [BCG03] propose des méthodes permettant de répondre à ce problème.

## 4.4 Applications

Après avoir établi l'étape $E$ et $M$ de l'algorithme $EM$ pour un mélange de $\kappa$ lois de Poisson, nous les mettons en pratique dans cette section. Tout d'abord sur des échantillons simulés, l'estimateur obtenu sera comparé avec la vraie densité, ou le vrai paramètre. Enfin nous appliquons l'algorithme sur les observations des biologistes. Cela permettra de déterminer les caractéristiques des sous-populations (le nombre moyen de

cellules cibles éliminée et la proportion dans la population totale) dont est composée la population de CTL.

### 4.4.1 Données simulées

Puisqu'ici nous connaissons la vraie valeur des paramètres du mélange de lois, un estimateur est évalué par sa distance en variation totale avec la vraie densité. Voici un rappel de sa définition pour des probabilités discrètes.

**Définition 4.4.1.** *La* **distance en variation totale** *de deux probabilités discrètes $P$ et $Q$ est*

$$d_{TV}(P,Q) = \frac{1}{2}\sum_{i=1}^{\infty}|P(i) - Q(i)|.$$

**Convergence de l'algorithme à échantillon fixé** Dans un premier algorithme, un échantillon de taille 10000 d'un mélange de 3 lois de Poisson de vecteur des paramètres aléatoires $\theta = (p, \lambda)$ est généré. Nous avons arrêté l'algorithme quand la différence entre $\theta^t$ et $\theta^{t+1}$ est inférieure à $\varepsilon = 0.01$.

Notons que l'algorithme s'arrête très vite, pour $t = 43$, et de plus, sa distance en variation totale avec la densité réelle est $||h_{\theta^{43}} - h_\theta||_{TV} = 0,08$. Remarquons que pour le vrai vecteur des paramètres $\theta = (0.44, 0.49, 0.07, 9.13, 6.32, 0.97)$, l'estimateur obtenu est $\theta^{43} = (0.43, 0.49, 0.08, 9.13, 6.46, 1.12)$ et l'erreur relative est $\frac{||\theta - \theta^{43}||_1}{||\theta||_1} = 0.02$.

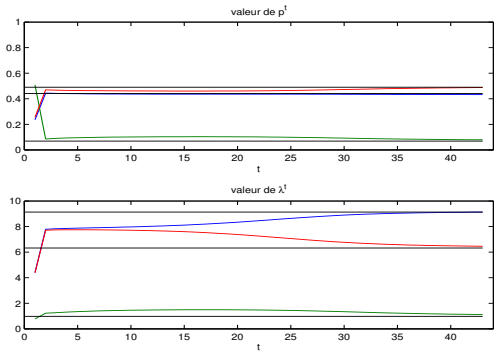



FIGURE 4.3 – **convergence de l'algortithme.** Evolution au cours des itérations des différentes valeurs prises par le vecteur des paramètres $\theta^t = (p^t, \lambda^t)$. Les barres horizontales noires sont les valeurs des paramètres de la vraie densité.

**Convergence de l'algorithme en augmentant la taille de l'échantillon** Afin d'étudier la convergence de l'algorithme $EM$ quand la taille de l'échantillon augmente, une densité de mélange $h^*$ est fixée. Puis, nous calculons la distance en variation totale entre $h^*$ les estimateurs obtenus sur des échantillons dont la taille varie entre 100 et 10000 (sur 100 simulations Monte-Carlo).

Dans un premier temps, les paramètres des lois de Poisson de $h^*$ sont relativement petits $\leq 10$ (voir la Figure 4.4 (A)). Puis dans un second temps, les paramètres des lois de Poisson sont grands $\leq 100$ (voir la Figure 4.4 (B)). Dans les deux cas, la convergence est assez rapide. Par contre, dans le deuxième cas l'estimateur de $h^*$ est moins bon que dans le premier cas. En effet, avec un échantillon de taille 10000 la distance est $\sim 1,2$ pour de grands paramètres, alors que pour de petits paramètres elle est autour de 0,2.

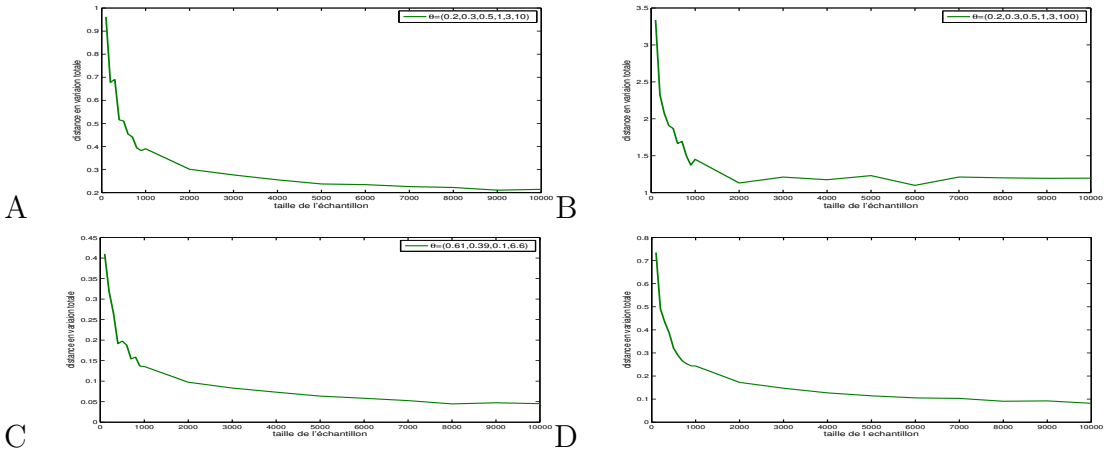



FIGURE 4.4 – **Convergence de l'algorithme** $EM$**.** Distance en variation total entre la vraie densité et la densité estimée par l'algorithme $EM$ (100 simulations Monte-Carlo), en fonction de la taille de l'échantillon (abscisse). (A) La valeur des paramètres des lois de Poisson est $\leq$10, (B) la valeur des paramètres des lois $\leq$100. (C) La vraie densité est un mélange de 2 lois de Poisson, (D) la vraie densité est un mélange de 10 lois de Poisson dont les moyennes sont inférieures à 10.

### 4.4.2 Données observées

Pour finir, nous mettons en application l'algorithme $EM$ sur les observations biologiques (Figure 4.1(C)). Pour cela, nous supposons que les observations sont issues d'un mélange de $\kappa$ lois de Poisson, sans avoir d'a priori sur $\kappa$ à par que $\kappa \in \mathbb{N}^*$ et $\kappa > 1$. L'algorithme $EM$ nous permettra de connaître les caractéristiques de ces sous-groupes.

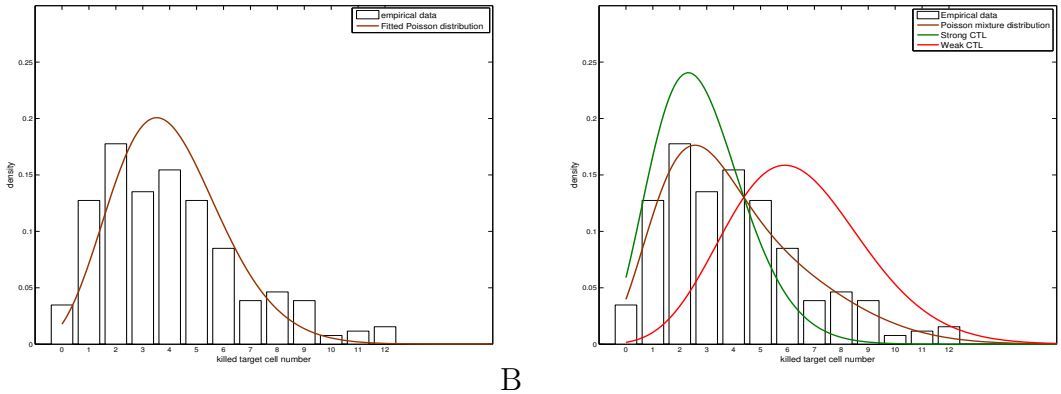
**Mélange de 2 lois de Poisson** Dans un premier temps, nous nous demandons si la population de CTL est divisée en 2 catégories : les "Weak CTL", qui détruisent en moyenne peu de cellules cibles, et les "Strong CTL", qui en éliminent plus en moyenne.

Nous avons appliqué l'algorithme $EM$ avec plusieurs valeurs de départ (différents $\theta^1$). Relativement à nos initialisations, l'algorithme est très stable, c'est-à-dire que les estimateurs sont toujours très proches de $\hat{\theta}$, où

$$\hat{\theta} = (\hat{p}, \hat{\lambda}), \quad \text{et } \hat{p} = (0.66, 0.34), \ \hat{\lambda} = (2.8, 6.4). \tag{4.6}$$

La distance en variation totale entre la densité de mélange obtenue avec cet estimateur et les données observées est de 0,15. Ce risque est de 0,35 en supposant que les observations sont issues d'une seule loi de Poisson. Ainsi, une densité de mélange a une meilleure adéquation aux observations qu'une densité de loi de Poisson. C'est assez naturel puisque l'estimateur $\hat{\theta}$ est obtenu sur l'espace des mélanges de 2 lois, qui est plus gros que celui de l'ensemble des lois de Poisson. Dans le chapitre suivant, nous étudions les propriété de ces estimateurs notamment en contrôlant le biais et la variance.

Une représentation graphique de la densité de mélange est donnée dans la Figure 4.5 (B), elle est superposée à l'histogramme des observations.

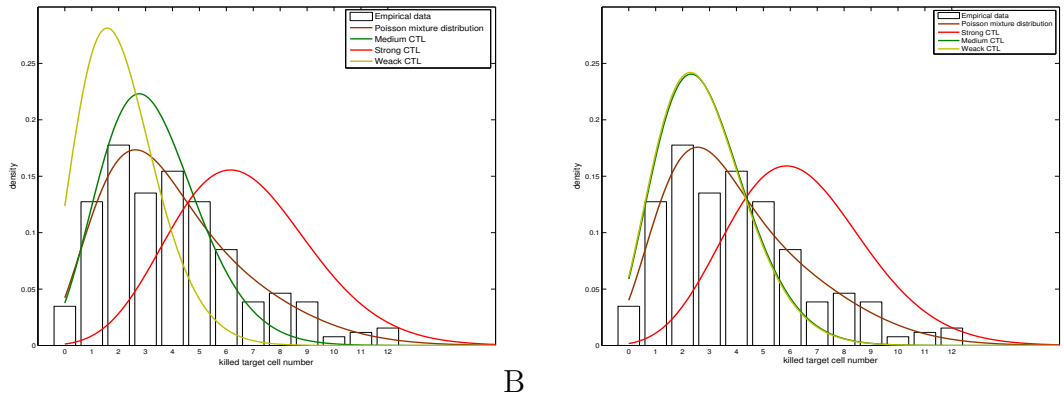



A B

FIGURE 4.5 – **Densité de 1 lois de Poisson et densité du mélange de 2 lois de Poisson.** Les données observées sont représentées par l'histogramme. (A) La courbe marron représente la densité de Poisson de paramètre la moyenne des observations : 4.03. (B) La courbe verte représente la densité de la loi de Poisson de faible moyenne, et la courbe rouge celle de grande moyenne. La courbe marron représente la densité du mélange des 2 lois (verte et rouge), obtenue grâce à l'algorithme $EM$.

**Mélange de 3 lois de Poisson** A présent, nous nous demandons si la population des CTL est divisée en 3 catégories.

Cette fois-ci, l'algorithme $EM$ donne des estimateurs très différents suivant la valeur initiale $\theta^1$, voir la Figure 4.6. Il existe des méthodes pour stabiliser l'estimateur obtenu par l'algorithme $EM$, par exemple dans [BCG03]. Mais nous ne développons pas dans ce sens car supposer que les observations sont issues d'un mélange de 3 lois n'améliore pas la prédiction. En effet, la distance en variation totale entre les densités de mélange de 3 lois de Poisson obtenues avec l'algorithme $EM$ et les valeurs observées est de 0.15. C'est la même distance que celle obtenue par la densité de mélange de 2 lois de Poisson. Nous n'étudions pas la possibilité que la population de CTL soit composée de plus de 4 sous-groupes.




A B

FIGURE 4.6 – **Mélange de 3 lois de Poisson.** Les données observées sont représentées par l'histogramme. Les courbes vertes, rouges et jaunes des Figures (A) et (B) sont les densités des lois de Poisson, dont le paramètre a été obtenu grâce à l'algorithme $EM$ pour différentes valeurs initiales, (A) $\theta^1 = (0.74, 0.1, 0.16, 3.4, 6.07, 1.48)$ et (B) $\theta^1 = (0.45, 0.05, 0.5, 6.25, 1.44, 0.24)$. Les courbes marrons représentent la densité de mélange des 3 courbes précédentes (A) $\theta^t = (0.55, 0.28, 0.17, 3.28, 6.66, 2.09)$ et $\theta^t = (0.34, 0.28, 0.38, 6.4, 2.83, 2.79)$.

**Test d'adéquation à un mélange de 2 lois de Poisson** En vue d'un critère pénalisé, que nous développons dans le chapitre suivant, nous avons choisi le modèle de mélange qui a sa distance en variation totale avec les observations la plus faible, et qui est le moins complexe. C'est-à-dire un modèle de mélange de 2 lois de Poisson. L'algorithme $EM$ permet d'estimer les caractéristiques de ces 2 sous-populations : (66%,2.8) pour le groupe des "Weak CTL" et (34%,6.4) pour le groupe des "strong CTL".

Afin de valider l'hypothèse d'hétérogénéité, nous avons procédé à un test du $\chi^2$ à une loi, comme dans la Section 4.1.2. Plus formellement nous souhaitions tester l'hypothèse

$H_0^{(2)}$ : les observations sont issues d'un mélange de 2 lois de Poisson de vecteur des paramètres $\hat{\theta}$ défini en (4.6).
La statistique de test a été définie par

$$T = \sum_{i=0}^{12} \frac{n \cdot (f_i - p_i)^2}{p_i},$$

où $n = 259$ comme dans la Sous-section 4.1.2, les $p_i = \mathbb{P}(\hat{p}X_{\hat{\lambda}} + (1-\hat{p})X_{\hat{\mu}} = i)$ pour tout $i = \{0, ..., 12\}$ et $p_{13} = \mathbb{P}(\hat{p}X_{\hat{\lambda}} + (1-\hat{p})X_{\hat{\mu}} \geq 13)$. Nous avons obtenu une statistique de test $T < 8,9$, qui est bien plus petite que le quantile d'ordre $1-\alpha$ d'une loi du $\chi_2$ à 9 degrés de liberté ($(13-1)-3$ car 3 paramètres sont estimés par l'algorithme $EM$) qui vaut 27,88. Nous acceptons donc l'hypothèse d'hétérogénéité $H_0^{(2)}$ avec une erreur de première espèce $\alpha = 0,001$.

## 4.5 Conclusions et perspectives

**Conclusion biologique** [5] Le test d'adéquation rejète l'hypothèse que la variabilité des observations (Figure 4.1 (C)) provienne de l'activité cytolytique d'une population homogène de CTL. Alors que ce même test accepte l'hypothèse que l'hétérogénéité des observations provienne de 2 sous-populations de CTL.

L'algorithme $EM$ estime les caractéristiques de ces 2 sous-populations :

- un sous-groupe de CTL représentant 66% de la population totale, qui détruit en moyenne 2,8 cellules cibles en 12 heures. Ils sont dit "Weak CTL" ;
- un second sous-groupe représentant 34% de la population et qui élimine en moyenne 6,4 cellules cibles. Ils sont dit "Strong CTL".

D'un point de vue biologique, il reste à déterminer d'où provient la différence entre ces deux sous-populations de CTL. En effet connaitre les mécanismes conduisant un CTL à devenir "strong CTL" plutôt que "Weak CTL" permettrait d'améliorer les fonctions de lyse d'un CTL et de développer une immunothérapie anti-tumorale.

Une première idée est de penser que cette différence est d'ordre génétique. Cette hypothèse est rejetée, car par des expériences supplémentaires, les biologistes ont retrouvé la même l'hétérogénéité de la capacité cytolytique des CTL fils [6] d'un "Strong CTL". Une étude approfondie des temps de killing des 259 CTL a mis en évidence que les CTL qui tuent plus de 5 cibles, sont ceux qui, pendant les 12 heures d'observation, ont au moins une phase où ils tuent 3 cibles ou plus en moins de 3 heures. Cette capacité à "tuer vite" serait un premier paramètre permettant d'identifier un "strong CTL" d'un "weak CTL".

**Conclusion mathématique** L'algorithme $EM$ est souvent utilisé pour des mélanges gaussiens. Dans ce chapitre, nous l'avons mis en place pour un mélange de $\kappa$ lois de Poisson, où $\kappa \in \mathbb{N}^*$.

Remarquons que, sur les données simulées, pour $n = 200$, si $\kappa = 3$, la distance en variation totale entre la densité estimée par l'algorithme $EM$ et la vraie densité est deux fois plus grande que si $\kappa = 2$ (voir Figure 4.4 (A) et (C)). Cela peut expliquer l'instabilité de l'estimateur obtenu par l'algorithme $EM$ pour les 259 observations de la Figure 4.1 (C), si nous supposons qu'elles sont issues d'un mélange de 3 lois. Ainsi, la dimension maximale des modèles de mélanges qu'il est possible de considérer dépend de la taille de l'échantillon.

Comme souligné dans la sous-section consacrée aux données simulées, si au moins un des paramètres des lois de Poisson est grand, l'estimateur obtenu par l'algorithme $EM$ n'est pas très bon (Figure 4.4 (B)). Notons que si $\kappa = 10$ mais que les moyennes des lois intervenant dans le mélange sont $\leq 10$, alors l'algorithme $EM$ fournit de bons

5. Ces travaux font l'objet d'un article soumis, écrit en collaboration avec Zilton Vasconcelos, Sabina Müler, Delphine Guipouy, Yu Wong, Sébastien Gadat, Salvatore Valitutti et Loïc Dupré

6. CTL obtenus après la division cellulaire d'un "Strong CTL"

estimateurs (Figure 4.4 (D)). Ainsi, pour avoir une bonne estimation de la densité, il est important d'avoir une taille maximale $\Lambda$ des paramètres des lois pas trop grande. On peut expliquer l'importance du choix de $\Lambda$ par le fait plus le paramètre de la loi de Poisson est grand plus la variance est grande. Ainsi sur un échantillon de mélange de lois, même si la proportion de la loi de grand paramètre est grande, l'échantillon sera éparpillé ce qui rend l'estimation difficile.

**Perspective 1** L'algorithme $EM$ permet d'estimer les caractéristiques de $\kappa$ groupes composant la population totale des CTL. Cependant, il ne permet pas de déterminer le nombre $\kappa$ de sous-populations. Le test d'adéquation du $\chi^2$ informe sur la probabilité ou non que les observations soient issues d'un modèle (hypothèse $H_0$). Mais il n'informe pas sur la qualité de l'estimateur puisqu'il ne compare pas les modèles entre eux.

Pour comparer les modèles, nous avons calculé la distance en variation totale entre les observations et les densités estimées. Puis nous avons choisi le modèle le plus "simple" parmi les modèles atteignant la distance minimale. Nous avons fait de manière informelle de la sélection de modèle. Plus formellement il convient de définir un critère permettant de choisir le nombre $\kappa$ de sous-groupes qui maximise la log-vraisemblance et favorise le modèle le plus simple. Cela permettra de montrer la consistance de modèle. Ce point est développé dans le chapitre suivant.

**Perspective 2** On remarque (Figures 4.5 et 4.6) que les densités obtenues par l'algorithme $EM$ pour un mélange de 2 ou 3 lois sont très proches (elles ont la même distance en variation totale avec les observations). Notamment, la densité de mélange de la Figure 4.6 (B) a ses paramètres très proches de ceux du mélange de 2 lois de Poisson. En effet, nous retrouvons la loi de Poisson de moyenne de 6.4 et de proportion de 34% dans ces deux densités. Les 2 sous-groupes restant du mélange de 3 lois ont quasiment la même moyenne 2,79 et 2,83 qui correspond à la moyenne du second sous-groupe du mélange de 2 lois. Il semble alors difficile de différencier ces 2 lois.

Dans un cadre plus simple, cela revient à résoudre le problème de test suivant.

Soit $X$ un échantillon de taille $n$, notons $f$ la densité de la loi dont est issue $X$. On souhaite tester l'hypothèse, pour $\mathcal{F}$ l'ensemble des lois de Poisson (défini en (4.1)),

$H_0 \ : \ f \in \mathcal{F}$,

contre

$H_{\varepsilon_n,\delta_n} \ : \ f \in \{p_n f_{\lambda_n} + (1-p_n) f_{\mu_n}, \text{ où } f_{\lambda_n}, f_{\mu_n} \in \mathcal{F}, \ |\lambda_n - \mu_n| \geq \delta_n, \ p_n \wedge (1-p_n) > \varepsilon_n\}$

Pour quelles valeurs de $\varepsilon_n > 0$ et $\delta_n > 0$ telles que $p_n \wedge (1-p_n) > \varepsilon_n$, $|\lambda_n - \mu_n| > \delta_n$, est-il possible de différencier $H_0$ de $H_{\varepsilon_n,\delta_n}$ ?

Pour répondre à cette question, il faut déterminer un test (test du rapport de vraisemblance, autres tests ?), définir un risque... C'est un thème de recherche très actuel, [LMMR13 ; ACBL12 ; AB+10 ; DJ04] évoquent les contextes gaussiens.

**Perspective 3** L'étude que nous avons réalisée utilise uniquement les données à 12 heures alors que nous avons à notre disposition le nombre de cellules cibles éliminées toutes les 2h pendant 12h (voir la Figure 4.1 (B)). Réaliser une étude cinétique de l'activité cytolytique permettrait de définir plus précisément ce qui distingue les 2 sous-types de CTL.

Au vu de l'étude précédente, les données seraient issues d'un mélange de 2 processus de Poisson. Cependant, nous n'avons pas de raison de supposer que le temps mis par un CTL pour détruire une cellule cible est constant au cours du temps. Ainsi, nous privilégions une modélisation par des processus de Poisson inhomogènes. C'est-à-dire que nous supposons que les paramètres des processus $\lambda$ et $\mu$ ne sont pas linéaires par rapport au temps, mais des fonctions appartenant à $\mathcal{C}^s$, où $s > \frac{1}{2}$. Un modèle adéquat serait alors

$$p\mathcal{P}(\lambda(t)) + (1-p)\mathcal{P}(\mu(t)),$$

où $p = 66\%$ est estimée par l'algorithme $EM$. Il reste à déterminer les fonctions $\lambda$ et $\mu$.

Nous sommes confrontés au problème d'estimation non paramétrique des fonctions $\lambda$ et $\mu \in \mathcal{C}^s$, où $s > \frac{1}{2}$. Le théorème de Girsanov pour les processus de Poisson permet de donner un sens à la vraisemblance des observations pour différentes fonction de $\mathcal{C}^s$. Il semble donc envisageable de développer une stratégie $EM$ fonctionnelle non paramétrique pour estimer les intensités $\lambda$ et $\mu$.

# Chapitre 5

# Sélection de modèle non asymptotique pour des mélanges poissoniens

A nombre de lois fixé dans le mélange, c'est-à-dire à taille de modèle fixé, l'algorithme $EM$ est traditionnellement utilisé pour estimer les paramètres d'un mélange. Mais le nombre de lois dans un mélange est rarement disponible, comme dans le modèle biologique présenté au chapitre précédent. Ainsi, la question du meilleur estimateur, ou du meilleur modèle, se pose. En suivant le cadre donné dans [Mas07], un critère pénalisé non asymptotique est proposé pour sélectionner le nombre de lois de Poisson entrant dans un mélange. Pour obtenir ce critère de pénalité, il est nécessaire de majorer l'entropie à crochet pour l'ensemble des mélanges finis et infinis de lois de Poisson.

## 5.1 Introduction

L'étude menée au chapitre précédent remet en question l'homogénéité de la population des CTL au travers de leur activité cytolytique. En effet, le modèle le plus en adéquation avec les observations du nombre de cibles éliminées par un CTL est un modèle de mélange de lois de Poisson. Cela sous-entend que la population de CTL est divisée en plusieurs sous-populations. Elles sont caractérisées par le nombre moyen de cibles éliminées en 12 heures et leur proportion dans la population totale.

A nombre de composante du mélange fixé, l'algorithme $EM$, mis en place pour les mélanges poissoniens au chapitre précédent, estime les paramètres d'un mélange. Par contre, nous n'avons pas d'information sur la qualité du modèle.

Si au chapitre précédent nous avons choisi le modèle d'un mélange de 2 lois de Poisson, nous n'avons pas démontré la consistance de notre approche dans ce cadre. C'est ce que nous proposons de faire dans ce chapitre grâce à l'approche traditionnelle de la sélection de modèle.

La sélection de modèle trouve son origine dans les années 70 avec les travaux d'Akaike et Schwartz [Aka73 ; Sch78]. Dans ces deux cas, la méthode consiste à minimiser le contraste empirique pénalisé sur un ensemble de modèles fixés. Les critères AIC (Akaike's Information Criterion) et BIC (Bayesian Information Criterion) sont de type log-vraisemblance, dont les pénalités sont proportionnelles aux nombres de paramètres.

Pour ces deux critères, la consistance n'est pas assurée. En effet, le critère AIC a tendance à sur-ajuster, il a été conçu pour être efficace (choisir le modèle réalisant le meilleur compromis biais variance). Le critère BIC est consistant s'il est possible de garantir que le vrai modèle fait partie de l'ensemble des modèles fixés, ce qui n'est pas toujours le cas, sinon il a tendance à sous-ajuster [BA02 ; MP00].

Ces deux critères reposent sur une approximation asymptotique, c'est-à-dire que le modèle choisi par le critère sera d'autant meilleur que le nombre d'observations augmente, ce qui nécessite un ensemble de modèles fixes. Mais dans de nombreux problèmes, il est préférable d'augmenter la taille des modèles avec la taille de l'échantillon, comme dans notre cas. En effet, dans les modèles de mélange il est inutile de supposer que les observations sont issues d'un mélange de plus de $n$ lois, où $n$ est la taille de l'échantillon.

Pour pallier ce problème, [Mas07] propose de faire de la sélection de modèle non asymptotique. Ce qui est entendu par *non asymptotique* est que la taille des modèles peut croître avec la taille de l'échantillon.

Dans ce chapitre, nous proposons un critère pénalisé non asymptotique définissant le nombre de lois et les paramètres du meilleur mélange poissonnien pour un échantillon donné.

Dans un premier temps, nous rappelons le principe de la sélection de modèle, puis nous énonçons le résultat non asymptotique de [Mas07]. Dans un deuxième temps, nous calculons une borne supérieure de l'entropie à crochet de l'ensemble des modèles de mélange fini et infini de lois de Poisson. Cela permettra de définir dans un troisième temps le critère pénalisé à minimiser pour sélectionner le meilleur modèle. Nous terminons par une application des résultats aux observations biologiques.

## 5.2 Sélection de modèle

Nous nous intéressons à l'estimation de densité de mélange de lois de Poisson, qui est absolument continue par rapport à la mesure de comptage. Nous nous plaçons dans ce cadre pour le rappel sur la sélection de modèle.

### 5.2.1 Estimation par minimisation du critère pénalisé

Notons $h^*$ la vraie densité dont sont issues les observations $x = (x_1, ..., x_n)$, $\mathcal{H}$ un ensemble de densités absolument continues par rapport à la mesure de comptage et $h$ un élément de $\mathcal{H}$ .

#### Estimation par minimisation du contraste

**Information de Kulback-Leibler** Dans notre contexte de mélange, nous cherchons à utiliser la log-vraisemblance comme critère d'attache aux données. La fonction de contraste qui est naturellement associée pour mesurer l'erreur d'estimation est donc $\gamma(t, .) = -\ln(t(.))$. Considérons la fonction perte correspondante : l'information de Kullback-Leibler, définie pour $s$ et $t$ deux densités absolument continues par rapport à la mesure de comptage,

$$KL(s, t) = \sum_{i=0}^{\infty} s(i) \ln \frac{s(i)}{t(i)}.$$

La fonction perte permet de comparer les estimateurs entre eux. En effet, pour deux estimateurs $h_n^{(1)}$ et $h_n^{(2)}$, celui qui a sa fonction de perte par rapport à $h^*$ la plus petite est préférable.

**Contraste empirique** On note que minimiser $KL(h^*, h)$, pour $h \in \mathcal{H}$, revient à minimiser $\sum_{i=0}^{\infty} -h^*(i) \ln(h(i))$. Bien que $h_o = \text{argmin}_{h \in \mathcal{H}} \sum_{i=0}^{\infty} -h^*(i) \ln(h(i))$ soit la meilleure densité sur $\mathcal{H}$ pour approcher $h^*$, ce n'est pas un estimateur de $h^*$ car $h_o$ n'est pas mesurable aux données et dépend de $h^*$, qui est inconnue. Afin de construire un estimateur de $h^*$, définissons le contraste empirique :

$$\gamma_n(h) = -\frac{1}{n} \sum_{i=0}^{n} \ln(h(x_i)). \tag{5.1}$$

Ainsi $\hat{h} = \text{argmin}_{h \in \mathcal{H}} \gamma_n(h)$ est un estimateur de $h^*$.

**Le paradigme du choix de modèle** Une des difficultés soulevées par l'estimateur obtenu $\hat{h}$ est de choisir le bon ensemble $\mathcal{H}$. En effet, si $\mathcal{H}$ est trop "petit", il se peut que $h^*$ soit très loin de l'ensemble $\mathcal{H}$, et l'estimateur peut ne pas converger vers 0 en raison du biais imcompressible ainsi créé. Par contre, si $\mathcal{H}$ est trop "gros", par exemple dans notre cas si $\mathcal{H}$ est l'ensemble de toutes les densités absolument continues pas rapport à la mesure de comptage, la consistance de l'estimateur n'est pas garantie, ou le taux de convergence peut ne pas être optimal [Bah58 ; BM93].

**Sélection de modèle par pénalisation**

Notons $(\mathcal{H}_m)_{m\in\mathcal{M}}$ un ensemble dénombrable de modèles paramétriques, et pour chaque modèle, $\hat{h}_m$ l'estimateur du maximum de vraisemblance dans cet ensemble.

**Oracle** Le meilleur estimateur parmi l'ensemble des estimateurs du maximum de vraisemblance $\left(\hat{h}_m\right)_{m\in\mathcal{M}}$ serait celui qui minimise le risque, c'est-à-dire $\hat{h}_{m(h^*)}$ où

$$m(h^*) = \underset{m\in\mathcal{M}}{\operatorname{argmin}}\, \mathbb{E}\left(KL(h^*, \hat{h}_m)\right).$$

La fonction $\hat{h}_{m(h^*)}$ est appelée oracle, mais n'est toujours pas un estimateur car il dépend de la vraie densité, on ne peut donc pas le calculer. Le risque de l'oracle sert de marqueur pour la suite.

**Inégalité oracle** Le but est alors de choisir $\hat{m}$ aussi proche que possible de $m(h^*)$, dans le sens suivant : le risque de l'estimateur $\hat{h}_{\hat{m}}$ doit être le plus proche possible du risque de $\hat{h}_{m(h^*)}$. Ainsi, $\hat{m} \in \mathcal{M}$ doit être tel que $\hat{h}_{\hat{m}}$ vérifie l'inégalité oracle : pour $C \in \mathbb{R}_+^*$

$$\mathbb{E}\left(KL\left(h^*, \hat{h}_{\hat{m}}\right)\right) \leq C\mathbb{E}\left(KL\left(h^*, \hat{h}_{m(h^*)}\right)\right) + O\left(\frac{1}{n}\right).$$

Cela garantit que le risque est proche du risque minimal à constante près et avec un terme résiduel d'ordre $\frac{1}{n}$.

**Fonction de pénalité** Dans l'inégalité oracle, les risques ne peuvent pas être calculés car ils dépendent de la vraie densité inconnue. Il faut alors définir un critère de sélection pour remplacer le terme $\mathbb{E}\left(KL\left(h^*, \hat{h}_{\hat{m}}\right)\right) - \mathbb{E}\left(KL\left(h^*, \hat{h}_{m(h^*)}\right)\right)$, et une pénalité (pour définir $O\left(\frac{1}{n}\right)$) afin de sélectionner le meilleur $\hat{h}_{\hat{m}}$ parmi les $\left(\hat{h}_m\right)_{m\in\mathcal{M}}$.
On procède à la sélection de modèle via une pénalisation

$$pen \;:\; \mathcal{M} \to \mathbb{R}_+ : m \mapsto pen(m).$$

**Critère pénalisé** Ainsi, le meilleur estimateur de $m(h^*)$ sera celui qui minimise le critère pénalisé :

$$crit\,(m) = \gamma_n\left(\hat{h}_m\right) + pen(m).$$

c'est-à-dire $(\hat{m}) = \operatorname{argmin}_{m\in\mathcal{M}} crit(m) = \operatorname{argmin}_{m\in\mathcal{M}} \gamma_n\left(\hat{h}_m\right) + pen(m)$.

Pour définir la fonction de pénalité dans un cadre non asymptotique, [Mas07] utilise des inégalités de concentration.

### 5.2.2 Critère pénalisé non asymptotique

Dans cette sous-partie nous rappelons le Théorème de [Mas07, section 7.4], qui propose une fonction de pénalité ainsi que l'inégalité oracle associée.

Soit $(\mathcal{H}_m)_{m\in\mathcal{M}}$ un ensemble dénombrable de modèles tels que, pour chaque $m$, les éléments de $\mathcal{H}_m$ sont des densités absolument continues par rapport à la mesure de comptage. Deux hypothèses sont faites sur les modèles $\mathcal{H}_m$.

Nous supposerons que chaque ensemble n'est pas trop "gros". Autrement dit nous considérons que l'entropie à crochet, notée $H_{[.]}(\varepsilon, \mathcal{H}_m, d)$ pour $\varepsilon > 0$ et $d$ une distance sur $\mathcal{H}_m$, est intégrable en 0. Nous renvoyons à la section suivante pour une définition précise de l'entropie à crochet. Pour tout $m \in \mathcal{M}$, nous considérons une fonction qui vérifie les propriétés suivantes :

1. $\Psi_m$ est croissante sur $\mathbb{R}_+$ ;
2. l'application $\xi \mapsto \frac{\Psi_m(\xi)}{\xi}$ est décroissante sur $\mathbb{R}_+$ ;
3. pour $\xi \in \mathbb{R}_+$, $\int_0^\xi \sqrt{H_{[.]}(x, \mathcal{H}_m, d_H)}\,\mathrm{d}x \leq \Psi_m(\xi)$.

**Théorème 5.2.1** ([Mas07])**.** *Soient $x_1, ..., x_n$ des variables aléatoires i.i.d. de densité inconnue $h^*$ absolument continue par rapport à la mesure de comptage. Soit $(\mathcal{H}_m)_{m\in\mathcal{M}}$ un ensemble dénombrable de modèles vérifiant les deux hypothèses précédentes et $\hat{h}_m$ l'estimateur de maximum de vraisemblance du modèle $\mathcal{H}_m$. Soit $(\rho_m)_{m\in\mathcal{M}}$ un famille de nombres positifs ou nuls telle que*

$$\sum_{m\in\mathcal{M}} e^{-\rho_m} = \Upsilon < \infty.$$

*Pour tout $m \in \mathcal{M}$, considérons la fonction $\Psi_m$ qui vérifie les propriétés 1, 2 et 3, et notons $\xi_m$ l'unique solution positive de l'équation*

$$\Psi_m(\xi) = \sqrt{n}\xi^2.$$

*Soit la fonction de pénalité $pen : \mathcal{M} \to \mathbb{R}_+$, et le critère pénalisé*

$$crit(m) = \gamma_n(\hat{h}_m) + pen(m).$$

*Alors, s'il existe une constante universelle $c > 0$ telles que, pour tout $m \in \mathcal{M}$*

$$pen(m) \geq c\left(\xi_m^2 + \frac{\rho_m}{n}\right),$$

*et s'il existe une variable aléatoire $\hat{m}$ qui minimise crit sur $\mathcal{M}$. Alors pour toute densité*

$h$,

$$\mathbb{E}\left[d_H^2(h,\hat{h}_{\hat{m}})\right] \leq C\left[\inf_{m\in\mathcal{M}}\left\{KL(h,\mathcal{H}_m)+pen(m)\right\}+\frac{\Upsilon}{n}\right],$$

*où* $C>0$ *et* $KL(h,\mathcal{H}_m)=\inf_{t\in\mathcal{H}_m} KL(h,t)$ *pour tout* $m\in\mathcal{M}$.

**Sélection de modèle non asymptotique et mélange de lois** Ce théorème s'applique dans de nombreux domaines. Dans le cadre de mélanges de lois, [Mey13] propose un critère pour un mélange de $K$, fixé, régressions gaussiennes. Ce critère sélectionne sur la norme $L^1$ des moyennes des gaussiennes. [MB11] définit un critère pour des mélanges gaussiens multivariés particuliers qui sélectionne sur le nombre de composantes dans le mélange, ainsi que sur la dimension des variables gaussiennes multivariées pertinentes. [MRM13] affirme que l'estimateur pénalisé obtenu pour des mélanges de lois gaussiennes univariées est un estimateur minimax $\beta$ adaptatif.

### 5.2.3 Notation

A présent, soit $x=(x_1,...,x_n)$, $x_i\in\mathbb{N}$, les réalisations d'un $n$ échantillon, que nous supposons être issu d'une densité $h^*$ de mélange fini de lois de Poisson. Mais nous n'avons pas d'a priori sur le nombre de lois qui compose ce mélange. Ainsi $h^*$ appartient à l'ensemble suivant

$$\mathcal{H}=\left\{h_\kappa(.)=\sum_{k=0}^{\kappa} p_k f_{\lambda_k}(.),\quad \forall\kappa\geq 1,\ \theta_\kappa=(p_1,...,p_\kappa,\lambda_1,..,\lambda_\kappa)\in\Theta_\kappa\right\},$$

où, en notant $\mathfrak{S}_{\kappa-1}$ le simplexe de dimension $\kappa-1$,

$$\Theta_\kappa=\left\{\theta=(\underbrace{p_1,...,p_\kappa}_{p},\underbrace{\lambda_1,...,\lambda_\kappa}_{\lambda});\ p\in\mathfrak{S}_{\kappa-1}\text{ et }\ \lambda\in\mathbb{R}_+^\kappa\right\}.$$

Il est difficile de faire de l'inférence statistique sur $\mathcal{H}$, car cet ensemble est très gros. De plus, la Figure 4.4 (B) a démontré que la qualité de la convergence de l'estimateur obtenu par l'algorithme $EM$ dépend de la taille maximale des paramètres des lois de Poisson. C'est pour cela que nous allons borner les paramètres des lois de Poisson par $\Lambda\in\mathbb{N}^*$. Puis nous considérons des ensembles dont les modèles ont la même dimension, c'est-à-dire le même nombre de lois de Poisson. Soient $\kappa,\Lambda\in\mathbb{N}^*$, considérons l'ensemble des mélanges de $\kappa$ lois, dont les paramètres sont compris entre 0 et $\Lambda$

$$\mathcal{H}_{\kappa,\Lambda}=\left\{\sum_{k=1}^{\kappa} p_k f_{\lambda_k}(.),\quad (p,\lambda)\in\Theta_\kappa\text{ et }\forall k,\ \lambda_k\leq\Lambda\right\},\tag{5.2}$$

On note que la dimension de $\mathcal{H}_{\kappa,\Lambda}$ est $2\kappa-1$ et que $\mathcal{H}=\cup_{\kappa,\Lambda\in\mathbb{N}^*}\mathcal{H}_{\kappa,\Lambda}$.

Ainsi, pour chaque $\kappa, \Lambda \in \mathbb{N}^*$, nous allons sélectionner le meilleur modèle parmi $\mathcal{H}_{\kappa,\Lambda}$.

**Variables de selections** Nous allons sélectionner sur les deux paramètres suivants.

Le nombre $\kappa$ de lois dans le modèle, ce qui est le but premier car cela détermine le nombre de sous-groupes dans la population totale.

Le nombre $\Lambda$, ce qui est moins naturel à priori, mais pourtant essentiel car la qualité de l'estimateur en dépend. De plus, nous n'avons pas d'a priori sur cette valeur, excepté qu'il est certainement inférieur au maximum des observations.

Cela imposera en particulier que si $\kappa$ est grand, pour avoir un bon estimateur, il faut que les moyennes des lois ne soient pas trop grandes.

## 5.3 Borne supérieure de l'entropie à crochet

L'entropie à crochet intervenant dans le Théorème 5.2.1, est un moyen de mesurer la taille d'un ensemble de modèles. La valeur exacte de l'entropie n'est pas toujours explicite ou calculable. Dans cette section, nous proposons une borne supérieure de l'entropie à crochet de l'ensemble $\mathcal{H}_{\kappa,\Lambda}$ défini en (5.2).

Commençons par donner une définition de l'entropie à crochet.

**Définition 5.3.1.** *Soient $\varepsilon > 0$ et $\mathcal{F}$ un ensemble de fonctions ordonné et muni d'une distance $d$.*

- *Pour deux fonctions $l$ et $u$, le* **crochet** *$[l, u]$ est l'ensemble des fonctions $f$ de $\mathcal{F}$ tel que $l \leq f \leq u$.*
- *Un $\varepsilon-$***crochet** *pour la distance $d$ est un crochet $[l, u]$ vérifiant $d(l, u) \leq \varepsilon$.*
- *On note $N_{[.]}(\varepsilon, \mathcal{F}, d)$ le nombre minimum de $\varepsilon-$ crochets nécessaires pour recouvrir $\mathcal{F}$ (les bornes des crochets ne sont pas forcément des points de $\mathcal{F}$).*
- *On appelle* **entropie à crochet** *la quantité $\ln\left(N_{[.]}(\varepsilon, \mathcal{F}, d)\right)$, notée $H_{[.]}(\varepsilon, \mathcal{F}, d)$.*

La taille de l'ensemble $\mathcal{H}_{\kappa,\Lambda}$ dépend de $\kappa$, le nombre maximal de lois constituant le mélange, et de $\Lambda$, la valeur maximale des paramètres des lois. Pour mesurer $H_{[.]}(\varepsilon, \mathcal{H}_{\kappa,\Lambda}, d_h)$, il faut d'abord mesurer $N_{[.]}(\varepsilon, \mathcal{F}_\Lambda, d_H)$ de toutes les densités de loi de Poisson de paramètre inférieur à $\Lambda$. Plus formellement, nous définissons

$$\mathcal{F}_\Lambda = \{f_\lambda(.), 0 \leq \lambda \leq \Lambda\}, \tag{5.3}$$

où $f_\lambda$ est la densité de la loi de Poisson de paramètre $\lambda$.

Ensuite, il faut calculer la taille du mélange de ces lois. Pour cela, il y a deux méthodes. La première, utilisée dans l'article [GW00], consiste à mesurer la taille du

simplexe $\mathfrak{S}_{\kappa-1}$. Une seconde méthode consiste à contrôler la taille des mélanges par la taille de $\Lambda$, comme le suggère l'article [GV01]. Cette dernière méthode, qui exploite les propriétés de régularité en $\lambda$ de la loi de Poisson, sera avantageuse pour majorer l'entropie à crochet quand le nombre de lois composant le mélange, $\kappa$, est grand.

Nous utilisons deux distances pour contrôler l'entropie à crochet. Pour calculer le nombre minimum de $\varepsilon-$crochet des ensembles $\mathcal{F}_\Lambda$ et de $\mathcal{H}_{\kappa,\Lambda}$ par la première méthode, nous utilisons la distance d'Hellinger entre deux fonctions.

**Définition 5.3.2.** *La* **distance d'Hellinger** *entre deux probabilités $P$ et $Q$ à densités $f$ et $g$ par rapport à une mesure dominante $\nu$, est définie par :*

$$d_H(f,g)^2 = \frac{1}{2}\int_{\mathbb{R}} \left(\sqrt{f(x)} - \sqrt{g(x)}\right)^2 \mathrm{d}\nu(x).$$

Pour le calcul de l'entropie à crochet de $\mathcal{H}_{\kappa,\Lambda}$ par la seconde méthode, nous utilisons la distance en variation totale, dont la définition est rappelée en 4.4.1.

Ces deux distances sont reliées par la relation suivante : si et $f$ et $g$ sont deux fonctions positives ou nulles telles que, pour tout $x \in \mathbb{R}$, $g(x) \leq f(x)$, alors

$$d_H(f,g) \leq \sqrt{d_{TV}(f,g)} \text{ et } d_{TV}(f,g) \leq 2d_H(f,g)\sqrt{\int f}.$$

Dans la suite, $\lesssim$ désigne une inégalité à constante multiplicative universelle près.

### 5.3.1 L'entropie de $\mathcal{F}_\Lambda$

**Théorème 5.3.1.** *Soient $\Lambda \in \mathbb{N}^*$ et $\varepsilon \in ]0,1[$*

$$N_{[.]}(\varepsilon, \mathcal{F}_\Lambda, d_H) \leq \frac{4\Lambda}{\varepsilon^2}.$$

*Et*

$$H_{[.]}(\varepsilon, \mathcal{F}_\Lambda, d_H) \lesssim \ln(1/\varepsilon) + \ln(\Lambda).$$

La preuve de ce théorème nécessite différents lemmes, dont les démonstrations sont placées après la preuve principale.

*Démonstration.* Soit

$$\eta = \frac{\varepsilon^2}{2\sqrt{2\Lambda}+1}.$$

Notons $\xi = \frac{\eta}{\sqrt{2}} > 0$. Le Lemme 5.3.2 affirme que la famille $(l_i, u_i)_{i \in \left\{0,\ldots,\left\lfloor\sqrt{\Lambda/(2\xi^2)}\right\rfloor - 1\right\}}$ est une famille de $\delta-$crochets recouvrant $\mathcal{F}_\Lambda$ , où $l_i = \frac{1}{1+\delta} f_{\lambda_i}$, $u_i = (1+\delta) f_{\lambda_{i+1}}$, $\lambda_i = 2\xi^2 i^2$ et $\delta = e^{2\sqrt{2\Lambda}\xi + 2\xi^2} - 1$.

Vérifions que $d_H^2(l_i, u_i) \leq \varepsilon^2$. Remarquons d'abord que pour deux densités de probabilités $f$ et $g$,

$$d_H^2((1+\delta)f, (1+\delta)^{-1}g) \leq \delta^2 + d_H^2(f, g).$$

En effet

$$\begin{aligned} d_H^2((1+\delta)f, (1+\delta)^{-1}g) &= \frac{1}{2}\int \left(\sqrt{(1+\delta)f(x)} - \sqrt{\frac{1}{1+\delta}g(x)}\right)^2 \mathrm{d}x \\ &= \frac{1}{2}\left(1+\delta+\frac{1}{1+\delta} - 2\int \sqrt{f(x)g(x)}\,\mathrm{d}x\right) \\ &= \frac{1}{2}\left(\delta+\frac{1}{1+\delta}-1\right) + 1 - \int \sqrt{f(x)g(x)}\,\mathrm{d}x \\ &\leq \delta^2 + d_H^2(f, g). \end{aligned}$$

Ainsi, par le Lemme 5.3.1, il vient $d_H^2(l_i, u_i) \leq \delta^2 + d_H^2(f_{\lambda_i}, f_{\lambda_{i+1}}) \leq \delta^2 + \xi^2$.

Soit l'inégalité :

$$\forall\ x \in [0, 0.8], \quad (e^x - 1)^2 \leq 2x.$$

Comme $\xi \leq \frac{1}{4\sqrt{2\Lambda}+2}$ alors $0 \leq 2\sqrt{2\Lambda}\xi + 2\xi^2 \leq 0.8$, en appliquant l'inégalité précédente pour $x = 2\sqrt{2\Lambda}\xi + 2\xi^2$, nous obtenons

$$\begin{aligned} \xi^2 + \delta^2 &\leq \xi^2 + 4\sqrt{2\Lambda}\xi + 4\xi^2 \\ &\leq 2\sqrt{2\Lambda}\eta + 3\eta^2 \\ &\leq (1 + 2\sqrt{2\Lambda})\eta \\ &\leq \varepsilon^2. \end{aligned}$$

Ainsi, $N_{[.]}(\varepsilon, \mathcal{F}_\Lambda, d_H) = \left\lfloor \sqrt{\frac{\Lambda}{2\xi^2}} \right\rfloor \leq \sqrt{\frac{\Lambda}{\eta^2}} = \frac{2\sqrt{2}\Lambda + \sqrt{\Lambda}}{\varepsilon^2} \leq \frac{4\Lambda}{\varepsilon^2}$.

□

**Lemme 5.3.1.** *Soient* $\varepsilon \in ]0, 1[$, $\lambda_i = 2\varepsilon^2 i^2$, *pour* $i \in \left\{0, ..., \left\lfloor \sqrt{\frac{\Lambda}{2\varepsilon^2}} \right\rfloor \right\}$. *Alors, pour tout* $i \in \left\{0, ..., \left\lfloor \sqrt{\frac{\Lambda}{2\varepsilon^2}} \right\rfloor - 1 \right\}$

$$d_H^2(f_{\lambda_i}, f_{\lambda_{i+1}}) \leq \varepsilon^2.$$

*Démonstration.* Soit $i \in \left\{0, ..., \left\lfloor \sqrt{\frac{\Lambda}{2\varepsilon^2}} \right\rfloor - 1\right\}$,

$$\begin{aligned} d_H^2(f_{\lambda_i}, f_{\lambda_{i+1}}) &= \frac{1}{2}\sum_{k=1}^{\infty}\left(\sqrt{e^{-\lambda_i}\frac{\lambda_i^k}{k!}} - \sqrt{e^{-\lambda_{i+1}}\frac{\lambda_{i+1}^k}{k!}}\right)^2 \\ &= 1 - \sum_{k=1}^{\infty} e^{-\frac{\lambda_i+\lambda_{i+1}}{2}}\frac{\left(\sqrt{\lambda_i\lambda_{i+1}}\right)^k}{k!} \\ &= 1 - e^{-\frac{\lambda_i+\lambda_{i+1}}{2}} e^{\sqrt{\lambda_i\lambda_{i+1}}} \\ &= 1 - e^{-\frac{1}{2}\left(\sqrt{\lambda_i} - \sqrt{\lambda_{i+1}}\right)^2} \\ &= 1 - e^{-\frac{1}{2}\left(\sqrt{2}\varepsilon\right)^2} \\ &= 1 - e^{-\varepsilon^2}. \end{aligned}$$

On conclut en rappelant que $e^{-\varepsilon^2} = 1 - \varepsilon^2 + \frac{\varepsilon^4}{2} + o(\varepsilon^4)$, d'où $1 - e^{-\varepsilon^2} \leq \varepsilon^2 - \frac{\varepsilon^4}{2} + o(\varepsilon^4) \leq \varepsilon^2$. □

**Lemme 5.3.2.** *Soient $\varepsilon \in ]0,1[$ et $\delta$ tel que $e^{2\sqrt{2\Lambda}\varepsilon + 2\varepsilon^2} - 1 \leq \delta$.*
*Soit la famille $((l_i, u_i))_{i \in \left\{0,...,\left\lfloor\sqrt{\Lambda/2\varepsilon^2}\right\rfloor - 1\right\}}$, où $l_i = \frac{1}{1+\delta} f_{\lambda_i}$, $u_i = (1+\delta) f_{\lambda_{i+1}}$ et $\lambda_i = 2\varepsilon^2 i^2$.*
*Alors, pour tout $\lambda \in [0, \Lambda]$, il existe $i$ compris entre 0 et $\left\lfloor \sqrt{\frac{\Lambda}{2\varepsilon^2}} \right\rfloor - 1$, tel que pour tout $k \in \mathbb{N}$,*

$$l_i(k) = \frac{1}{1+\delta} e^{-\lambda_i}\frac{\lambda_i^k}{k!} \leq e^{-\lambda}\frac{\lambda^k}{k!} \leq (1+\delta) e^{-\lambda_{i+1}}\frac{\lambda_{i+1}^k}{k!} = u_i(k).$$

*Démonstration.* On cherche $i \in \mathbb{N}$, tel que $\lambda_i \leq \lambda \leq \lambda_{i+1}$. C'est-à-dire d'une part, $\lambda_i = 2\varepsilon^2 i^2 < \lambda \iff i < \frac{1}{\varepsilon}\sqrt{\frac{\lambda}{2}}$, d'autre part $\lambda \leq \lambda_{i+1} = 2\varepsilon^2(i+1)^2 \iff \frac{1}{\varepsilon}\sqrt{\frac{\lambda}{2}} - 1 \leq i$. Donc $i = \left\lfloor \frac{1}{\varepsilon}\sqrt{\frac{\lambda}{2}} \right\rfloor$.

- Montrons la première inégalité.

Soit $k \in \mathbb{N}$,

$$\frac{1}{1+\delta} e^{-\lambda_i}\frac{\lambda_i^k}{k!} \leq e^{-\lambda}\frac{\lambda^k}{k!} \iff \frac{1}{1+\delta} e^{\lambda - \lambda_i} \leq \left(\frac{\lambda}{\lambda_i}\right)^k.$$

On souhaite que cette inégalité soit vraie pour tout $k \in \mathbb{N}$, c'est-à-dire,

$$\frac{1}{1+\delta} e^{\lambda - \lambda_i} \leq \min_{k \in \mathbb{N}} \left(\frac{\lambda}{\lambda_i}\right)^k.$$

Or remarquons que $\frac{\lambda}{\lambda_i} \geq 1$. La fonction qui à $k$ associe $\left(\frac{\lambda}{\lambda_i}\right)^k$ est donc croissante. Elle

atteint donc son minimum en $k = 0$. Pour cela

$$\begin{aligned}\frac{1}{1+\delta}e^{-\lambda_i}\frac{\lambda_i^k}{k!} \leq e^{-\lambda}\frac{\lambda^k}{k!} &\iff \frac{1}{1+\delta}e^{\lambda-\lambda_i} \leq \left(\frac{\lambda}{\lambda_i}\right)^0 \\ &\iff \frac{1}{1+\delta}e^{\lambda-\lambda_i} \leq 1 \\ &\iff e^{\lambda-\lambda_i} \leq 1+\delta \\ &\iff e^{\lambda-\lambda_i} - 1 \leq \delta.\end{aligned}$$

De plus, $|\lambda - \lambda_i| \leq |\lambda_{i+1} - \lambda_i| \leq 2\varepsilon^2((i+1)^2 - i^2) = 2\varepsilon^2(2i+1)$. Or, $i \leq \left\lfloor \sqrt{\frac{\Lambda}{2\varepsilon^2}} \right\rfloor$, donc $|\lambda - \lambda_i| \leq 2\varepsilon^2 \left(2\left\lfloor \sqrt{\frac{\Lambda}{2\varepsilon^2}} \right\rfloor + 1\right) \leq 2\sqrt{2}\varepsilon\sqrt{\Lambda} + 2\varepsilon^2$.

- Montrons la seconde inégalité.

Soit $k \in \mathbb{N}$,

$$e^{-\lambda}\frac{\lambda^k}{k!} \leq (1+\delta)e^{-\lambda_{i+1}}\frac{\lambda_{i+1}^k}{k!} \iff \left(\frac{\lambda}{\lambda_{i+1}}\right)^k \leq (1+\delta)e^{\lambda-\lambda_{i+1}}.$$

On souhaite que cette inégalité soit vraie pour tout $k \in \mathbb{N}$, c'est-à-dire,

$$\max_{k\in\mathbb{N}} \left(\frac{\lambda}{\lambda_{i+1}}\right)^k \leq (1+\delta)e^{\lambda-\lambda_{i+1}}.$$

On remarque que la fonction qui à $k$ associe $\left(\frac{\lambda}{\lambda_{i+1}}\right)^k$ est décroissante car $\frac{\lambda}{\lambda_{i+1}} \leq 1$, le maximum est donc atteint pour $k = 0$. Ainsi

$$\begin{aligned}e^{-\lambda}\frac{\lambda^k}{k!} \leq (1+\delta)e^{-\lambda_{i+1}}\frac{\lambda_{i+1}^k}{k!} &\iff 1 \leq (1+\delta)e^{\lambda-\lambda_{i+1}} \\ &\iff e^{\lambda_{i+1}-\lambda} \leq (1+\delta) \\ &\iff e^{\lambda_{i+1}-\lambda} - 1 \leq \delta.\end{aligned}$$

Pour les mêmes raisons que pour la minoration, il faut que $e^{2\sqrt{2}\varepsilon\sqrt{\Lambda}+2\varepsilon^2} - 1 \leq \delta$ pour que l'inégalité soit vérifiée.

□

**Première majoration de $N_{[.]}(\varepsilon, \mathcal{H}_{\kappa,\Lambda}, d_H)$**

**Théorème 5.3.2.** *Soient $\varepsilon \in ]0,1[$ et $\kappa, \Lambda \in \mathbb{N}^*$,*

$$N_{[.]}(\varepsilon, \mathcal{H}_{\kappa,\Lambda}, d_H) \leq \frac{\kappa(108\sqrt{\pi e})^\kappa}{3}\frac{\Lambda^\kappa}{\varepsilon^{3\kappa-1}}.$$

*D'où* $H_{[.]}(\varepsilon, \mathcal{H}_{\kappa,\Lambda}, d_H) \lesssim \kappa\left(\ln\frac{1}{\varepsilon} + \ln\Lambda\right)$.

*Démonstration.* Par [GW00, Lemme 2], une borne supérieure du nombre de $\varepsilon-$crochet du simplexe d'ordre $\kappa - 1$ est $N_{[.]}(\varepsilon, \mathfrak{S}_{\kappa-1}, d_H) = \frac{\kappa(2\pi e)^{\kappa/2}}{\varepsilon^{\kappa-1}}$.

Puis par [GW00, Théorème 2] et le Lemme 5.3.1,

$$\begin{aligned} N_{[.]}(\varepsilon, \mathcal{H}_{\kappa,\Lambda}, d_H) &\leq N_{[.]}\left(\frac{\varepsilon}{3}, \mathfrak{S}_{\kappa-1}, d_H\right) \prod_{i=1}^{\kappa} N_{[.]}\left(\frac{\varepsilon}{3}, \mathcal{F}_{\Lambda}, d_H\right) \\ &= \frac{\kappa(9\pi e)^{\kappa/2}}{3\varepsilon^{\kappa-1}} \left(\frac{36\Lambda}{\varepsilon^2}\right)^{\kappa} \\ &= \frac{\kappa(\pi e)^{\kappa/2} 108^{\kappa}}{3} \frac{\Lambda^{\kappa}}{\varepsilon^{3\kappa-1}}. \end{aligned}$$

□

### 5.3.2 Pour l'ensemble des mélanges infinis de lois

Dans cette sous-partie, nous nous intéressons à l'entropie à crochet de l'ensemble des mélanges quelconques de lois de Poisson. Cet ensemble est noté :

$$\mathcal{H}_{\Lambda} = \left\{ h_G(.) = \int_0^{\Lambda} f_{\lambda}(.) \, \mathrm{d}G(\lambda), \quad G \in \mathcal{G} \right\},$$

où $\mathcal{G}$ est l'ensemble des probabilités sur $\mathbb{R}_+$ à support fini ou infini.

Nous sortons un peu du cadre de notre modèle, puisqu'on suppose que les observations sont issues d'un mélange fini de lois de Poisson. Mais la borne supérieure de l'entropie à crochet que nous obtenons ne dépend pas du nombre de lois dans le mélange, uniquement de $\Lambda$ la taille maximale du paramètre des lois de Poisson. Ainsi, à partir d'une certaine valeur $K$, pour tout $\kappa \geq K$, la borne obtenue dans cette sous-section sera choisie car elle donne de meilleurs résultats que le Théorème 5.3.2.

**Proposition 5.3.3.** *Soient $\varepsilon \in ]0, 1[$ et $h_G \in \mathcal{H}_{\Lambda}$. Il existe un mélange discret de probabilité $G'$ avec au plus $\kappa = C_{\Lambda} \ln(4/\varepsilon) + 1$ lois de Poisson, où $C_{\Lambda} = \max(\Lambda, 21) + 2e\Lambda$, tel que*

$$d_{TV}\left(h_G, h_{G'}\right) \leq \varepsilon.$$

*Démonstration.* Cette démonstration, inspirée des travaux de [BG14], se décompose en deux étapes.

La première étape de la démonstration consiste à montrer que l'on peut tronquer toutes les lois de Poisson de paramètre inférieur à $\Lambda$ à partir d'un certain rang $N$, qui dépend de $\Lambda$. En effet, la loi de Poisson est une loi à queue exponentielle (utilisation d'une propriété de la loi de Poisson).

La seconde étape consiste à approcher toute loi de mélange continue par un mélange fini dont le nombre de composantes dépend de $\Lambda$. En effet, soit certains paramètres sont très proches et nous pouvons les considérer égaux, soit les poids sont très petits et nous pouvons les considérer nuls (utilisation d'une propriété sur la loi de mélange).

Comme la loi Poisson n'est pas une loi à queue lourde, à partir d'un certain rang $N \in \mathbb{N}_+^*$, pour tout $x > N$, les densités $h_G(x)$ et $h_{G'}(x)$ sont petites. Coupons alors la distance en deux,

$$d_{TV}(h_G, h_{G'}) = \frac{1}{2}\sum_{x=0}^{N} |h_G(x) - h_{G'}(x)| + \frac{1}{2}\sum_{x=N+1}^{\infty} |h_G(x) - h_{G'}(x)|\,.$$

• *1 ère étape :* montrons qu'il existe $N \in \mathbb{N}^*$ tel que

$$\sum_{x=N+1}^{\infty} |h_G(x) - h_{G'}(x)| \leq \frac{\varepsilon}{2}. \tag{5.4}$$

Soient $g$ et $g'$ les densités respectives des lois de mélange $G$ et $G'$,

$$\begin{aligned}
\sum_{x=N+1}^{\infty} |h_G(x) - h_{G'}(x)| &= \sum_{x=N+1}^{\infty} \left| \int_0^{\Lambda} f_z(x)[g(z) - g'(z)]\,\mathrm{d}z \right| \\
&= \sum_{x=N+1}^{\infty} \left| \int_0^{\Lambda} e^{-z}\frac{z^x}{x!}[g(z) - g'(z)]\,\mathrm{d}z \right| \\
&\leq \int_0^{\Lambda} |g(z) - g'(z)| \sum_{x=N+1}^{\infty} e^{-z}\frac{z^x}{x!}\,\mathrm{d}z \\
&\leq \sum_{x=N+1}^{\infty} e^{-\Lambda}\frac{\Lambda^x}{x!} \int_0^{\Lambda} |g(z) - g'(z)|\,\mathrm{d}z \qquad (5.5) \\
&\leq 2\sum_{x=N+1}^{\infty} e^{-\Lambda}\frac{\Lambda^x}{x!} \qquad (5.6) \\
&= 2\mathbb{P}(X_\Lambda \geq N+1).
\end{aligned}$$

On obtient l'inégalité (5.5) car $\sum_{x=N+1}^{\infty} e^{-z}\frac{z^x}{x!}$ est la queue de distribution de la loi de Poisson. Celle-ci est maximale pour le plus grand paramètre des lois de Poisson possible, dans notre cas $\Lambda$. L'inégalité (5.6) est obtenue car $g$ et $g'$ sont des densités de probabilité.

Chercher un $N \in \mathbb{N}^*$ qui vérifie l'inégalité (5.4) revient, par le Lemme 5.3.4, à chercher $N$ tel que

$$\begin{aligned}
e^{-(N+1)\left(\ln\frac{N+1}{\Lambda}-1\right)-\Lambda} \leq \frac{\varepsilon}{4} &\iff -(N+1)\left(\ln\frac{N+1}{\Lambda} - 1\right) - \Lambda \leq \ln\frac{\varepsilon}{4} \\
&\iff \underbrace{(N+1)\left(\ln\frac{N+1}{\Lambda} - 1\right) + \Lambda}_{L_\Lambda(N+1)} \geq \ln\frac{4}{\varepsilon}.
\end{aligned}$$

On remarque tout d'abord que la fonction $L_\Lambda(N+1)$ est croissante dès que $N+1 \geq \Lambda$. En effet, $L_\Lambda(N+1)' = \ln \frac{N+1}{\Lambda} + (N+1)\frac{1/\Lambda}{(N+1)/\Lambda} - 1 = \ln \frac{N+1}{\Lambda}$ ainsi $L_\Lambda(N+1) \geq 0 \iff N+1 \geq \Lambda$.

Montrons ensuite que $N+1 = \max(21, \Lambda) \ln \frac{4}{\varepsilon}$, qui est supérieur ou égale à $\Lambda$ car $\varepsilon < 1$, convient.

Si $\Lambda \geq 21$, alors

$$\begin{aligned} L_\Lambda\left(\Lambda \ln \frac{4}{\varepsilon}\right) &= \left(\ln \frac{\Lambda \ln(4/\varepsilon)}{\Lambda} - 1\right) \Lambda \ln \frac{4}{\varepsilon} + \Lambda \\ &= \ln \frac{4}{\varepsilon} \left(\Lambda \ln\left(\ln \frac{4}{\varepsilon}\right) - \Lambda + \frac{\Lambda}{\ln(4/\varepsilon)}\right) \\ &= \ln \frac{4}{\varepsilon} \left(\Lambda \underbrace{\left(\ln\left(\ln \frac{4}{\varepsilon}\right) - 1 + \frac{1}{\ln(4/\varepsilon)}\right)}_{\ell(\varepsilon)}\right). \end{aligned}$$

Il reste à montrer que $\Lambda\ell(\varepsilon) \geq 1$. Or, étudier la fonction $\ell(\varepsilon)$, pour $\varepsilon \in ]0,1[$, revient à étudier $\ell_1(x) = \ln x - 1 + \frac{1}{x}$ pour $x = \ln \frac{4}{\varepsilon} > \ln 4$. Notons que $\ell_1(x)' = \frac{1}{x} - \frac{1}{x^2} \geq 0$ si est seulement si $x^2 \geq x \iff x \geq 1$. Or, dans notre cas, $x > 1$. Ainsi, $\ell_1$ est strictement croissante pour tout $x > \ln 4$. Nous en déduisons que $\ell$ est décroissante pour tout $\varepsilon \in ]0,1[$. Le minimum de $\ell$ est alors atteint en 1 et vaut $\ell(1) = \ln(\ln 4) - 1 + \frac{1}{\ln 4} \geq \frac{1}{21}$. Or comme $\Lambda \geq 21$, $\Lambda\ell(\varepsilon) \geq \Lambda\ell(1) \geq 1$. Donc $L_\Lambda\left(\Lambda \ln \frac{4}{\varepsilon}\right) \geq \ln \frac{4}{\varepsilon}$.

Si $\Lambda \leq 21$, alors

$$\begin{aligned} L_\Lambda\left(21 \ln \frac{4}{\varepsilon}\right) &= \left(\ln \frac{21 \ln(4/\varepsilon)}{\Lambda} - 1\right) 21 \ln \frac{4}{\varepsilon} + M \\ &= \ln \frac{4}{\varepsilon} \left(21 \ln \frac{21 \ln(4/\varepsilon)}{\Lambda} - 21 + \frac{\Lambda}{\ln(4/\varepsilon)}\right) \\ &\geq \ln \frac{4}{\varepsilon} \left(\underbrace{21 \ln \frac{21}{\Lambda} + \frac{\Lambda - 21}{\ln(4/\varepsilon)}}_{\ell_2(\Lambda)} + 21 \underbrace{\left(\ln\left(\ln \frac{4}{\varepsilon}\right) - 1 + \frac{1}{\ln(4/\varepsilon)}\right)}_{\ell(\varepsilon)}\right). \end{aligned}$$

Par les calculs précédents, nous savons déjà que $21\ell(\varepsilon) \geq 1$, il reste à vérifier que $\ell_2(\Lambda) \geq 0$ pour tout $\Lambda \in [0, 21]$. Notons que $\ell_2(\Lambda)' = \frac{-21^2/\Lambda^2}{21/\Lambda} + \frac{1}{\ln(4/\varepsilon)} = \frac{-21}{\Lambda} + \frac{1}{\ln(4/\varepsilon)} \geq 0 \iff \Lambda \geq 21 \ln \frac{4}{\varepsilon} \geq 21$. Donc sur $[0, 21]$, $\ell_2$ est décroissante, et son minimum est atteint en 21 et vaut $\ell_2(21) = 0$. Donc $\ell_2(\Lambda) + 21\ell(\varepsilon) \geq \ell_2(21) + 21\ell(1) \geq 1$ d'où $L_\Lambda\left(21 \ln \frac{4}{\varepsilon}\right) \geq \ln \frac{4}{\varepsilon}$.

• *2 ème étape :* montrons à présent qu'il existe une loi de mélange finie $G'$ telle que

$$\sum_{x=0}^{N} |h_G(x) - h_{G'}(x)| \leq \frac{3\varepsilon}{2}.$$

Pour cela, montrons que la loi de Poisson tronquée peut s'approcher par un polynôme de degré $J$, que l'on va définir plus tard. Puis nous approcherons la loi de mélange continu par un mélange fini.

Par le développement de Taylor de la fonction exponentielle, pour tout $z \in \mathbb{R}_+$,

$$\left| e^{-z} - \sum_{j=0}^{J-1} \frac{(-1)^j z^j}{j!} \right| = \left| \sum_{j \geq J} \frac{(-1)^j z^j}{j!} \right| \leq \frac{z^J}{j!}.$$

Ainsi, nous approchons la densité d'une loi de Poisson tronquée par $\sum_{j=0}^{J-1} \frac{(-1)^j z^{j+x}}{j!x!}$ et

$$\left| f_z(x) - \sum_{j=0}^{J-1} \frac{(-1)^j z^{j+x}}{j!x!} \right| \leq \frac{z^{J+x}}{j!x!}.$$

D'où

$$\begin{aligned}
\sum_{x=0}^{N} |p_G(x) - p_{G'}(x)| &= \sum_{x=0}^{N} \left| \int_0^\Lambda f_z(x)[g(z) - g'(z)] \, \mathrm{d}z \right| \\
&\leq \sum_{x=0}^{N} \left| \int_0^\Lambda \sum_{j=0}^{J-1} \frac{(-1)^j z^{j+x}}{j!x!} [g(z) - g'(z)] \, \mathrm{d}z \right| \\
&\quad + \sum_{x=0}^{N} \int_0^\Lambda \left| f_z(x) - \sum_{j=0}^{J-1} \frac{(-1)^j z^{j+x}}{j!x!} \right| |g(z) - g'(z)| \, \mathrm{d}z \\
&\leq \underbrace{\sum_{x=0}^{N} \left| \int_0^\Lambda \sum_{j=0}^{J-1} \frac{(-1)^j z^{j+x}}{j!x!} [g(z) - g'(z)] \, \mathrm{d}z \right|}_{A} \\
&\quad + \underbrace{\sum_{x=0}^{N} \int_0^\Lambda \frac{z^{J+x}}{J!x!} |g(z) - g'(z)| \, \mathrm{d}z}_{B}.
\end{aligned}$$

Si $G'$ est telle que

$$\int z^l g(z) \, \mathrm{d}z = \int z^l g'(z) \, \mathrm{d}z, \text{ pour tout } l \in \{0, ..., N + J - 1\},$$

alors $A$ est nulle. Or, par le Théorème de Carathéodory (voir [GV01, Lemme A1]) , $G'$, vérifiant l'inégalité au dessus, peut être choisie comme une probabilité discrète sur $[0, \Lambda]$ avec au plus $N + J + 1$ points.

Il reste à définir $J$ de telle sorte que $B \leq \frac{3\varepsilon}{2}$.
Or, $B \leq 2\sum_{x=0}^{N} \frac{\Lambda^{J+x}}{J!x!}$, car $z \in [0,\Lambda]$, et, $g$ et $g'$ sont des densités. Comme $\sum_{x=0}^{N} \frac{\Lambda^x}{x!} \leq e^{\Lambda}$, majorons $B$ par $2\frac{e^{\Lambda}\Lambda^J}{J!}$. Ainsi, cherchons $J$ tel que $\frac{e^{\Lambda}\Lambda^J}{J!} \leq \frac{3\varepsilon}{4}$. Par la formule de Stirling, $J! \geq \sqrt{2\pi J}\left(\frac{J}{e}\right)^J$,

$$\frac{e^{\Lambda}(e\Lambda)^J}{J^J\sqrt{2\pi J}} \leq \frac{3\varepsilon}{4} \iff \frac{(e\Lambda)^J e^{\Lambda}}{J^J\sqrt{J}} \leq \frac{3\sqrt{2\pi}\varepsilon}{4}.$$

Pour simplifier, cherchons $J$ tel que

$$\begin{aligned}\left(\frac{e\Lambda}{J}\right)^J e^{\Lambda} = e^{\Lambda - J\ln\left(\frac{J}{e\Lambda}\right)} \leq \frac{3\sqrt{2\pi}}{4}\varepsilon \quad &\iff \quad J\ln\frac{J}{e\Lambda} - \Lambda \geq \ln\frac{4}{\varepsilon} + \ln\frac{1}{3\sqrt{2\pi}} \\ &\impliedby \quad L_{\Lambda}^{(2)}(J) = J\ln\frac{J}{e\Lambda} - \Lambda \geq \ln\frac{4}{\varepsilon},\end{aligned}$$

car $\frac{1}{3\sqrt{2\pi}} \leq 1$ donc $\ln\frac{1}{3\sqrt{2\pi}} \leq 0$.

On remarque tout d'abord que $L_{\Lambda}^{(2)}(J)$ est croissante pour tout $J \geq \Lambda$. En effet, $L_{\Lambda}^{(2)'}(J) = \ln\frac{J}{e\Lambda} + J\frac{1/e\Lambda}{J/e\Lambda} = \ln\frac{J}{e\Lambda} + 1 \geq 0$. Montrons ensuite que $J = \Lambda 2e\ln\frac{4}{\varepsilon}$ convient.

$$\begin{aligned}L_{\Lambda}^{(2)}\left(2e\Lambda\ln\frac{4}{\varepsilon}\right) &= \ln\frac{4}{\varepsilon}\left(2e\Lambda\ln\left(\frac{2e\Lambda\ln(4/\varepsilon)}{e\Lambda}\right) - \frac{\Lambda}{\ln(4/\varepsilon)}\right) \\ &= \ln\frac{4}{\varepsilon}\left(\Lambda\underbrace{\left(2e\ln\left(2\ln\frac{4}{\varepsilon}\right) - \frac{1}{\ln(4/\varepsilon)}\right)}_{\ell_3(\varepsilon)}\right).\end{aligned}$$

Reste à voir que pour tout $\varepsilon \in ]0,1[$, $\Lambda\ell_3(\varepsilon) \geq 1$. On note que la fonction $\ell_3(\varepsilon)$ est décroissante, atteint son minimum en 1 et vaut $\ell_3(1) \geq 4$. De plus $\Lambda \in \mathbb{N}^*$ donc l'inégalité est vérifiée.

• Pour conclure la démonstration de la proposition, nous avons montré dans la deuxième étape, que l'on peut approcher le mélange $G$ par un mélange d'au plus $N + J + 2 = C_{\Lambda}\ln(4/\varepsilon) + 1$, où $C_{\Lambda} = \max(\Lambda, 21) + 2e\Lambda$.

□

**Lemme 5.3.4.** *Soient $X$ une variable aléatoire de Poisson de paramètre $\lambda$ et $N \in \mathbb{R}_+$ alors,*

$$\mathbb{P}(X \geq N) \leq e^{-N\ln\frac{N}{\lambda} + N - \lambda}.$$

*Démonstration.* Tout d'abord, montrons que pour tout $t > 0$,

$$\mathbb{P}(X \geq N) \leq e^{-tN+\lambda(e^t-1)}.$$

Soit $t \in \mathbb{R}_+$, comme $\mathbb{P}(X \geq N) = \mathbb{P}(e^{tX} \geq e^{tN})$, alors

$$\begin{aligned}\mathbb{E}(e^{tX}) &= \mathbb{E}\left(e^{tX}\mathbb{1}_{e^{tX}\geq e^{tN}}\right) + \mathbb{E}\left(e^{tX}\mathbb{1}_{e^{tX}<e^{tN}}\right) \\ &\geq e^{tN}\mathbb{E}\left(\mathbb{1}_{e^{tX}\geq e^{tN}}\right) + \mathbb{E}\left(e^{tX}\mathbb{1}_{e^{tX}<e^{tN}}\right) \\ &\geq e^{tN}\mathbb{P}\left(e^{tX} \geq e^{tN}\right).\end{aligned}$$

Par la transformée de Laplace d'une loi de Poisson, nous obtenons l'inégalité souhaitée.

Il reste à trouver $t$ qui minimise $e^{-tN+\lambda(e^t-1)}$. La fonction exponentielle étant croissante, cherchons le minimum de $\phi(t) = -tN + \lambda(e^t - 1)$, or $\phi'(t) = \lambda e^t - N$, alors $\phi'(t) \leq 0 \iff e^t \leq \frac{N}{\lambda} \iff t \leq \ln\left(\frac{N}{\lambda}\right)$. Ainsi, le minimum de $\phi(t)$ est atteint pour $t = \ln\left(\frac{N}{\lambda}\right)$, ce qui conclut la preuve du lemme.

□

**Théorème 5.3.3.** *Soient* $\varepsilon \in ]0,1[$, $\Lambda \in \mathbb{N}^*$

$$N_{[.]}\left(\varepsilon, \mathcal{H}_\Lambda, d_H\right) \leq \frac{\left[C'_\Lambda \ln(4/\varepsilon) + 1\right]\left(108\sqrt{\pi e}\right)^{\left(C'_\Lambda \ln(4/\varepsilon)+1\right)}}{3} \frac{(\Lambda)^{C'_\Lambda \ln(4/\varepsilon)+1}}{\varepsilon^{3C'_\Lambda \ln(4/\varepsilon)-2}},$$

*et*

$$H_{[.]}(\varepsilon, \mathcal{H}_\Lambda, d_H) \lesssim C'_\Lambda \ln\frac{1}{\varepsilon}\left(\ln\frac{1}{\varepsilon} + \ln(\Lambda)\right)$$

*où* $C'_\Lambda = 2\left(\max(\Lambda, 21) + 2e\Lambda\right)$.

*Démonstration.* On rappelle que $d_H^2 \leq d_{TV}$. Par la proposition précédente, nous approchons à une distance en variation totale $\delta$ tous mélanges infinis par un mélange à $N + J + 2 = C_\Lambda \ln\frac{4}{\delta} + 1$ lois, où $C_\Lambda = \max(\Lambda, 21) + 2e\Lambda$. Le nombre minimal de $\varepsilon-$crochets pour couvrir $\mathcal{H}_\Lambda$ avec la distance d'Hellinger est $N + J + 2 = 2C_\Lambda \ln\frac{4}{\varepsilon} + 1$. En effet,

$$d_{TV}(l_i, u_i) \leq \delta \iff d_H^2(l_i, u_i) \leq \delta \iff d_H(l_i, u_i) \leq \sqrt{\delta} \leq \varepsilon.$$

C'est-à-dire $\delta = \varepsilon^2$.

Puis, le Théorème 5.3.2 donne un borne supérieure du nombre de $\varepsilon-$crochet associé à la distance de Hellinger et à l'ensemble des mélanges composés de $\kappa = 2\left(\max(\Lambda, 21) + 2e\Lambda\right)$ lois, d'où le résultat. □

### 5.3.3 Pour l'ensemble des mélanges finis de lois

En rassemblant les résultats des Théorèmes 5.3.2 et 5.3.3 nous obtenons,

**Corollaire 5.3.5.** *Soient $\varepsilon \in ]0,1[$ et $\kappa, \Lambda \in \mathbb{N}^*$,*

$$H_{[.]}\left(\varepsilon, \mathcal{H}_{\kappa,\Lambda}, d_H\right) \lesssim \left(\kappa \wedge C_\Lambda \ln \frac{1}{\varepsilon}\right)\left(\ln \frac{1}{\varepsilon} + \ln \Lambda\right),$$

*où $C_\Lambda = \Lambda + 1$.*

## 5.4 Critère pénalisé non asymptotique pour les mélanges poissoniens

En suivant le Théorème 5.2.1, grâce aux calculs précédents, commençons par définir la fonction $\Psi_{\kappa,\Lambda}$ qui vérifie les propriétés 1, 2, 3. Puis, calculons la fonction de pénalité et l'inégalité oracle associée. Nous notons $a \vee b$ le maximum entre $a$ et $b$, et $a \wedge b$ le minimum entre $a$ et $b$.

**Proposition 5.4.1.** *Soient $\Lambda \in \mathbb{N}^*\backslash\{1,2\}$ et , $\kappa \in \mathbb{N}^*$, $\xi > 0$, il existe $c > 0$ telle que*

$$\int_0^\xi \sqrt{H_{[.]}(x, \mathcal{H}_{\kappa,\Lambda}, d_H)}\,\mathrm{d}x \leq c\xi\left(\sqrt{\kappa} + \sqrt{\Lambda}\right)\sqrt{\ln \Lambda}\left[\sqrt{\ln \frac{1}{\xi \wedge 1} \vee \frac{\kappa}{(\Lambda+1)\ln \Lambda}} + 1\right] \quad (5.7)$$

*Démonstration.* Dans la démonstration, $c$ indiquera une constante positive.

Suivant que $\kappa$ est plus grand ou non que $(\Lambda+1)\ln \frac{1}{x}$, $H_{[.]}(x, \mathcal{H}_{\kappa,\Lambda}, d_H)$ n'a pas la même valeur. Nous allons distinguer différents cas :

– Si $\kappa \leq \Lambda$ et $x \leq 1/e \leq 1$ alors $\kappa < (\Lambda+1)\ln \frac{1}{x}$ et

$$H_{[.]}\left(x, \mathcal{H}_{\kappa,\Lambda}, d_H\right) \lesssim \kappa \ln \frac{1}{x} + \kappa \ln \Lambda.$$

Ainsi, pour $\xi \leq 1/e$,

$$\begin{aligned}
\int_0^\xi \sqrt{H_{[.]}(x, \mathcal{H}_{\kappa,\Lambda}, d_H)}\,\mathrm{d}x &\leq c\int_0^\xi \sqrt{\kappa \ln \frac{1}{x} + \kappa \ln \Lambda}\,\mathrm{d}x \\
&\leq c\sqrt{\kappa}\int_0^\xi \sqrt{\ln \frac{1}{x}}\,\mathrm{d}x + \sqrt{\kappa}\sqrt{\ln \Lambda}\xi \\
&\leq c\xi\sqrt{\kappa}\left(\sqrt{\ln \frac{1}{\xi}} + \sqrt{\pi} + \sqrt{\ln \Lambda}\right) \\
&\leq c\xi\sqrt{\kappa}\left(\sqrt{\ln \frac{1}{\xi}} + \sqrt{\ln \Lambda}\right).
\end{aligned}$$

L'avant-dernière inégalité provient de [MB11, Lemme A.2].

– Si $\kappa \geq \Lambda + 1$ et $x \geq 1/e$, alors $\kappa \geq (\Lambda+1)\ln\frac{1}{x}$ et

$$H_{[.]}\left(x, \mathcal{H}_{\kappa,\Lambda}, d_H\right) \lesssim (\Lambda+1)\left(\ln\frac{1}{x}\right)^2 + (\Lambda+1)\ln\Lambda\ln\frac{1}{x}.$$

Ainsi, pour $\xi \geq 1/e$,

$$\begin{aligned}
\int_0^\xi \sqrt{H_{[.]}(x, \mathcal{H}_{\kappa,\Lambda}, d_H)}\,\mathrm{d}x &\leq c\int_0^{1\wedge\xi} \sqrt{(\Lambda+1)\left(\ln\frac{1}{x}\right)^2 + (\Lambda+1)\ln\Lambda\ln\frac{1}{x}}\,\mathrm{d}x \\
&\leq c\sqrt{\Lambda+1}\left(\int_0^{1\wedge\xi}\ln\frac{1}{x}\,\mathrm{d}x + \sqrt{\ln\Lambda}\int_0^{1\wedge\xi}\sqrt{\ln\frac{1}{x}}\,\mathrm{d}x\right) \\
&\leq c\xi\sqrt{\Lambda+1}\left(\ln\frac{1}{1\wedge\xi} + 1 + \sqrt{\ln\Lambda}\left(\sqrt{\ln\frac{1}{1\wedge\xi}} + \sqrt{\pi}\right)\right).
\end{aligned}$$

Notons que $\xi \geq \frac{1}{e} \iff \ln\frac{1}{\xi} \leq 1 \iff \ln\frac{1}{\xi} \leq \sqrt{\ln\frac{1}{\xi}}$, et $\Lambda \geq 3$, d'où

$$\int_0^\xi \sqrt{H_{[.]}(x, \mathcal{H}_{\kappa,\Lambda}, d_H)}\,\mathrm{d}x \leq \xi c\sqrt{\Lambda\ln\Lambda}\left(\sqrt{\ln\frac{1}{1\wedge\xi}} + 1\right).$$

– Si $\kappa \geq \Lambda + 1$ et $x \leq 1/e \leq 1$. On note que $\kappa \leq (\Lambda+1)\ln\frac{1}{x} \iff \frac{\kappa}{(\Lambda+1)} \leq \ln\frac{1}{x} \iff x \leq e^{-\frac{\kappa}{\Lambda+1}}$. Ainsi, pour $x \leq 1/e$

$$\begin{aligned}
\int_0^\xi \sqrt{H_{[.]}(x, \mathcal{H}_{\kappa,\Lambda}, d_H)}\,\mathrm{d}x &\leq c\sqrt{\kappa}\int_0^{\xi\wedge e^{-\frac{\kappa}{\Lambda+1}}} \sqrt{\left(\ln\frac{1}{x} + \ln\Lambda\right)}\,\mathrm{d}x \\
&\quad + c\sqrt{\Lambda+1}\int_{\xi\wedge e^{-\frac{\kappa}{\Lambda+1}}}^{\xi} \sqrt{\left(\ln\frac{1}{x}\right)^2 + \ln\Lambda\ln\frac{1}{x}}\,\mathrm{d}x \\
&\leq c\sqrt{\kappa}\left(\int_0^{\xi\wedge e^{-\frac{\kappa}{\Lambda+1}}} \sqrt{\ln\frac{1}{x}}\,\mathrm{d}x + \sqrt{\ln\Lambda}\left(\xi\wedge e^{-\frac{\kappa}{\Lambda+1}}\right)\right) \\
&\quad + c\sqrt{\Lambda+1}\left(\int_0^{\xi}\ln\frac{1}{x}\,\mathrm{d}x + \sqrt{\ln\Lambda}\int_0^{\xi}\sqrt{\ln\frac{1}{x}}\,\mathrm{d}x\right) \\
&\leq c\xi\sqrt{\kappa}\left(\sqrt{\ln\frac{1}{\xi\wedge e^{-\frac{\kappa}{\Lambda+1}}}} + \sqrt{\pi} + \sqrt{\ln\Lambda}\right) \\
&\quad + c\xi\sqrt{\Lambda\ln\Lambda}\left(\sqrt{\ln\frac{1}{\xi}} + \sqrt{\pi}\right) \\
&\leq c\xi\sqrt{\kappa\ln\Lambda}\left(\sqrt{\frac{\ln(1/\xi)\vee(\kappa/(\Lambda+1))}{\ln\Lambda}} + 1\right) \\
&\quad + \xi c\sqrt{\Lambda\ln\Lambda}\left(\sqrt{\ln\frac{1}{\xi}} + 1\right).
\end{aligned}$$

– Si $\kappa \leq \Lambda$ et $x \geq 1/e$. Notons que $\kappa \leq (\Lambda+1)\ln\frac{1}{x} \iff \frac{\kappa}{(\Lambda+1)} \leq \ln\frac{1}{x} \iff x \leq$

$e^{-\frac{\kappa}{\Lambda+1}}$. Ainsi, pour $\xi \geq 1/e$

$$\begin{aligned}
\int_0^{\xi} \sqrt{H_{[.]}(x, \mathcal{H}_{\kappa,\Lambda}, d_H)}\, \mathrm{d}x &\leq c\sqrt{\kappa} \int_0^{\xi \wedge e^{-\frac{\kappa}{\Lambda+1}}} \sqrt{\left(\ln \frac{1}{x} + \ln \Lambda\right)}\, \mathrm{d}x \\
&\quad + c\sqrt{\Lambda+1} \int_{\xi \wedge e^{-\frac{\kappa}{\Lambda+1}}}^{1 \wedge \xi} \sqrt{\left(\ln \frac{1}{x}\right)^2 + \ln \Lambda \ln \frac{1}{x}}\, \mathrm{d}x \\
&\leq c\xi\sqrt{\kappa} \left(\sqrt{\ln \frac{1}{\xi} \vee \frac{\kappa}{\Lambda+1}} + \sqrt{\ln \Lambda}\right) \\
&\quad + c\xi\sqrt{\Lambda \ln \Lambda} \left(\sqrt{\ln \frac{1}{1 \wedge \xi}} + \sqrt{\pi}\right) \\
&\leq c\xi \left(\sqrt{\kappa} + \sqrt{\Lambda}\right) \left(\sqrt{\ln \frac{1}{\xi \wedge 1} \vee \frac{\kappa}{(\Lambda+1)\ln \Lambda}} + 1\right).
\end{aligned}$$

Dans les quatre cas, l'inégalité (5.7) est vérifiée.

□

**Lemme 5.4.2.** *Soit $\kappa \in \mathbb{N}^*$, $\Lambda \in \mathbb{N}^* \backslash \{1, 2\}$, pour tout $\xi \in \mathbb{R}_+^*$, nous définissons*

$$\Psi_{\kappa,\Lambda}(\xi) = \xi \left(\sqrt{\kappa} + \sqrt{\Lambda}\right) \sqrt{\ln \Lambda} \left(\sqrt{\ln \frac{1}{\xi \wedge 1} \vee \frac{\kappa}{(\Lambda+1)\ln \Lambda}} + 1\right)$$

*La fonction $\Psi_{\kappa,\Lambda}$ vérifie les propriétés 1, 2 et 3.*

*Démonstration.* Par définition de la fonction, à constante multiplicative près, la condition 3 est vérifiée. Pour la condition 2, remarquons que

$$\mathbb{R}_+ \to \mathbb{R} \ : \ x \mapsto \frac{\Psi_{\kappa,\Lambda}(\xi)}{\xi} = \left(\sqrt{\kappa} + \sqrt{\Lambda}\right) \sqrt{\ln \Lambda} \left(\sqrt{\ln \frac{1}{\xi \wedge 1} \vee \frac{\kappa}{(\Lambda+1)\ln \Lambda}} + 1\right)$$

ne dépend de $\xi$ qu'au travers du terme $\sqrt{\ln \frac{1}{\xi \wedge 1}}$ qui est une fonction décroissante.

Pour la condition 1, notons que la fonction $f \ : \ \mathbb{R}_+ \to \mathbb{R} \ : \ x \to x \left(\ln \frac{1}{x \wedge 1} + 1\right)$ est croissante. En effet si $x \geq 1$, c'est évident. Si $x < 1$, $f'(x) = \ln \frac{1}{x} - 1 + 1 = \ln \frac{1}{x} \geq 0$, donc $f$ est croissante. Nous concluons de même pour la fonction $\Psi_{\kappa,\Lambda}$.

□

**Théorème 5.4.1.** *Soient $\kappa \in \mathbb{N}^*$, $\Lambda \in \mathbb{N}^* \backslash \{1, 2\}$, notons $\hat{h}_{\kappa,\Lambda}$ l'estimateur du maximum de vraisemblance de l'ensemble $\mathcal{H}_{\kappa,\Lambda}$. Soit $c > 1$ une constante positive. Supposons que pour tout $\kappa > 0$, $\Lambda > 2$,*

$$pen(\kappa, \Lambda) \geq c \left(\frac{(\kappa + \Lambda)\ln \Lambda}{n} \left(\left[\frac{\kappa}{(\Lambda+1)\ln \Lambda} \vee \ln \frac{n}{[(\kappa \vee \Lambda)\ln \Lambda] \wedge n}\right] + 1\right) + \frac{\kappa + \Lambda}{n}\right). \tag{5.8}$$

*Si* $(\hat{\kappa}, \hat{\Lambda})$ *minimise le critère pénalisée :*

$$(\hat{\kappa}, \hat{\Lambda}) = \underset{\mathbb{N}^* \times \mathbb{N}^* \setminus \{1,2\}}{\operatorname{argmin}} \; crit(\kappa, \Lambda) = \underset{\mathbb{N}^* \times \mathbb{N}^* \setminus \{1,2\}}{\operatorname{argmin}} \; \gamma_n(\hat{h}_{\kappa,\Lambda}) + pen(\kappa, \Lambda).$$

*Nous obtenons alors l'inégalité oracle suivante :*

$$\mathbb{E}(d_H^2(h^*, \hat{h}_{\hat{\kappa},\hat{\Lambda}})) \leq C \left[ \inf_{\kappa,\Lambda \in \mathbb{N}^*, \Lambda > 2} \{KL(h^*, \mathcal{H}_{\kappa,\Lambda}) + pen(\kappa, \Lambda)\} + \frac{\Gamma}{n} \right],$$

*où* $\Gamma = \frac{e^{-4}}{(1-e^{-1})^2}$*, et* $C = C(c)$ *dépend de* $c$*.*

*Démonstration.* Dans toute la preuve, $\alpha$ désigne une constante positive.

Comme la fonction $\Psi_{\kappa,\Lambda}$ vérifie les propriétés 1, 2, 3, on cherche $\xi^*$ tel que

$$\sqrt{n}\,(\xi^*)^2 = \Psi_{\kappa,\Lambda}(\xi^*) \iff \xi^* = \frac{\left(\sqrt{\kappa} + \sqrt{\Lambda}\right)\sqrt{\ln \Lambda}}{\sqrt{n}} \left( \sqrt{\ln \frac{1}{\xi^* \wedge 1} \vee \frac{\kappa}{(\Lambda+1)\ln \Lambda}} + 1 \right).$$

En minorant $\xi^*$ par $\bar{\xi} = \frac{\left(\sqrt{\kappa}+\sqrt{\Lambda}\right)\sqrt{\ln \Lambda}}{\sqrt{n}}$,

$$\begin{aligned}
(\xi^*)^2 &\leq \left( \frac{\left(\sqrt{\kappa} + \sqrt{\Lambda}\right)\sqrt{\ln \Lambda}}{\sqrt{n}} \left( \sqrt{\ln \frac{1}{\bar{\xi} \wedge 1} \vee \frac{\kappa}{(\Lambda+1)\ln \Lambda}} + 1 \right) \right)^2 \\
&\leq \alpha \frac{(\kappa + \Lambda)\ln \Lambda}{n} \left( \ln \frac{1}{\left[\sqrt{(\kappa \vee \Lambda)\ln \Lambda / n}\right] \wedge 1} \vee \frac{\kappa}{(\Lambda+1)\ln \Lambda} + 1 \right) \\
&\leq \alpha \frac{(\kappa + \Lambda)\ln \Lambda}{n} \left( \ln \frac{n}{[\kappa \ln \Lambda] \wedge n} \vee \frac{\kappa}{(\Lambda+1)\ln \Lambda} + 1 \right).
\end{aligned}$$

En choisissant $\rho_{\kappa,\Lambda} = \kappa + \Lambda$, suivant le Théorème de sélection de modèle 5.2.1, nous obtenons

$$pen(\kappa, \Lambda) \geq c \left[ \frac{(\kappa + \Lambda)\ln \Lambda}{n} \left( \frac{\kappa}{(\Lambda+1)\ln \Lambda} \vee \ln \frac{n}{[(\kappa \vee \Lambda)\ln \Lambda] \wedge n} + 1 \right) + \frac{(\kappa + \Lambda)}{n} \right].$$

Le calcul de la constante $\Gamma$ dans l'inégalité oracle donne

$$\begin{aligned}
\Gamma = \sum_{\Lambda=3}^{\infty} \sum_{\kappa \in \mathbb{N}^*} e^{-\rho_{\kappa,\Lambda}} &= \sum_{\Lambda=3}^{\infty} e^{-\Lambda} \sum_{\kappa=1}^{\infty} e^{-\kappa} \\
&= \left( \frac{e^{-3}}{1 - e^{-1}} \right) \left( \frac{e^{-1}}{1 - e^{-1}} \right) \\
&= \frac{e^{-4}}{(1 - e^{-1})^2}.
\end{aligned}$$

□

**Consistance de l'estimateur** Dans notre cadre, la consistance de l'estimateur c'est la décroissance vers 0 de la pénalité quand le nombre d'observations croit vers $\infty$. C'est à dire la décroissance du risque de l'estimateur vers le risque de l'oracle quand $n$ tend vers $\infty$, à constante multiplicative près.

Il n'est pas raisonnable de supposer que le nombre de lois d'un mélange est supérieur au nombre d'observations. Ainsi, soit $\alpha > 0$, supposons $\kappa = O(n^{1-\alpha})$. De même, il n'est pas raisonnable de supposer que la taille maximale du paramètre des lois de Poisson est d'ordre plus grand que $n$. En effet, si la taille maximale des densités d'un mélange est grand, alors les $n$ observations sont très dispersées, il est difficile de retrouver les lois. Supposons alors, pour $\beta > 0$, $\Lambda = O(n^{1-\beta})$.

1. Si $\beta \leq \alpha$, alors $pen(\kappa, \Lambda) = O\left(\frac{n^{(1-\beta)+(1-\alpha)-(1-\beta)}}{n}\right) = O\left(n^{-\alpha}\right) \to 0$ quand $n \to \infty$.
2. Si $\beta > \alpha > \frac{\beta}{2}$, alors $pen(\kappa, \Lambda) = O\left(\frac{n^{(1-\alpha)+(1-\alpha)-(1-\beta)}}{n}\right) = O\left(n^{-2\alpha+\beta}\right) \to 0$ quand $n \to \infty$.

Dans ces deux cas, la consistance est assurée.

## 5.5 Applications

Pour un entier $n$ représentant la taille de l'échantillon et $X = (X_1, ..., X_n)$ les réalisations, nous faisons varier $\kappa \in [\![0, n/10]\!]$ et $\Lambda \in [\![3, \max_i(x_i)]\!]$.

Pour minimiser le critère pénalisé, il faut calculer l'estimateur du maximum de vraisemblance des ensembles $\mathcal{H}_{\kappa,\Lambda}$. Grâce à l'algorithme $EM$ défini au Chapitre 4, nous obtenons un estimateur du maximum de vraisemblance spécifié par le vecteur des paramètres $\breve{\theta} = (\breve{p}_1, ..., \breve{p}_\kappa, \breve{\lambda}_1, ..., \breve{\lambda}_\kappa)$. Mais, pour tout $k$, $\breve{\lambda}_k$ n'est pas borné. Ainsi le vecteur des paramètres de l'estimateur du maximum de vraisemblance de l'ensemble de $\mathcal{H}_{\kappa,\Lambda}$ est

$$\hat{\theta}_{\kappa,\Lambda} = (\breve{p}_1, ..., \breve{p}_\kappa, \breve{\lambda}_1 \wedge \Lambda, ..., \breve{\lambda}_\kappa \wedge \Lambda).$$

### 5.5.1 Calibration de la constante de pénalité

Pour pouvoir appliquer le Théorème 5.4.1 il reste à calibrer la pénalité, c'est-à-dire trouver une valeur pour la constante $c$. Récemment pour surmonter cette difficulté, [BM07] ont proposé une méthode mélangeant la théorie et une idée heuristique pour définir une pénalité optimale en fonction des données. Cette méthode est appliquée pour les mélanges gaussiens dans [MM11]. Cela consiste à choisir comme pénalité optimale deux fois la pénalité minimale. Cette pénalité minimale est obtenue grâce à l'heuristique de la pente.

Habituellement, la sélection de modèle se fait sur la dimension des modèles. La méthode proposée n'échappe pas à la règle. Pour les modèles de mélange, cela revient à sélectionner sur $\kappa$ le nombre de lois composant le mélange. Soit $\Lambda$ fixé, pour pouvoir appliquer l'heuristique de la pente, il faut s'assurer que l'application $\kappa \mapsto -\gamma_n(\hat{h}_{\kappa,\Lambda})$ est

linéaire à partir d'un certain seuil $\kappa_0$ à déterminer. Ensuite, nous calculons la pente, soit le coefficient de linéarité, noté $c_{min}$, et $pen(\kappa) = 2c_{min}pen(\kappa)$.

**Algorithme** En plus de sélectionner sur la dimension du modèle, nous suggérons de sélectionner aussi sur la taille $\Lambda$ des paramètres possibles pour les lois de Poisson. Nous proposons l'algorithme suivant pour sélectionner le meilleur modèle :

1. *Pour chaque* $\kappa$
   (a) Calculer $\gamma_n(\kappa, \Lambda)$.
   (b) Sélectionner $\hat{\Lambda}_\kappa = \text{argmin}_{\mathbb{N}^* \setminus \{1,2\}} crit(\kappa, \Lambda)$.
   *Problème :* Calibrer la constante de pénalité. Remarquons que $\Lambda$ n'est pas une dimension. Or, le graphe $\Lambda \mapsto -\gamma_n(\hat{h}_{\kappa,\Lambda})$, à partir d'un certain $\Lambda_0$, est constant. Appliquer l'heuristique de la pente donnerait une pénalité nulle en $\Lambda$.
   *Suggestion :* Choisir $\hat{\Lambda}_k$ comme le plus petit minimiseur de $\gamma_n(\hat{h}_{\kappa,\Lambda})$.
   (c) Calculer la borne inferieure de la pénalité à constante près (5.8).

2. *Sélectionner sur* $\kappa$
   *Calibrer la constante de pénalité :* Appliquer la méthode de l'heuristique de la pente.
   *Selectionner* $\hat{\kappa} = \text{argmin}_{\mathbb{N}^*} crit(\kappa, \hat{\Lambda}_\kappa)$.

**Cas où** $\kappa \leq \Lambda$ La loi de Poisson permet de modéliser le nombre d'occurrences d'un évènement pendant un intervalle de temps fixé. Le paramètre de la loi de Poisson est le nombre moyen d'évènements survenus pendant cet intervalle de temps.

Ainsi, dans le contexte de mélange de $\kappa$ lois de Poisson de paramètres $(\lambda_i)_{i\in\{1,\dots,\kappa\}}$, nous pouvons supposer que pour tout $i \neq j$, $|\lambda_i - \lambda_j| \geq 1$. Cela correspond à l'idée intuitive que pour différencier deux sous-populations, il faut qu'il y ait au moins en moyenne un évènement de plus qui intervient entre les 2 sous-populations. Dans le cas des observations biologiques, cela signifie que les 2 sous-populations de CTL sont différentes si une des sous-population détruit en moyenne 1 cellule cible de plus que la seconde sous-population.

Pour la suite des calculs, nous nous plaçons dans le cas où $\kappa \leq \Lambda$. Il est alors possible d'optimiser la minoration de la fonction de pénalité par

$$pen(\kappa, \Lambda) \geq c \left[ \frac{\kappa}{n} \left( \ln \frac{n}{\kappa} + \ln \Lambda \right) + \frac{\kappa + \Lambda}{n} \right].$$

### 5.5.2 Données observées

Nous avons appliqué le critère pénalisé aux observations des biologistes (Figure 4.1), nous obtenons que $\hat{\kappa} = 2$ et $\hat{\Lambda} = 7$. L'estimateur étant celui obtenu par l'algorithme $EM$ nous retrouvons donc $\hat{\theta}$ défini en (4.6). Cela nous a permis de démontrer la consistance de cet estimateur, parmi tous les estimateurs de mélange de lois possibles.

### 5.5.3 Perspectives

**Perspectives 1** Une étude numérique comparant notre critère pénalisé avec ceux déjà existant $AIC$, $BIC$... serait à réaliser.

Il serait également pertinent de rechercher dans quel cadre notre critère, qui pénalise aussi sur la taille maximale des lois de Poisson $\Lambda$ , peut se révéler meilleur par rapport aux autres critères.

**Perspectives 2** Il serait intéressant d'obtenir un résultat théorique donnant la constante de pénalité à choisir. [AM09] se sont penchés sur ce problème dans le cas de la régression des moindres carrés.

# Bibliographie


[AB+10] Louigi Addario-Berry et al. “On combinatorial testing problems”. Dans : *Ann. Statist.* 38.5 (2010), p. 3063–3092. issn : 0090-5364 (cf. p. 103).

[ACBL12] Ery Arias-Castro, Sébastien Bubeck et Gábor Lugosi. “Detection of correlations”. Dans : *Ann. Statist.* 40.1 (2012), p. 412–435. issn : 0090-5364 (cf. p. 103).

[AG13] Helen Angell et Jérôme Galon. “From the immune contexture to the Immunoscore : the role of prognostic and predictive immune markers in cancer”. Dans : *Current opinion in immunology* 25.2 (2013), p. 261–267 (cf. p. 4).

[Aka73] H. Akaike. “Information theory and an extension of the maximum likelihood principle”. Dans : *Second International Symposium on Information Theory (Tsahkadsor, 1971)*. Akadémiai Kiadó, Budapest, 1973, p. 267–281 (cf. p. 106).

[AM09] Sylvain Arlot et Pascal Massart. “Data-driven calibration of penalties for least-squares regression”. Dans : *The Journal of Machine Learning Research* 10 (2009), p. 245–279 (cf. p. 128).

[APP05] L. Alili, P. Patie et J. L. Pedersen. “Representations of the first hitting time density of an Ornstein-Uhlenbeck process”. Dans : *Stoch. Models* 21.4 (2005), p. 967–980. issn : 1532-6349 (cf. p. 45, 46).

[BA02] Kenneth P. Burnham et David R. Anderson. *Model selection and multimodel inference*. Second. A practical information-theoretic approach. Springer-Verlag, New York, 2002, p. xxvi+488. isbn : 0-387-95364-7 (cf. p. 106).

[Bab12] Charles F Babbs. “Predicting success or failure of immunotherapy for cancer : insights from a clinically applicable mathematical model.” Dans : *Am. J. Cancer Res* 2 (2012), p. 204–213 (cf. p. 4).

[Bah58] R. R. Bahadur. “Examples of inconsistency of maximum likelihood estimates”. Dans : *Sankhyā* 20 (1958), p. 207–210. issn : 0972-7671 (cf. p. 107).

[BBF12] Mariusz Bieniek, Krzysztof Burdzy et Sam Finch. “Non-extinction of a Fleming-Viot particle model”. Dans : *Probab. Theory Related Fields* 153.1-2 (2012), p. 293–332. issn : 0178-8051 (cf. p. 74).

[BCG03] C. Biernacki, G. Celeux et G. Govaert. “Choosing starting values for the EM algorithm for getting the highest likelihood in multivariate Gaussian mixture models”. Dans : *Comput. Statist. Data Anal.* 41.3-4 (2003). Recent developments in mixture models (Hamburg, 2001), p. 561–575. issn : 0167-9473 (cf. p. 97, 100).

[Beu+10] Hélène Beuneu et al. “Visualizing the Functional Diversification of CD8+ T Cell Responses in Lymph Nodes”. Dans : *Immunity* 33.3 (2010), p. 412–423 (cf. p. 84).

[BG14] Dominique Bontemps et Sébastien Gadat. “Bayesian methods for the Shape Invariant Model”. Dans : *Electronic Journal of Statistics* (2014), to appear (cf. p. 116).

[BGB12] Veit R Buchholz, Patricia Gräf et Dirk H Busch. “The origin of diversity : studying the evolution of multi-faceted CD8+ T cell responses”. Dans : *Cellular and Molecular Life Sciences* 69.10 (2012), p. 1585–1595 (cf. p. 84).

[Bie+06] Christophe Biernacki et al. “Model-based cluster and discriminant analysis with the MIXMOD software”. Dans : *Comput. Statist. Data Anal.* 51.2 (2006), p. 587–600. issn : 0167-9473 (cf. p. 86).

[BM07] Lucien Birgé et Pascal Massart. “Minimal penalties for Gaussian model selection”. Dans : *Probab. Theory Related Fields* 138.1-2 (2007), p. 33–73. issn : 0178-8051 (cf. p. 126).

[BM93] Lucien Birgé et Pascal Massart. “Rates of convergence for minimum contrast estimators”. Dans : *Probab. Theory Related Fields* 97.1-2 (1993), p. 113–150. issn : 0178-8051 (cf. p. 107).

[BMM12] Jean-Patrick Baudry, Cathy Maugis et Bertrand Michel. “Slope heuristics : overview and implementation”. Dans : *Stat. Comput.* 22.2 (2012), p. 455–470. issn : 0960-3174 (cf. p. 86).

[BMR13] T. Byczkowski, J. Małecki et M. Ryznar. “Hitting times of Bessel processes”. Dans : *Potential Anal.* 38.3 (2013), p. 753–786. issn : 0926-2601 (cf. p. 43, 44, 55).

[Boo+94] Thierry Boon et al. “Tumor antigens recognized by T lymphocytes”. Dans : *Annual review of immunology* 12.1 (1994), p. 337–365 (cf. p. 4).

[BS02] Andrei N. Borodin et Paavo Salminen. *Handbook of Brownian motion—facts and formulae*. Second. Probability and its Applications. Birkhäuser Verlag, Basel, 2002, p. xvi+672. isbn : 3-7643-6705-9 (cf. p. 43, 46, 48).

[Bud+10] Sadna Budhu et al. “CD8+ T cell concentration determines their efficiency in killing cognate antigen–expressing syngeneic mammalian cells in vitro and in mouse tissues”. Dans : *The Journal of experimental medicine* 207.1 (2010), p. 223–235 (cf. p. 22, 30).

[Bur+96] Krzysztof Burdzy et al. “Configurational transition in a fleming-viot-type model and probabilistic interpretation of laplacian eigenfunctions”. Dans : *Journal of Physics A : Mathematical and General* 29.11 (1996), p. 2633 (cf. p. 74).

[Car+09] Íris Caramalho et al. “Visualizing CTL/melanoma cell interactions : Multiple hits must be delivered for tumour cell annihilation”. Dans : *Journal of cellular and molecular medicine* 13.9b (2009), p. 3834–3846 (cf. p. 9, 12, 23, 35).

[Cat+04] Marta Catalfamo et al. “Human CD8+ T Cells Store RANTES in a Unique Secretory Compartment and Release It Rapidly after TcR Stimulation”. Dans : *Immunity* 20.2 (2004), p. 219–230 (cf. p. 31).

[Cat+09] Patrick Cattiaux et al. “Quasi-stationary distributions and diffusion models in population dynamics”. Dans : *Ann. Probab.* 37.5 (2009), p. 1926–1969. issn : 0091-1798 (cf. p. 72).

[CF13] Jennifer Couzin-Frankel. “Cancer immunotherapy”. Dans : *Science* 342.6165 (2013), p. 1432–1433 (cf. p. xiii).

[CM13] Daniel S Chen et Ira Mellman. “Oncology meets immunology : the cancer-immunity cycle”. Dans : *Immunity* 39.1 (2013), p. 1–10 (cf. p. 4).

[CMSM13] Pierre Collet, Servet Martínez et Jaime San Martín. *Quasi-stationary distributions*. Probability and its Applications (New York). Markov chains, diffusions and dynamical systems. Springer, Heidelberg, 2013, p. xvi+280. isbn : 978-3-642-33130-5 ; 978-3-642-33131-2 (cf. p. 69–72).

[CPB08] Edward A Codling, Michael J Plank et Simon Benhamou. “Random walk models in biology”. Dans : *Journal of the Royal Society Interface* 5.25 (2008), p. 813–834 (cf. p. 5).

[Cre+13] Joel Crespo et al. “T cell anergy, exhaustion, senescence, and stemness in the tumor microenvironment”. Dans : *Current opinion in immunology* 25.2 (2013), p. 214–221 (cf. p. xiv, 4, 29).

[Das+09] Jayajit Das et al. “Digital signaling and hysteresis characterize ras activation in lymphoid cells”. Dans : *Cell* 136.2 (2009), p. 337–351 (cf. p. 31).

[Dav+07] Mark M Davis et al. “T cells as a self-referential, sensory organ”. Dans : *Annu. Rev. Immunol.* 25 (2007), p. 681–695 (cf. p. 30).

[DJ04] David Donoho et Jiashun Jin. “Higher criticism for detecting sparse heterogeneous mixtures”. Dans : *Ann. Statist.* 32.3 (2004), p. 962–994. issn : 0090-5364 (cf. p. 103).

[DL10] Michael L Dustin et Eric O Long. “Cytotoxic immunological synapses”. Dans : *Immunological reviews* 235.1 (2010), p. 24–34 (cf. p. xiii, 4, 35, 83).

[DLM99] Bernard Delyon, Marc Lavielle et Eric Moulines. “Convergence of a stochastic approximation version of the EM algorithm”. Dans : *Ann. Statist.* 27.1 (1999), p. 94–128. issn : 0090-5364 (cf. p. 97).

[DLR77] A. P. Dempster, N. M. Laird et D. B. Rubin. “Maximum likelihood from incomplete data via the EM algorithm”. Dans : *J. Roy. Statist. Soc. Ser. B* 39.1 (1977). With discussion, p. 1–38. issn : 0035-9246 (cf. p. 86, 92, 93, 96, 97).

[Far+03] Mustapha Faroudi et al. “Lytic versus stimulatory synapse in cytotoxic T lymphocyte/target cell interaction : manifestation of a dual activation threshold”. Dans : *Proceedings of the National Academy of Sciences* 100.24 (2003), p. 14145–14150 (cf. p. 29).

[Gad+14] Saikrishna Gadhamsetty et al. “A General Functional Response of Cytotoxic T Lymphocyte-Mediated Killing of Target Cells”. Dans : *Biophysical journal* 106.8 (2014), p. 1780–1791 (cf. p. 37).

[Gaj07] Thomas F Gajewski. “Failure at the effector phase : immune barriers at the level of the melanoma tumor microenvironment”. Dans : *Clinical Cancer Research* 13.18 (2007), p. 5256–5261 (cf. p. 29).

[Gaj+13] Thomas F Gajewski et al. “Cancer immunotherapy strategies based on overcoming barriers within the tumor microenvironment”. Dans : *Current opinion in immunology* 25.2 (2013), p. 268–276 (cf. p. 4).

[GBDB11] Vitaly V Ganusov, Daniel L Barber et Rob J De Boer. “Killing of targets by CD8+ T cells in the mouse spleen follows the law of mass action”. Dans : *PloS one* 6.1 (2011), e15959 (cf. p. 30).

[GR09] Frederik Graw et Roland R Regoes. “Investigating CTL mediated killing with a 3D cellular automaton”. Dans : *PLoS computational biology* 5.8 (2009), e1000466 (cf. p. 29).

[Gra+98] R. L. Graham et al. “Dense packings of congruent circles in a circle”. Dans : *Discrete Math.* 181.1-3 (1998), p. 139–154. issn : 0012-365X (cf. p. 39).

[GV01] S Ghosal et A. W. van der Vaart. “Entropies and rates of convergence for maximum likelihood and Bayes estimation for mixtures of normal densities”. Dans : *Ann. Statist.* 29.5 (2001), p. 1233–1263. issn : 0090-5364 (cf. p. 86, 112, 119).

[GW00] C. R. Genovese et L. Wasserman. “Rates of convergence for the Gaussian mixture sieve”. Dans : *Ann. Statist.* 28.4 (2000), p. 1105–1127. issn : 0090-5364 (cf. p. 111, 116).

[Har+12] Tajie H Harris et al. “Generalized Lévy walks and the role of chemokines in migration of effector CD8+ T cells”. Dans : *Nature* 486.7404 (2012), p. 545–548 (cf. p. 5, 42).

[HM13] Yuji Hamana et Hiroyuki Matsumoto. “The probability distributions of the first hitting times of Bessel processes”. Dans : *Trans. Amer. Math. Soc.* 365.10 (2013), p. 5237–5257. issn : 0002-9947 (cf. p. 43).

[Hua+13] Jun Huang et al. “A Single Peptide-Major Histocompatibility Complex Ligand Triggers Digital Cytokine Secretion in CD4+T Cells”. Dans : *Immunity* 39.5 (2013), p. 846–857 (cf. p. 31).

[HW00] Douglas Hanahan et Robert A Weinberg. “The hallmarks of cancer”. Dans : *cell* 100.1 (2000), p. 57–70 (cf. p. xii).

[HW11] Douglas Hanahan et Robert A Weinberg. “Hallmarks of cancer : the next generation”. Dans : *Cell* 144.5 (2011), p. 646–674 (cf. p. xiv).

[IW81] Nobuyuki Ikeda et Shinzo Watanabe. *Stochastic differential equations and diffusion processes*. T. 24. North-Holland Mathematical Library. North-Holland Publishing Co., Amsterdam-New York ; Kodansha, Ltd., Tokyo, 1981, p. xiv+464. isbn : 0-444-86172-6 (cf. p. 43).

[Kan+00] AR Kansal et al. “Cellular automaton of idealized brain tumor growth dynamics”. Dans : *Biosystems* 55.1 (2000), p. 119–127 (cf. p. 36).

[KJ13] Michael Kalos et Carl H June. “Adoptive T cell transfer for cancer immunotherapy in the era of synthetic biology”. Dans : *Immunity* 39.1 (2013), p. 49–60 (cf. p. 4).

[KS71] Evelyn F Keller et Lee A Segel. “Model for chemotaxis”. Dans : *Journal of Theoretical Biology* 30.2 (1971), p. 225–234 (cf. p. 11).

[LMMR13] Béatrice Laurent, Clément Marteau et Cathy Maugis-Rabusseau. “Non-asymptotic detection of two-component mixtures with unknown means”. Dans : *arXiv :1304.6924 [math.ST]* (2013) (cf. p. 103).

[Mas07] Pascal Massart. *Concentration inequalities and model selection*. T. 1896. Lecture Notes in Mathematics. Lectures from the 33rd Summer School on Probability Theory held in Saint-Flour, July 6–23, 2003, With a foreword by Jean Picard. Springer, Berlin, 2007, p. xiv+337. isbn : 978-3-540-48497-4 ; 3-540-48497-3 (cf. p. 105, 106, 108, 109).

[MB11] C. Maugis et M. Bertrand. “A non asymptotic penalized criterion for Gaussian mixture model selection”. Dans : *ESAIM Probab. Stat.* 15 (2011), p. 41–68. issn : 1292-8100 (cf. p. 110, 122).

[McL+99] Geoff J McLachlan et al. “The EMMIX software for the fitting of mixtures of normal and t-components”. Dans : *Journal of Statistical Software* 4.2 (1999), p. 1–14 (cf. p. 86).

[MDP06] Daniel G Mallet et Lisette G De Pillis. “A cellular automata model of tumor-immune system interactions”. Dans : *Journal of Theoretical Biology* 239.3 (2006), p. 334–350 (cf. p. 29, 36).

[Mey13] Caroline Meynet. “An $\ell_1$-oracle inequality for the Lasso in finite mixture Gaussian regression models”. Dans : *ESAIM Probab. Stat.* 17 (2013), p. 650–671. issn : 1292-8100 (cf. p. 110).

[MK08] Geoffrey J. McLachlan et Thriyambakam Krishnan. *The EM algorithm and extensions*. Second. Wiley Series in Probability and Statistics. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, 2008, p. xxviii+359. isbn : 978-0-471-20170-0 (cf. p. 92, 97).

[MM11] Cathy Maugis et Bertrand Michel. “Data-driven penalty calibration : a case study for Gaussian mixture model selection”. Dans : *ESAIM Probab. Stat.* 15 (2011), p. 320–339. issn : 1292-8100 (cf. p. 126).

[MP00] Geoffrey McLachlan et David Peel. *Finite mixture models*. Wiley Series in Probability and Statistics : Applied Probability and Statistics. Wiley-Interscience, New York, 2000, p. xxii+419. isbn : 0-471-00626-2 (cf. p. 86, 92, 106).

[MRM13] C. Maugis-Rabusseau et B. Michel. “Adaptive density estimation for clustering with Gaussian mixtures”. Dans : *ESAIM Probab. Stat.* 17 (2013), p. 698–724. issn : 1292-8100 (cf. p. 86, 110).

[MV12] Sylvie Méléard et Denis Villemonais. “Quasi-stationary distributions and population processes”. Dans : *Probab. Surv.* 9 (2012), p. 340–410. issn : 1549-5787 (cf. p. 69).

[Nae+07] Dieter Naeher et al. “A constant affinity threshold for T cell tolerance”. Dans : *The Journal of experimental medicine* 204.11 (2007), p. 2553–2559 (cf. p. 31).

[PB82] AS Perelson et GI Bell. “Delivery of lethal hits by cytotoxic T lymphocytes in multicellular conjugates occurs sequentially but at random times.” Dans : *The Journal of Immunology* 129.6 (1982), p. 2796–2801 (cf. p. 37, 83).

[PTSV87] M Poenie, RY Tsien et AM Schmitt-Verhulst. “Sequential activation and lethal hit measured by [Ca2+] i in individual cytolytic T cells and targets.” Dans : *The EMBO journal* 6.8 (1987), p. 2223 (cf. p. 29, 83).

[Pur+04] Marco A Purbhoo et al. “T cell killing does not require the formation of a stable mature immunological synapse”. Dans : *Nature immunology* 5.5 (2004), p. 524–530 (cf. p. 29).

[Red81] Richard Redner. “Note on the consistency of the maximum likelihood estimate for nonidentifiable distributions”. Dans : *Ann. Statist.* 9.1 (1981), p. 225–228. issn : 0090-5364 (cf. p. 97).

[Rot+78] Thomas L Rothstein et al. “Cytotoxic T lymphocyte sequential killing of immobilized allogeneic tumor target cells measured by time-lapse microcinematography”. Dans : *The Journal of Immunology* 121.5 (1978), p. 1652–1656 (cf. p. 37, 83).

[RY99] Daniel Revuz et Marc Yor. *Continuous martingales and Brownian motion*. Third. T. 293. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Springer-Verlag, Berlin, 1999, p. xiv+602. isbn : 3-540-64325-7 (cf. p. 43).

[Sch78] Gideon Schwarz. “Estimating the dimension of a model”. Dans : *Ann. Statist.* 6.2 (1978), p. 461–464. issn : 0090-5364 (cf. p. 106).

[SOS11] Robert D Schreiber, Lloyd J Old et Mark J Smyth. “Cancer immunoediting : integrating immunity’s roles in cancer suppression and promotion”. Dans : *Science* 331.6024 (2011), p. 1565–1570 (cf. p. xiv, 4, 6, 22, 29, 35).

[Spr+13] Stefani Spranger et al. “Up-regulation of PD-L1, IDO, and Tregs in the melanoma tumor microenvironment is driven by CD8+ T cells”. Dans : *Science translational medicine* 5.200 (2013), 200ra116–200ra116 (cf. p. 4, 29).

[Tei54] H. Teicher. “On the convolution of distributions”. Dans : *Ann. Math. Statistics* 25 (1954), p. 775–778. issn : 0003-4851 (cf. p. 90).

[Tei61] H. Teicher. “Identifiability of mixtures”. Dans : *Ann. Math. Statist.* 32 (1961), p. 244–248. issn : 0003-4851 (cf. p. 90).

[TSM85] D. M. Titterington, A. F. M. Smith et U. E. Makov. *Statistical analysis of finite mixture distributions*. Wiley Series in Probability and Mathematical Statistics : Applied Probability and Statistics. John Wiley & Sons, Ltd., Chichester, 1985, p. x+243. isbn : 0-471-90763-4 (cf. p. 86).

[VE05] Salvatore Valitutti et Nicolas Espagnolle. “Immunological Synapse”. Dans : *eLS* (2005) (cf. p. xiv).

[Vil13] Denis Villemonais. “General approximation method for the distribution of Markov processes conditioned not to be killed”. Dans : *arXiv :1106.0878v3* (2013) (cf. p. 74).

[Wie+06] Aurelie Wiedemann et al. “Cytotoxic T lymphocytes kill multiple targets simultaneously via spatiotemporal uncoupling of lytic and stimulatory synapses”. Dans : *Proceedings of the National Academy of Sciences* 103.29 (2006), p. 10985–10990 (cf. p. 29, 37, 83).

[Wu83] C.-F. Jeff Wu. “On the convergence properties of the EM algorithm”. Dans : *Ann. Statist.* 11.1 (1983), p. 95–103. issn : 0090-5364 (cf. p. 96).

[YYS13] V.I. Yukalov, E.P. Yukalova et D Sornette. “Utility rate equations of group population dynamics in biological and social systems”. Dans : *PLoS One* 8.12 (2013), e83225 (cf. p. 9).