Estimation de paramètres pour un modèle de propagation 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. Estimation de paramètres pour un modèle de propagation de la fièvre typhoïde à Mayotte. Journées Math Bio Santé 2022, Oct 2022, Besançon, France. hal-04424171

HAL Id: hal-04424171
https://hal.science/hal-04424171v1

Submitted on 29 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

Estimation de paramètres pour un modèle de propagation de la fièvre typhoïde à Mayotte

Ibrahim Bouzalmat[1], Benoîte de Saporta[1], Solym Manou-Abi[1,2]
[1] *IMAG, Univ Montpellier, CNRS, Montpellier, France.* [2] *CUFR Mayotte, France*

Résumé

L'objectif de ce travail est de modéliser la propagation de la fièvre typhoïde à Mayotte à partir d'un jeu de données d'hospitalisations fourni par l'Agence Régionale de Santé. Nous utilisons un processus de naissance et mort linéaire avec immigration comptabilisant les personnes infectées par la maladie. L'objectif est alors d'estimer les taux de contamination de personne à personne, contamination par l'environnement et guérison à partir des données. L'originalité de notre approche et la difficulté du problème provient de deux sources. D'une part, les observations ne sont pas disponibles en temps continu mais seulement à des dates fixes (hospitalisations journalières), et d'autre part à ces dates le nombre total de personnes infectées n'est pas observé directement, seuls les nouveaux cas depuis la date précédente sont comptabilisés. Pour traiter ces spécificités, nous obtenons d'abord une expression explicite des lois de transition à pas fixe pour construire des estimateurs de nos paramètres basés sur les fréquences des transitions pour le nombre d'infectés. Ensuite nous nous plaçons dans le cadre des chaînes de Markov cachées et adaptons l'algorithme de Baum-Welch à notre cas.

① Contexte et Modélisation

La fièvre typhoïde est une maladie infectieuse causée par une bactérie de la famille Entérobactérie,du genre des salmonelles.

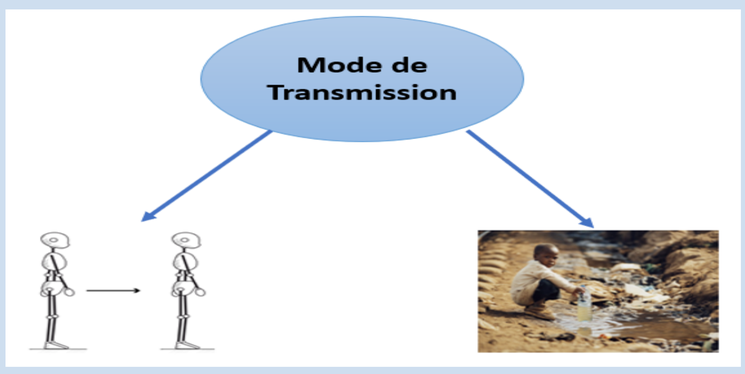

Figure 1. Mode de transmission de la Typhoide .

La fièvre typhoïde est une maladie endémique à Mayotte. Au cours de l'année 2021,35 cas de fièvre typhoïde confirmés par hémoculture et/ou coproculture ont été signalés par le laboratoire du Centre Hospitalier de Mayotte (CHM).

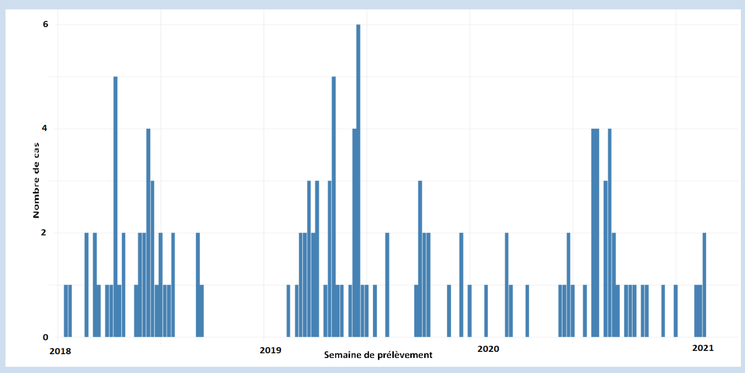

Figure 2. Nouveaux cas hebdomadaires entre 2018 et 2020 à Mayotte (ARS) .

Notre objectif est donc de proposer un modèle simple pour décrire la dynamique de transmission de la typhoïde à Mayotte et d'estimer ses paramètres à partir d'un jeu de données d'hospitalisation.
La prévalence de la fièvre typhoïde à Mayotte est caractérisée par des niveaux faibles et fluctuants, et des résurgences de la maladie après des plages sans nouvelles infections, voir Fig.2. Un modèle stochastique de type processus de naissance et mort avec immigration est donc bien adapté pour modéliser ces données.

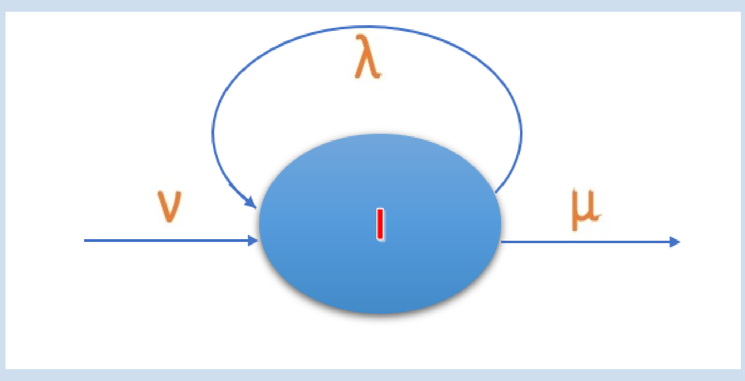

Figure 3. Diagramme du modèle .

Chaque individu contamine un nouvel individu à un taux de naissance $\boldsymbol{\lambda}$ et guérit avec un taux de mort $\boldsymbol{\mu}$ indépendamment des autres individus. La contamination via l'environnement génère également de nouveaux infectés par un taux d'immigration $\boldsymbol{\nu}$. Notre objectif est d'estimer ces trois paramètres $\boldsymbol{\lambda, \mu, \nu}$.

② Observations et Originalité du problème

Nous disposons de 1086 cumuls journaliers de nouveaux cas positifs de la fièvre typhoïde à Mayotte pour les années 2018 à 2020. Ces cumuls compris entre 0 et 3.
L'originalité de notre approche et la difficulté du problème provient de deux sources. D'une part, les observations ne sont pas disponibles en temps continu mais seulement à des dates fixes (hospitalisations journalières), et d'autre part à ces dates le nombre total de personnes infectées n' est pas observé directement, seuls les nouveaux cas depuis la date précédente sont comptabilisés. Ainsi, il n'est pas possible de mettre en oeuvre une méthode d'estimation directe par maximum de vraisemblance avec ce schéma d'observations.

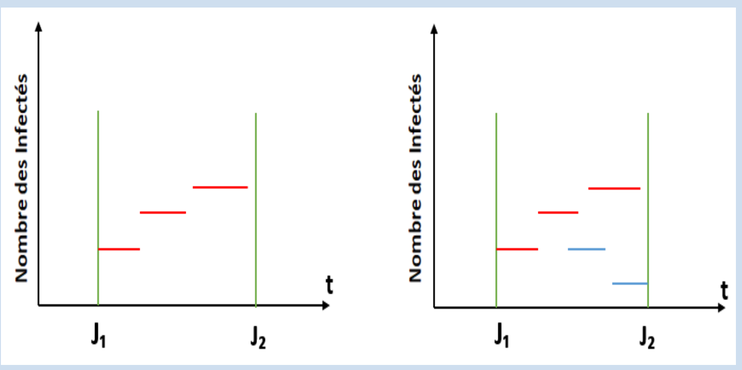

Figure 4. Les cumuls vs donnees complètes

③ Strategie de Résolution

1- Construire (analytiquement) des estimateurs de $\boldsymbol{\lambda, \nu}$ et $\boldsymbol{\mu}$ en fonction des probabilités de transition $\boldsymbol{p_{i,j}}$ du processus $\boldsymbol{I_t}$ pour la structure d'observations complètes en temps discret en résolvant une équation de Kolmogorov:

$$\boldsymbol{\lambda = \frac{ln\left(\frac{u}{p}\right)(p-1)}{\Delta t\,(p-u)}, \quad \mu = \frac{ln\left(\frac{u}{p}\right)(u-1)}{\Delta t\,(p-u)}, \quad \nu = r\lambda,}$$

avec

$$\boldsymbol{p = \frac{p_{0,1}}{p_{0,0} ln(p_{0,0})} W\left(\frac{p_{0,0}^{\left(\frac{p_{0,0}}{p_{0,1}}+1\right)} ln(p_{0,0})}{p_{0,1}}\right), \quad r = \frac{ln(p_{0,0})}{ln(p)}, \quad u = 1 - \frac{p_{1,0}}{p_{0,0}},}$$

où $\boldsymbol{W}$ est la fonction W de Lambert et $\boldsymbol{\Delta t}$ un pas de temps fixé.
2- Formuler le problème d'estimation comme un Modèle de Markov caché (MMC) non standard en temps discret: Notons $\boldsymbol{O_n}$ pour $\boldsymbol{n \in \mathbb{N}^*}$ le nombre de nouveaux infectés sur l'intervalle de temps $\boldsymbol{[(n-1)\Delta t, n\Delta t[}$ qui correspond aux observations disponibles. Ce modèle de Markov caché est caractérisé par la probabilité de transition du processus d'état $\boldsymbol{(I_{n-1}, I_n)}$: $\boldsymbol{Q_{(i,j),(i',j')} = p_{i',j'}\delta_{i'=j}}$, et la probabilité d'émission du processus $\boldsymbol{O}$ sachant le processus $\boldsymbol{(I_{n-1}, I_n)}$: $\boldsymbol{\psi_{(i,j)}(o)}$.

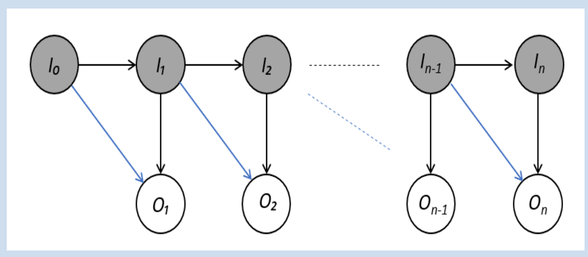

Figure 5. Schéma d'émission du MMC.

3- Construire (numériquement) en adaptant les algorithmes du MMC (type EM) les estimateurs pour notre cas, temps discret, observations cachées : Partant d'un modèle initial $\boldsymbol{M^{(0)}}$, les ré-estimations sont données par :

$$\boldsymbol{p_{i,j}^{(n)} = \frac{1}{N}\sum_{i'=0}^{N} Q_{(i',i),(i,j)}^{(n)},}$$

où $\boldsymbol{Q_{(i,j),(i',j')}^{(n)} = \frac{\sum_{t=0}^{T-1}\xi_{(i,j),(i',j')}(t)}{\sum_{t=0}^{T-1}\gamma_{(i,j)}(t)}\delta_{i'=j}}$, $\boldsymbol{\psi_{(i,j)}^{(n)}(o) = \frac{\sum_{t=0}^{T}1_{\{O_t=o\}}\gamma_{(i,j)}(t)}{\sum_{t=0}^{T}\gamma_{(i,j)}(t)}}$, $\boldsymbol{\rho_{(i,j)}^{(n)} = \gamma_{(i,j)}(0),}$
avec

$$\boldsymbol{\xi_{(i,j),(i',j')}(t) = \frac{\alpha_{(i,j)}(t)p_{(i',j')}\psi_{(i',j')}(o_{t+1})\beta_{(i',j')}(t+1)}{\mathbb{P}(O_{0:T} = o_{0:T}|M^{(n-1)})}\delta_{i'=j},\ \gamma_{(i,j)}(t) = \frac{\alpha_{(i,j)}(t)\beta_{(i,j)}(t)}{\mathbb{P}(O_{0:T} = o_{0:T}|M^{(n-1)})}.}$$

$\boldsymbol{\alpha_{(i,j)}(t)}$ et $\boldsymbol{\beta_{(i,j)}(t)}$ représentent respectivement the Forward and Backward probability.

④ Résultats

Nous choisissons les paramètres initiaux dans les intervalles suivants $\boldsymbol{\lambda^{(0)} \in [0.05, 0.08]}$, $\boldsymbol{\mu^{(0)} \in [0.11, 0.25]}$ et $\boldsymbol{\nu^{(0)} \in [0.015, 0.03]}$. Les valeurs des paramètres maximisant la vraisemblance des observations pour toutes les valeurs initiales testées sont $\boldsymbol{\hat{\lambda} = 0.05, \hat{\mu} = 0.11}$ et $\boldsymbol{\hat{\nu} = 0.015}$.
En particulier, on a $\boldsymbol{\hat{\lambda} < \hat{\mu}}$ qui correspond à un régime récurrent positif sous lequel nos estimations sont valables, et où le processus retourne à $\boldsymbol{0}$ infiniment souvent ce qui est cohérent avec les observations.
Le temps moyen de guérison $\boldsymbol{1/\hat{\mu} = 9}$ est cohérent avec les durées d'hospitalisations et de traitements.