



**HAL**  
open science

## Modélisation et estimation des paramètres d'un modèle multi-chaîne cachée de la fièvre typhoïde à Mayotte

Ibrahim Bouzalmat, Benoîte de Saporta, Solym Manou-Abi

### ► To cite this version:

Ibrahim Bouzalmat, Benoîte de Saporta, Solym Manou-Abi. Modélisation et estimation des paramètres d'un modèle multi-chaîne cachée de la fièvre typhoïde à Mayotte. 54es Journées de Statistique de la SFDS, la Société Française de Statistique (SFdS), Jul 2023, Bruxelles, Belgique. hal-04424077

**HAL Id: hal-04424077**

**<https://hal.science/hal-04424077v1>**

Submitted on 31 Jan 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# MODÉLISATION ET ESTIMATION DES PARAMÈTRES D'UN MODÈLE MULTI-CHAÎNE CACHÉE DE LA FIÈVRE TYPHOÏDE À MAYOTTE

Ibrahim Bouzalmat<sup>1</sup> & Benoîte de Saporta<sup>1</sup> & Solym Manou-Abi<sup>2</sup>

<sup>1</sup> *IMAG, Univ Montpellier, CNRS, Montpellier, France,  
ibrahim.bouzalmat@umontpellier.fr , benoite.de-saporta@umontpellier.fr*

<sup>2</sup> *IMAG, Univ Montpellier, CNRS, CUFR Mayotte, France,  
solym.manou-abi@univ-mayotte.fr*

**Résumé.** L'objectif de ce travail est de proposer et d'étudier un nouveau modèle épidémiologique décrivant la dynamique de la transmission de la typhoïde à Mayotte à partir des caractéristiques médicales de cette maladie et d'un jeu de données d'hospitalisations fourni par l'Agence Régionale de Santé. Nous présentons une approche paramétrique qui consiste à utiliser un processus Markovien à deux compartiments pour comptabiliser les personnes exposées et infectées par la maladie, et à estimer les paramètres du modèle, tels que les taux de contamination de personne à personne, contamination par l'environnement, incubation et guérison. Notre méthodologie d'estimation est innovante pour deux raisons principales. Tout d'abord, les observations ne sont pas disponibles en temps continu, mais seulement à des dates fixes (hospitalisations journalières), ensuite, seuls les cumuls des nouveaux cas déclarés sont comptabilisés, ce qui rend l'estimation des paramètres plus complexe. Afin de surmonter les spécificités des données disponibles, des expressions explicites des estimateurs des paramètres sont obtenues en utilisant les moments du processus exposé-infecté. Ensuite, l'algorithme de Baum-Welch est adapté au cas de multi-chaîne de Markov cachée pour estimer ces paramètres.

**Mots-clés.** Propagation de la fièvre typhoïde, Modèle multi-chaîne cachée, processus de saut pur, Estimation paramétrique, Algorithme de Baum-Welch.

**Abstract.** The objective of this work is to propose and study a new epidemiological model describing the dynamics of typhoid transmission in Mayotte using the medical characteristics of this disease and a hospitalization dataset provided by the Regional Health Agency. We present a parametric approach that consists in using a two-compartments Markovian process to count the persons exposed and infected by the disease, and to estimate the parameters of the model, such as the rates of person-to-person contamination, environmental contamination, incubation and recovery. Our estimation methodology is innovative for two main reasons. First, observations are not available continuously but only at fixed dates (daily hospitalizations), and secondly, only the accumulations of newly reported cases are counted, which makes the estimation of parameters more complex. To overcome these data-specific issues, explicit expressions of parameter estimators are obtained using the moments of the exposed-infected process. Then, the Baum-Welch algorithm is adapted to the case of hidden multi-chain Markov model to estimate these parameters.

**Keywords.** Typhoid fever propagation, Hidden multi-chain model, pure jump process, Parametric estimation, Baum-Welch algorithm.

# 1 Introduction

La fièvre typhoïde est une maladie endémique à Mayotte, un département qui représente un contexte particulier de précarisation où l'accès à l'eau courante est restreint et les conditions hygiènes sont insuffisantes. Ce type de maladie est caractérisée par une durée d'incubation, c'est-à-dire la période entre la contamination par la bactérie *Salmonella Typhi* et l'apparition des premiers symptômes. Elle se transmet soit par une voie directe de personne à personne ou une voie indirecte par des bactéries environnementales par l'intermédiaire d'eau. Nous modélisons la dynamique de sa propagation à l'aide d'un processus Markovien bidimensionnel de saut pur comptabilisant les personnes exposées et infectées par la maladie. L'objectif est alors d'estimer les paramètres de ce modèle exposé-infecté : les taux de contamination de personne à personne, contamination par l'environnement, incubation et guérison. Les particularités de notre approche et la complexité du problème proviennent de deux facteurs. Tout d'abord, les observations ne sont pas disponibles en temps continu, mais seulement à des dates fixes (hospitalisations journalières). En outre, à ces dates, le nombre total de personnes exposées ou bien infectées n'est pas observé, seuls les cumuls des nouveaux cas déclarés sont comptabilisés.

Dans notre méthode d'estimation, la première étape consiste à étudier les moments du processus bidimensionnel exposé-infecté afin d'obtenir une forme analytique de leurs espérances en fonction des paramètres. Sous certaines conditions de stabilité, il est possible d'inverser le système d'espérances obtenu pour exprimer les paramètres en termes des moments. La deuxième étape prend en compte le fait que seules les nouveaux cas isolés sont comptabilisées, nous nous plaçons dans le cadre de multi-chaîne de Markov cachée ou modèle de Markov caché couplé qui est considéré comme une extension du modèle de Markov caché classique. Cette extension permet de capturer les dépendances complexes entre les différentes séquences d'états, ce qui les rend particulièrement utiles dans de nombreuses applications (P. Kumar et al (2017), S. Ghosh et al (2017), R. Sarno et al (2017)). Ensuite, nous procédons à l'adaptation de l'algorithme de Baum-Welch (Baum et Petrie (1966)), qui permet d'estimer la probabilité de transition de la chaîne de Markov cachée couplée à partir d'un schéma d'observations non standard, puis, nous estimons les moments du processus à l'aide de simulations de Monte Carlo et utilisons les formules analytiques obtenues lors de la première étape pour obtenir des estimations des paramètres du modèle exposé-infecté.

Ce document est organisé comme suit, dans la section 2, nous décrivons la formulation du modèle stochastique exposé-infecté, ainsi que ses propriétés mathématiques. Dans la section 3, nous présentons le schéma d'observations et ses caractéristiques. Nous dérivons la procédure d'estimation des paramètres de notre modèle en adaptant l'algorithme de Baum-Welch à nos observations dans la section 4. Dans la section 5, nous mettons en place des simulations numériques pour valider et d'illustrer les résultats obtenus.

## 2 Modèle et propriété

La prévalence de la fièvre typhoïde à Mayotte se caractérise par des niveaux faibles et fluctuants Fig.1. Il est donc approprié de modéliser ces données avec un processus stochastique

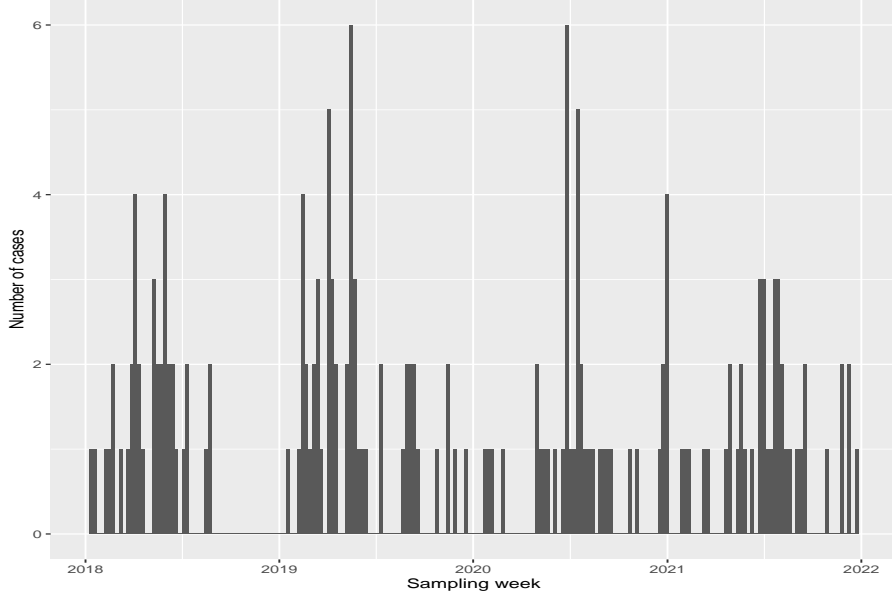


FIGURE 1 – Nombre hebdomadaire de nouveaux cas déclarés de fièvre typhoïde entre 2018 et 2021 à Mayotte (données Agence Régionale de Santé).

de comptage de type saut pur. Dans la suite, nous notons par  $(E_t, I_t)_{t \geq 0}$  le processus de Markov bidimensionnel en temps continu comptabilisant le nombre des individus exposés et infectés à l’instant  $t$  au sein de la population. La dynamique de la maladie est caractérisée par une phase d’incubation, de contagiosité et de guérison. Pendant la phase d’incubation, l’individu exposé (il a été en contact avec la maladie mais n’est pas encore infecté) ne présente pas encore de symptômes de la maladie, et il devient infecté après une durée moyenne  $1/\alpha$ , où  $\alpha > 0$  est le taux individuel d’incubation. Ensuite, il y a une phase contagieuse ou de l’infection où l’individu infecté devient contagieux et peut transmettre la maladie à un taux  $\lambda > 0$  à une autre personne. La phase de guérison où l’individu se rétablit à un taux  $\mu > 0$ . La contamination via l’environnement génère également de nouveaux infectés via un processus de Poisson de taux  $\nu > 0$ . Ainsi, étant donné l’état courant  $(e, i)$ , les transitions possibles du notre processus sont :

- $(e, i) \rightarrow (e + 1, i)$  avec un taux de  $\lambda i + \nu$ .
- $(e, i) \rightarrow (e - 1, i + 1)$  avec un taux de  $\alpha e$ .
- $(e, i) \rightarrow (e, i - 1)$  avec un taux de  $\mu i$ .

Nous étudions maintenant quelques propriétés mathématiques du processus  $(E_t, I_t)_{t \geq 0}$ .

**Théorème 1** (*Comportement moyen de  $(E_t, I_t)_{t \geq 0}$* ). *Pour tout  $e_0, i_0 \geq 0$ , si  $\lambda = \mu$ , alors*

$$\begin{cases} \mathbb{E}_{(e_0, i_0)}[E_t] = \frac{\lambda \nu}{\lambda + \alpha} t - c \left( e^{-(\lambda + \alpha)t} + \frac{\lambda}{\alpha} \right) + i_0 \frac{\lambda}{\alpha} + \frac{\nu}{\lambda + \alpha}, \\ \mathbb{E}_{(e_0, i_0)}[I_t] = \frac{\alpha \nu}{\lambda + \alpha} t + c \left( e^{-(\lambda + \alpha)t} - 1 \right) + i_0, \end{cases}$$

et si  $\lambda \neq \mu$ , alors

$$\begin{cases} \mathbb{E}_{(e_0, i_0)}[E_t] = c_1 e^{-\frac{\mu - \alpha + \sqrt{(\mu - \alpha)^2 + 4\alpha\lambda}}{2}t} + c_2 e^{-\frac{\mu + \alpha + \sqrt{(\mu - \alpha)^2 + 4\alpha\lambda}}{2}t} + \frac{\mu\nu}{\alpha(\mu - \lambda)}, \\ \mathbb{E}_{(e_0, i_0)}[I_t] = c_3 e^{-\frac{\mu - \alpha + \sqrt{(\mu - \alpha)^2 + 4\alpha\lambda}}{2}t} + c_4 e^{-\frac{\mu + \alpha + \sqrt{(\mu - \alpha)^2 + 4\alpha\lambda}}{2}t} + \frac{\nu}{\mu - \lambda}, \end{cases}$$

où

$$\begin{aligned} c &= \left( e_0 + i_0 - \frac{\nu}{\lambda + \alpha} \right) \frac{\alpha}{\lambda + \alpha}, \\ c_1 &= \frac{1}{a} \left( i_0 + \frac{\mu - \alpha + a}{2} e_0 - \frac{\nu}{\mu - \lambda} \left( 1 + \frac{\mu(\mu - \alpha + a)}{2\alpha} \right) \right), \\ c_2 &= e_0 - c_1 - \frac{\mu\nu}{\alpha(\mu - \lambda)}, \\ c_3 &= \frac{-\mu + \alpha + a}{2a} \left( i_0 + \frac{\mu - \alpha + a}{2} e_0 - \frac{\nu}{\mu - \lambda} \left( 1 + \frac{\mu(\mu - \alpha + a)}{2\alpha} \right) \right), \\ c_4 &= -\frac{\mu - \alpha + a}{2} c_2, \\ a &= \sqrt{(\mu - \alpha)^2 + 4\alpha\lambda}. \end{aligned}$$

**Corollaire 1** *Asymptotiquement, dans le cas où  $\lambda \geq \mu$ , le comportement moyen du processus  $(E_t, I_t)_{t \geq 0}$  est explosif. Lorsque  $\lambda > \mu$ , les nombres moyens des individus exposés et infectés augmentent de manière exponentielle au fil du temps. Si  $\lambda = \mu$ , les deux moyennes augmentent de manière linéaire. Cependant, le processus moyen se stabilise dans le cas où  $\lambda < \mu$  et*

$$\begin{cases} \lim_{t \rightarrow +\infty} \mathbb{E}_{(e_0, i_0)}[E_t] = \frac{\alpha\nu}{\mu(\mu - \lambda)}, \\ \lim_{t \rightarrow +\infty} \mathbb{E}_{(e_0, i_0)}[I_t] = \frac{\nu}{\mu - \lambda}. \end{cases}$$

**Théorème 2** *(Comportement asymptotique de  $(E_t, I_t)_{t \geq 0}$ ). Si  $\lambda < \alpha < \mu$  et  $\lambda < \frac{\alpha(\mu - \alpha)}{\mu}$ , alors  $(E_t, I_t)_{t \geq 0}$  est récurrent positif et admet une unique loi invariante.*

### 3 Observations

Les nombres d'individus exposés  $E_t$  et infectés  $I_t$  ne sont pas directement observable. Nous avons à notre disposition uniquement les cumuls de nouveaux cas déclarés par jour, correspondant aux sauts d'amplitude -1 du processus  $I_t$ . Ainsi, nous notons par  $O_{]s, t]}$  le nombre de sauts d'amplitude -1 du processus  $(I_t)_{t \geq 0}$  sur l'intervalle  $]s, t]$ . Les probabilités de transition  $p_{(e, i), (e', i', o)}(t) := \mathbb{P}((E_t, I_t) = (e', i'), O_{]0, t]} = o | (E_0, I_0) = (e, i))$  du processus joint  $(E_t, I_t, O_{]0, t]})_{t \geq 0}$  satisfont l'équation de Kolmogorov suivante : si  $i > i' + o$ , alors  $p_{(e, i), (e', i', o)} = 0$  et si  $i \leq i' + o$  alors

$$\begin{aligned} \frac{d}{dt} p_{(e,i),(e',i',o)}(t) &= (\lambda i + \nu) p_{(e+1,i),(e',i',o)}(t) + \alpha e p_{(e-1,i+1),(e',i',o)}(t) + \mu i p_{(e,i-1),(e',i',o)}(t) \\ &\quad - ((\lambda + \mu)i + \alpha e + \nu) p_{(e,i),(e',i',o)}(t). \end{aligned}$$

avec la condition initiale  $p_{(e,i),(e',i',o)}(0) = \delta_{i=i'} \delta_{o=0}$ . Cette équation n'a pas de solution analytique, ce qui nous empêche d'exprimer la distribution marginale des observations en fonction des paramètres  $\lambda, \mu, \alpha$  et  $\nu$ . Par conséquent, une méthode d'estimation directe par maximum de vraisemblance n'est pas réalisable avec ce schéma d'observations. Pour résoudre cette problématique, nous allons adopter une approche en deux étapes. Tout d'abord, nous allons aborder le cas où les variables  $E_t$  et  $I_t$  sont observées à chaque instant  $t$ . Ensuite, nous allons inclure ce cadre d'observation dans la catégorie des modèles de Markov cachés.

Fixons un pas de temps  $\Delta t$  ( $\Delta t = 1$  jour pour nos données), et notons par  $O_n := O_{]n\Delta t, (n+1)\Delta t]}$  le nombre des nouveaux déclarés pendant l'intervalle de temps  $]n\Delta t, (n+1)\Delta t]$ . Nous définissons pour la suite le processus  $N_t := \sum_{k=0}^{\lfloor t/\Delta t \rfloor} O_k$  décrivant le nombre cumulé de nouveaux cas isolés de 0 à l'instant  $t$ . Nous étudions le comportement moyen du processus joint  $(E_t, I_t, N_t)$ , nous sommes en mesure d'exprimer explicitement leurs moments en fonction des paramètres  $\lambda, \mu, \alpha$  et  $\nu$ . Sous la condition de stabilité mentionnée dans le corollaire 1, le système d'espérances obtenu est inversible, ce qui nous permet de déduire une expression explicite des paramètres en terms des moments du processus  $(E_t, I_t, N_t)$ .

$$\begin{aligned} \lambda &= \frac{N^* \left( \frac{R^*}{E^* I^*} - 1 \right) (E^* + I^*)}{I^* \left( 1 + \left( \frac{R^*}{E^* I^*} - 1 \right) (E^* + I^*) \right)}, \\ \mu &= \frac{N^*}{I^*}, \\ \alpha &= \frac{N^*}{E^*}, \\ \nu &= I^* (\mu - \lambda), \end{aligned} \tag{1}$$

où

$$N^* = \lim_{t \rightarrow +\infty} \frac{\mathbb{E}_{(e_0, i_0, n_0)}[N_t]}{t}, \quad E^* = \lim_{t \rightarrow +\infty} \mathbb{E}_{(e_0, i_0, n_0)}[E_t], \quad I^* = \lim_{t \rightarrow +\infty} \mathbb{E}_{(e_0, i_0, n_0)}[I_t],$$

et

$$R^* = \lim_{t \rightarrow +\infty} \mathbb{E}_{(e_0, i_0, n_0)}[E_t I_t].$$

## 4 Estimation des paramètres

Le but de cette section est d'adapter les algorithmes du Modèle de Markov caché (MMC) afin d'estimer les probabilités de transition de la chaîne cachée  $(E_n, I_n)_{n \in \mathbb{N}} := (E_{n\Delta t}, I_{n\Delta t})_{n \in \mathbb{N}}$  à partir des observations  $(O_n)_{n \in \mathbb{N}}$ . Ensuite, nous utiliserons ces estimations pour obtenir les valeurs  $E^*, I^*$  et  $R^*$  à l'aide d'une méthode de Monte Carlo, et donc estimer les paramètres  $\lambda, \mu, \alpha$  et  $\nu$  par plug-in. Dans notre modèle de Markov caché exposé-infecté (Figure 2), l'état caché des infectés à chaque instant  $n$  dépend de l'état de sa propre chaîne et de la chaîne de

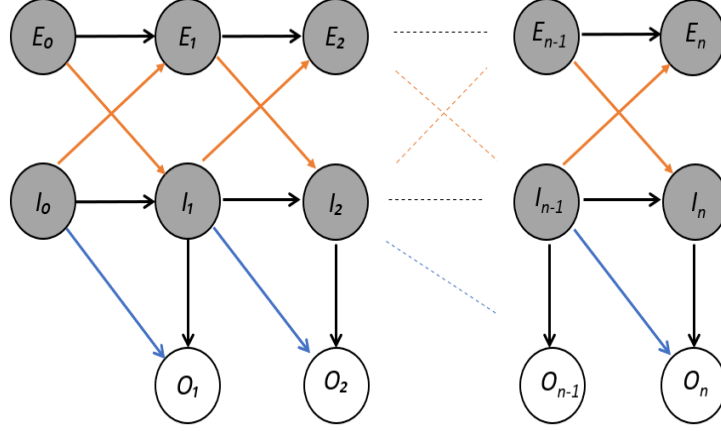


FIGURE 2 – modèle de Markov caché exposé-infecté.

couplage (des exposés) à l'instant  $n - 1$ . Cependant, la valeur observée  $O_n$  à chaque instant dépend non seulement de l'état actuel  $I_n$ , mais aussi de l'état précédent  $I_{n-1}$ , ce qui représente un schéma d'émission non standard.

Nous pouvons ramener notre Modèle de Markov caché Couplé (MMCC) à un MMC avec un schéma d'observations standard en introduisant la chaîne tridimensionnelle  $(Z_n)_{n \in \mathbb{N}^*} = (E_{n-1}, I_{n-1}, I_n)_{n \in \mathbb{N}^*}$ , qui représente une chaîne de Markov cachée à valeurs dans l'espace d'état  $\mathbb{N}^3$ . Cette transformation permet de démontrer que la séquence  $(Z_n, O_n)$  est un modèle de Markov caché standard avec les caractéristiques suivantes.

i) La matrice de probabilité de transition  $Q$  de la chaîne cachée  $(Z_n)$  est

$$Q_{(e,i,j),(e',i',j')} = \frac{P(e',i'),(\bullet,j')P(e,i),(e',i')}{P(e,i),(\bullet,j)} \delta_{i'=j} \quad (e, i, j), (e', i', j') \in \mathbb{N}^3,$$

où  $p_{(e,i),(e',i')} = p_{(e,i),(e',i')}(\Delta t)$  est la matrice de transition de la chaîne  $(E_n, I_n)_{n \in \mathbb{N}}$  et  $p_{(e,i),(\bullet,j)} = \sum_k P_{(e,i),(k,j)}$ .

ii) La probabilité d'émission  $\psi$  du processus  $O$  sachant le processus  $Z$  est

$$\psi_{(e,i,j)}(o) = \mathbb{P}(O_n = o | Z_n = (e, i, j)) = \frac{P_{(e,i),(\bullet,j,o)}}{P_{(e,i),(\bullet,j)}} \quad e, i, j, o \in \mathbb{N},$$

où  $p_{(e,i),(\bullet,j,o)} = p_{(e,i),(\bullet,j,o)}(\Delta t)$  est définie par

$$p_{(e,i),(\bullet,j,o)}(t) = \sum_k P_{(e,i),(k,j,o)}(t).$$

iii) La loi initiale  $\rho$  du processus d'état  $Z$  est

$$\rho_{e,i,j} = \mathbb{P}(Z_1 = (e, i, j)) = p_{(e,i),(\bullet,j)} \pi_{(e,i)} \quad e, i, j \in \mathbb{N},$$

où  $\pi$  est la loi initiale de la chaîne  $(E_n, I_n)_{n \in \mathbb{N}}$ .

Étant donné une suite d'observations  $(O_n)_{0 \leq n \leq T}$ , nous cherchons à déterminer par maximum

de vraisemblance les paramètres  $M = (Q, \psi, \rho)$  du modèle, et plus précisément la matrice de transition  $p$  de la chaîne couplée  $(E_n, I_n)_{n \in \mathbb{N}}$ . L'exécution directe de la procédure du MMC standard sur  $(Z_n, O_n)$  produit une grande instabilité numérique. Pour éviter ce problème, nous avons réécrit les fonctions principales de l'algorithme EM, en adaptant les formules récursives à la structure particulière de notre modèle pour obtenir des estimateurs explicites des probabilités de transition  $p$ .

**Théorème 3** *Étant donné les paramètres  $M^n = (Q^n, \psi^n, \rho^n)$ , les estimations du maximum de vraisemblance de  $M^{n+1} = (Q^{n+1}, \psi^{n+1}, \rho^{n+1})$  étant donné les observations  $(o_1, \dots, o_T)$  sont données, pour  $e, i, j, o \in \mathbb{N}$ , par*

$$\begin{aligned} Q_{(e,i,j),(e',i',j')}^{n+1} &= \frac{\sum_{t=1}^T \xi_{(e,i,j),(e',i',j')}^n(t)}{\sum_{t=1}^T \gamma_{(e,i,j)}^n(t)} \delta_{i'=j}, \\ \psi_{(e,i,j)}^{n+1}(o) &= \frac{\sum_{t=1}^T \mathbb{1}_{o_t=o} \gamma_{(e,i,j)}^n(t)}{\sum_{t=1}^T \gamma_{(e,i,j)}^n(t)}, \\ \rho_{e,i,j}^{n+1} &= \gamma_{(e,i,j)}^n(1), \\ p_{(e,i),(e',i')}^{n+1} &= \frac{\sum_{j' \in \mathbb{N}} Q_{(e,i,j),(e',j',j')}^{n+1} \rho_{e,i,j}^{n+1}}{\sum_{j \in \mathbb{N}} \rho_{e,i,j}^{n+1}}, \end{aligned}$$

où

$$\begin{aligned} \xi_{(e,i,j),(e',i',j')}^n(t) &= \frac{\alpha_{(e,i,j)}^n(t) p_{(e',i'),(\bullet,j')} p_{(e,i),(e',i')} \psi_{(e',i',j')}^n(o_{t+1}) \beta_{(e',i',j')}^n(t+1)}{p_{(e,i),(\bullet,j)} \sum_{(e,i,j) \in \mathbb{N}^3} \alpha_{(e,i,j)}^n(t) \beta_{(e,i,j)}^n(t)} \delta_{i'=j}, \\ \gamma_{(e,i,j)}^n(t) &= \frac{\alpha_{(e,i,j)}^n(t) \beta_{(e,i,j)}^n(t)}{\sum_{(e,i,j) \in \mathbb{N}^3} \alpha_{(e,i,j)}^n(t) \beta_{(e,i,j)}^n(t)}, \end{aligned}$$

et les probabilités forward et backward  $\alpha^n$  et  $\beta^n$  sont définies par

Pour tout  $e, i, j \in \mathbb{N}$  et  $2 \leq t \leq T$

$$\begin{cases} \alpha_{(e,i,j)}(1) = p_{(e,i),(\bullet,j)} \pi_{(e,i)} \psi_{(e,i,j)}(o_1), \\ \alpha_{(e,i,j)}(t) = \psi_{(e,i,j)}(o_t) \sum_{(e',i') \in \mathbb{N}^2} \frac{p_{(e,i),(\bullet,j)} p_{(e',i'),(e,i)}}{p_{(e',i'),(\bullet,i)}} \alpha_{(e',i',i)}(t-1). \end{cases}$$

Pour  $e, i, j \in \mathbb{N}$  et  $1 \leq t \leq T-1$

$$\begin{cases} \beta_{(e,i,j)}(T) = 1, \\ \beta_{(e,i,j)}(t) = \sum_{(e',j') \in \mathbb{N}^2} \frac{p_{(e',j'),(\bullet,j')} p_{(e,i),(e',j')}}{p_{(e,i),(\bullet,j)}} \psi_{(e',j',j')} (o_{t+1}) \beta_{(e',j',j')}(t+1). \end{cases}$$

À chaque itération, l'algorithme de Baum-Welch met à jour les paramètres du modèle en utilisant les formules données dans le théorème (3) et réévalue la log-vraisemblance des observations, jusqu'à ce qu'une convergence vers les valeurs optimales soit atteinte. L'obtention de la matrice de transition  $\hat{p}$  optimale permet de simuler la chaîne de Markov  $(E_n, I_n)$  avec une taille suffisamment grande de  $n$  pour récupérer les moments  $E^*$ ,  $I^*$  et  $R^*$ . Finalement, Nous utilisons (1) pour estimer les paramètres du modèle exposé-infecté  $\lambda, \mu, \alpha$  et  $\nu$ .



## 5 Étude numérique

Nous allons étudier les performances numériques des estimateurs  $\hat{\lambda}^n, \hat{\mu}^n, \hat{\alpha}^n$  et  $\hat{\nu}^n$  sur des données synthétiques. Nous commencerons par examiner le cadre où la chaîne bidimensionnelle  $(E_n, I_n)$  est observé. Pour estimer les moments  $E^*, I^*$  et  $R^*$  de la chaîne de Markov observée  $(E_n, I_n)$ , nous utilisons la méthode de Monte Carlo qui consiste à simuler un grand nombre de trajectoires de la chaîne à partir desquelles nous pouvons estimer les moments empiriques. Nous simulons plusieurs trajectoires du processus  $(E_t, I_t)$  partant d'un état initial  $(E_0, i_0) = (0, 0)$  des valeurs de paramètres fixées à  $\lambda = 0.05, \mu = 0.2, \alpha = 0.1$  et  $\nu = 0.015$ , sur différents horizons du temps  $H$ . Pour chaque trajectoire, avec un pas de temps fixé à  $\Delta t = 1j$ , nous obtenons la chaîne discrète  $(E_n, I_n)$  et les observations  $O_n$  qui décrivent le nombre de sauts négatifs du processus marginal  $I_t$  sur l'intervalle  $]n\Delta t, (n+1)\Delta t]$ . Nous utilisons ensuite les moyennes empiriques de la chaîne pour estimer les moments, puis nous obtenons des estimations des paramètres en les injectant dans la formule (1). Cette méthode nous permet d'obtenir une distribution des estimateurs, à partir de laquelle nous pouvons calculer des intervalles de confiance.

Afin de réduire le temps de calcul des simulations de Monte Carlo, nous utilisons la fonction "dopar" du package "doparallel" (calaway2015package) pour paralléliser nos calculs. Cette méthode permet de diviser le travail en plusieurs tâches indépendantes qui peuvent être exécutées simultanément, ce qui a permis de réduire considérablement le temps nécessaire pour effectuer les simulations.

Le tableau (1) présente les valeurs estimées des paramètres ainsi que leurs intervalles de confiance pour différents horizons  $H$  (nombre des observations), pour 10000 trajectoires simulées.

TABLE 1 – Estimations des paramètres et leurs intervalles de confiance sur 10000 trajectoires du processus exposé-infecté.

Nombre des observations $n = 1000$		
Les paramètres	Les estimations	Intervalle de confiance (95%)
$\lambda$	0.049	[0.047, 0.052]
$\mu$	0.1999	[0.1996, 0.2019]
$\alpha$	0.10	[0.0996, 0.1005]
$\nu$	0.0147	[0.0145, 0.0149]
Nombre des observations $n = 10000$		
$\lambda$	0.050	[0.049, 0.0506]
$\mu$	0.20	[0.1998, 0.2001]
$\alpha$	0.100	[0.0998, 0.1002]
$\nu$	0.015	[0.0149, 0.01506]

Il convient de noter que la précision des estimations dépend non seulement du nombre de trajectoires simulées, mais aussi de la longueur de chaque trajectoire. En effet, plus le nombre

de trajectoires simulées est élevé, plus la précision de l'estimation sera élevée et l'intervalle de confiance sera réduit, ce qui signifie que l'estimation des paramètres sera plus précise et plus fiable. Cependant, il est important de noter que la longueur de chaque trajectoire peut également affecter la précision de l'estimation, en particulier si la longueur est trop courte, ce qui peut entraîner une variance élevée des estimations. Il est donc important de trouver un compromis entre le nombre de trajectoires et la longueur de chaque trajectoire pour obtenir des estimations précises des paramètres avec des intervalles de confiance fiables.

## 6 Perspective

**A court terme** : Nous allons étudier la performance de nos estimateurs dans le cadre de multi-chaine cachée en mettant en œuvre le schéma itératif présenté dans le théorème 3. Nous aborderons les points de choix des conditions initiales, la régulation du nombre d'itérations et le tronçonnage de l'espace.

**A moyen terme** : Nous étendrons notre méthodologie d'estimation en prenant en compte l'effet de la pluviométrie sur l'évolution de la maladie à Mayotte. Nous examinerons les deux régimes saisonniers, à savoir la saison sèche et la saison humide, pour déterminer l'influence de ces conditions météorologiques sur notre modèle d'estimation.

## Bibliographie

- P. Kumar, H. Gauba, P. P. Roy, and D. P. Dogra (2017), Coupled hmm-based multi-sensor data fusion for sign language recognition," *Pattern Recognition Letters*, vol. 86, pp. 1-8.
- S. Ghosh, J. Li, L. Cao, and K. Ramamohanarao (2017), Septic shock prediction for icu patients via coupled hmm walking on sequential contrast patterns, *Journal of Biomedical Informatics*, vol. 66, pp. 19-31.
- R. Sarno and K. R. Sungkono (2017), Coupled hidden markov model for process discovery of nonfree choice and invisible prime tasks, *Procedia Computer Science*, vol. 124, pp. 134-141.
- L. E. Baum and T. Petrie (1966), Statistical inference for probabilistic functions of finite state markov chains, *The Annals of Mathematical Statistics*, vol. 37, no. 6, pp. 1554-1563.
- R. Calaway, R. Analytics, S. Weston, D. Tenenbaum, and M. R. Calaway (2015), Package-doparallel.