



Modeling and parameter estimation of a hidden multi-chain model of typhoid fever in Mayotte

Ibrahim Bouzalmat¹

¹ IMAG, Montpellier Univ, France



Typhoid Fever

Typhoid fever

Typhoid fever is a bacterial infection that is caused by the bacterium *Salmonella typhi*.

► Symptoms

- Fever that starts low and increases daily
- Headache, weakness and fatigue, muscle aches
- Stomach pain, diarrhea or constipation



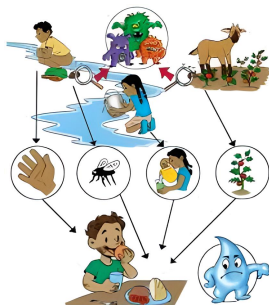
Typhoid Fever

► Incubation

- 1 to 3 weeks

► Transmission

- **direct** : Fecal-oral transmission by ingestion of bacteria from contaminated stool (dirty hands)
- **indirect** : by ingestion of water or food eaten raw (vegetables, seafood ...) and soiled by infected human stool (sewers, rainwater evacuation systems, etc.)



Typhoid Fever in Mayotte

- ▶ **Endemic** disease in Mayotte: Annual incidence varies between 35 cases (observed in 2017) and 100 cases (observed in 2022)
- ▶ **Notifiable** disease

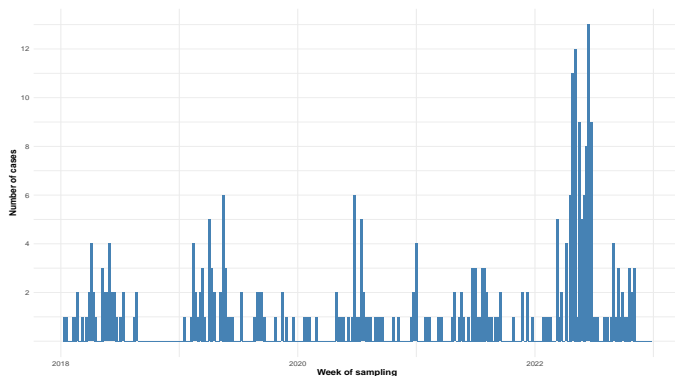
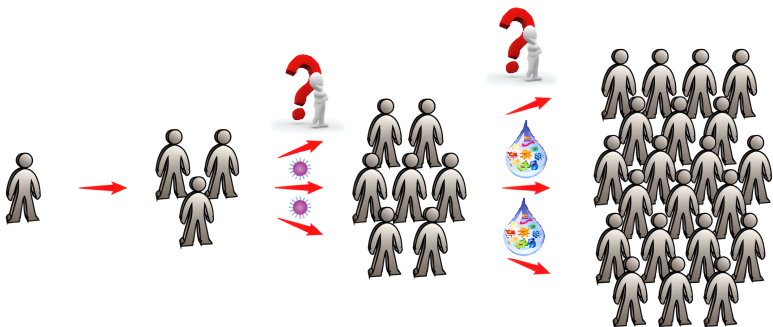


Figure: Weekly cumulative number of new reported cases of typhoid fever between 2018 and 2022 in Mayotte. (source: ARS Mayotte)

Objective

- ▶ Propose a **mathematical model** to describe the transmission dynamics of typhoid in Mayotte
- ▶ **Estimate** its parameters from a dataset of hospitalizations



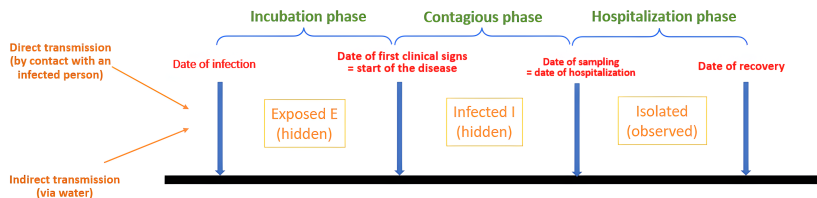
- **Population characteristics :**

- ▶ **Large** population of Susceptibles: $\sim 300\,000$ inhabitants

- ▶ **Small** population of Infected: ~ 60 cases per year on average

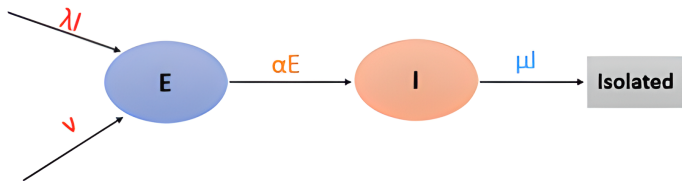
⇒ **Stochastic** model of the counting process type

- **Disease dynamics:**



⇒ Two-dimensional pure jump Markov process $(E_t, I_t)_{t \geq 0}$

• Two-compartment model : Exposed-Infected



- ▶ λ : Rate of contamination from **person to person**
- ▶ ν : Rate of exogenous **contamination**
- ▶ α : Rate of individual **incubation**
- ▶ μ : Rate of **isolation**

⇒ **Estimate** the rates describing the dynamics of typhoid fever transmission (λ , ν , α and μ)

Observations

- ▶ Number of exposed and infected: **non-observed**
- ▶ **Available observations O_n** : Daily cumulative number of new reported (isolated) cases = cumulated number of -1 jumps in the process I_t during a day
- ▶ Mathematically **complex and original**
- ▶ **Classic** in epidemiology

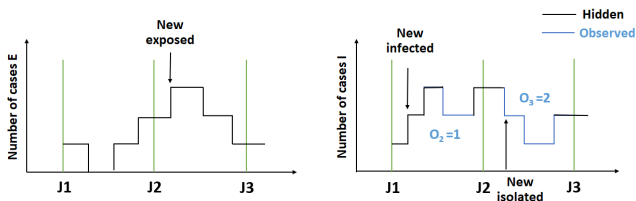


Figure: Example of a trajectory: exposed-infected

Challenges

- ▶ Multiple trajectories can generate the **same observations**
- ▶ Continuous-time model with **discrete-time observations**
- ▶ The numbers of exposed E_t and infected I_t individuals are hidden, only the **daily cumulative counts** are observed
- ▶ No direct **analytical/numerical** expression for the likelihood of observations

► Consider the tridimensional process (E_t, I_t, O_n)

Step 1 : Solve the estimation problem from **complete discrete-time observations** $(E_n, I_n, \text{ and } O_n)$

Step 2 : Utilize a **Hidden Markov Model (HMM)** framework to solve the estimation problem from **hidden discrete observations** (O_n)

Step 1: Complete Discrete Observations

Theorem: Analytical Expression of Parameters (BOUZALMAT et.al)

If $\lambda < \mu$,

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

where

$$N^* = \lim_{n \rightarrow +\infty} \frac{\mathbb{E}_{(e_0, i_0, n_0)}[N_n]}{n}, \quad E^* = \lim_{n \rightarrow +\infty} \mathbb{E}_{(e_0, i_0, n_0)}[E_n], \quad I^* = \lim_{n \rightarrow +\infty} \mathbb{E}_{(e_0, i_0, n_0)}[I_n],$$

and

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

$N_n := \sum_{k=0}^n O_k$: the cumulative number of new isolated cases from 0 to time n

- ▶ Proof idea: Kolmogorov equation
- ▶ An estimator of N^* is observed, but E^* , I^* , and R^* are hidden

- ▶ Estimate, via an **HMM**, the transition matrix of the hidden chain $(E_n, I_n)_{n \in \mathbb{N}}$, denoted $\rho_{(e,i),(e',i')}$, which is not analytically reachable
- ▶ Estimate moments E^* , I^* , R^* from $\hat{\rho}_{(e,i),(e',i')}$ using **Monte Carlo simulations**
- ▶ Construct estimators for λ , μ , α , and ν from the analytical relationship in the first step

Estimation Strategy

► Hidden multi-chain Markov model (HMCMM)

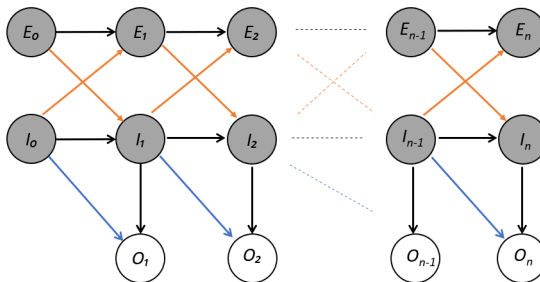


Figure: HMCMM emission scheme.

► $(Z_n)_{n \in \mathbb{N}^*} = (E_{n-1}, I_{n-1}, I_n)_{n \in \mathbb{N}^*} \Rightarrow (Z_n, O_n)$ is a **three-dimensional HMM** with a standard observation scheme whose characteristics are given by the triple $M = (Q, \psi, \rho)$

Adapted Baum-Welch Algorithm

Theorem (BOUZALMAT et.al)

Given the parameters $M^n = (Q^n, \psi^n, \rho^n)$ and observations (o_1, \dots, o_T) , the re-estimations of the HMM are given, for $e, i, j, o \in \mathbb{N}$, by

$$Q_{(e,i,j),(e',i',j')}^{n+1} = \frac{\sum_{t=1}^{T-1} \xi_{(e,i,j),(e',i',j')}^n(t)}{\sum_{t=1}^T \gamma_{(e,i,j)}^n(t)} \delta_{i'=j}$$

$$\psi_{(e,i,j)}^{n+1}(o) = \frac{\sum_{t=1}^T \mathbb{1}_{o_t=o} \gamma_{(e,i,j)}^n(t)}{\sum_{t=1}^T \gamma_{(e,i,j)}^n(t)}, \quad \rho_{e,i,j}^{n+1} = \gamma_{(e,i,j)}^n(1),$$

$$\rho_{(e,i),(e',i')}^{n+1} = \frac{\sum_{j' \in \mathbb{N}} Q_{(e,i,j),(e',j',j')}^{n+1} \rho_{e,i,j}^{n+1}}{\sum_{j \in \mathbb{N}} \rho_{e,i,j}^{n+1}},$$

$$\xi_{(e,i,j),(e',i',j')}^n(t) = \frac{\alpha_{(e,i,j)}^n(t) \rho_{(e',i'),(\cdot,j')}^n \rho_{(e,i),(e',i')}^n \psi_{(e',i',j')}^n(o_{t+1}) \beta_{(e',i',j')}^n(t+1)}{\rho_{(e,i),(\cdot,j)}^n \sum_{(e,i,j) \in \mathbb{N}^3} \alpha_{(e,i,j)}^n(t) \beta_{(e,i,j)}^n(t)},$$
$$\gamma_{(e,i,j)}^n(t) = \frac{\alpha_{(e,i,j)}^n(t) \beta_{(e,i,j)}^n(t)}{\sum_{(e,i,j) \in \mathbb{N}^3} \alpha_{(e,i,j)}^n(t) \beta_{(e,i,j)}^n(t)}.$$

Numerical Study : Complete observations

- ▶ We run a 10000 parallelized simulation of the process (E_t, I_t) starting from $(E_0, i_0) = (0, 0)$ with $\lambda = 0.05$, $\mu = 0.2$, $\alpha = 0.1$ and $\nu = 0.015$
 - Fixed time step $\Delta t = 1 \text{ day} \Rightarrow (E_n, I_n)$ and the observations O_n

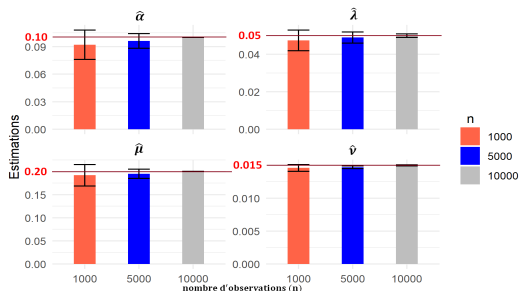


Figure: Parameter estimates and their CI for different n

⇒ The accuracy of the estimates depends on the number of simulated trajectories and the number of observations

▶ Hidden State Truncation N

- Choice of initial parameters: For each value of the quadruplet $(\lambda_i^{(0)}, \mu_i^{(0)}, \alpha_i^{(0)}, \nu_i^{(0)})$ for $i = 1, \dots, 15$:
 - Estimate initial matrices $\rho^{(0)}$ and $\psi^{(0)}$ through Monte Carlo simulations
 - Execute the adapted Baum-Welch algorithm
- Choose $\hat{p}^m = \max_{\hat{p}^n} \mathbb{P}(O|M^n)$

Numerical Study: Hidden Framework

► Number of observations: $n = 10000$

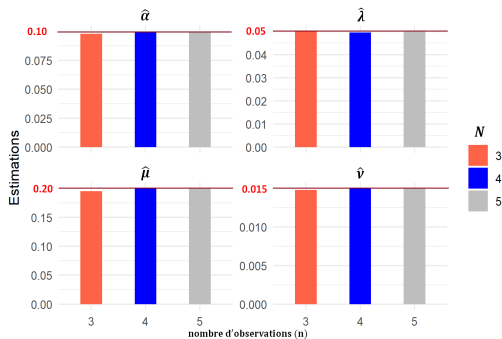


Figure: Impact of truncation N on parameter estimates

● Positive Aspects:

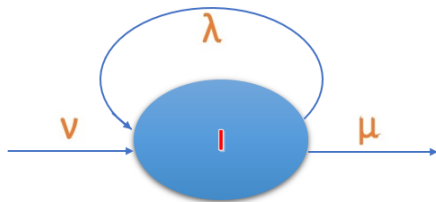
- Truncating the state space **does not have a significant impact** on the estimation.
- The **quality of parameter estimation** primarily depends on the **quantity of available data**:
 - Improves local convergence of EM
 - Enhances the estimation of N^* , which crucially depends on the number of observations.

● Limitations:

- Tuning of parameters.
- The algorithm's complexity may result in longer execution times.
- The dataset provided by the ARS of Mayotte proved insufficient for obtaining precise estimates of the **model parameters**.

Simplified model

- Linear birth-death-immigration (LBDI) model



⇒ **Aim:** Estimate the contamination and isolation rates (λ , ν and μ) from hospitalization data for Mayotte

▶ **Kolmogorov equation:** Explicit expression of λ , ν and μ as functions of $P_{0,0}(\Delta t)$, $P_{0,1}(\Delta t)$ and $P_{1,0}(\Delta t)$

$$(\lambda, \nu, \mu) = f(P_{0,0}(\Delta t), P_{0,1}(\Delta t), P_{1,0}(\Delta t))$$

Results on Hospitalization Data in Mayotte

- ▶ We have 1816 observations (299 reported cases over 5 years)

Table: Number of new reported cases in Mayotte between 2018 and 2022

Observation Value	0	1	2	3	4
Number of Observations	1596	167	34	12	7

Table: Transitions between States

Transition	$0 \rightarrow 0$	$0 \rightarrow 1$	$1 \rightarrow 0$	$1 \rightarrow 1$	$1 \rightarrow 2$
(%)	79.62	7.05	7.60	1.15	0.33

- ▶ Initialization: $\lambda^{(0)} \in [0.05, 0.08]$, $\mu^{(0)} \in [0.11, 0.25]$, and $\nu^{(0)} \in [0.015, 0.03]$

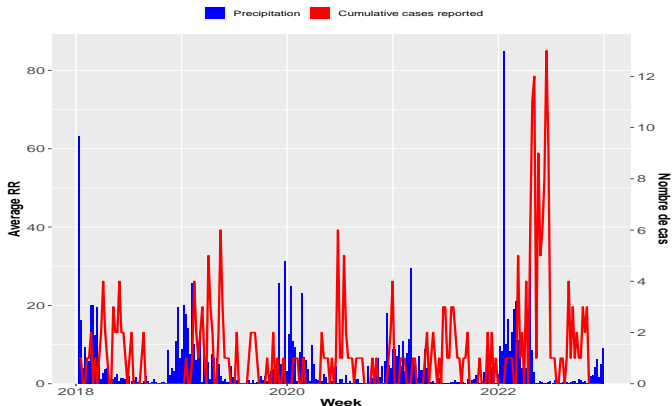
⇒ The estimated parameters are

$$\hat{\lambda} = 0.054, \quad \hat{\mu} = 0.132 \quad \text{and} \quad \hat{\nu} = 0.017$$

- ▶ $\hat{\lambda} < \hat{\mu}$: consistent with observations
- The average isolation duration is 7.5 days

Perspective

- ▶ Deepen our understanding of the impact of variations in rainfall levels on the development of typhoid fever
- ▶ Get additional data: date of first clinical signs, date of notification, date of hospitalisation, etc.



thank you

A row of ten light-colored wooden blocks, each with a single lowercase letter, spelling out the words 'thank you'. The blocks are arranged on a wooden surface. The background is a soft-focus bokeh of warm, golden light spots.