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

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

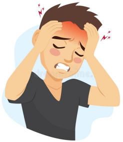
5-7 June 2023

Typhoid Fever

Typhoid fever

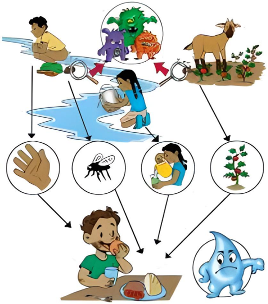
Typhoid fever is a bacterial infection that is caused by the bacterium Salmonella typhi.

- 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: around 100 cases in 2022.
- Notifiable disease

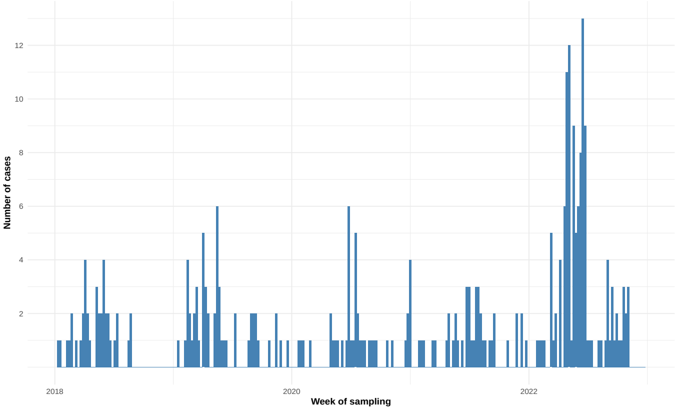
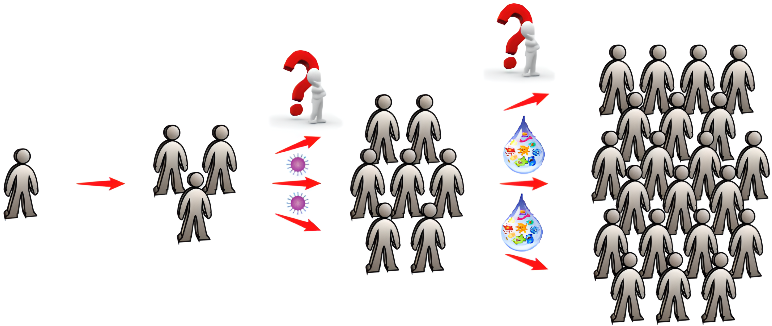


Figure: *Weekly cumulative number of new reported cases of typhoid fever between 2018 and 2022 in Mayotte. (source: ARS Mayotte)*

Objective

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

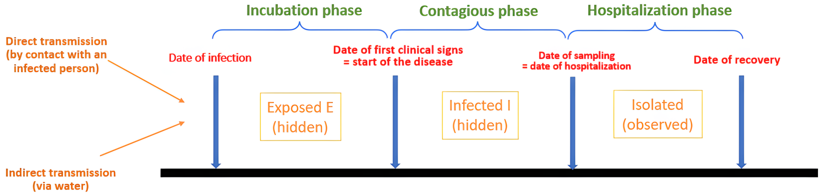


Model

- **Population characteristics :**
  - Large population of Susceptibles: $\sim$ 300 000 inhabitants
  - Small population of Infected: $\sim$ 40 cases per year on average

$\Rightarrow$ Stochastic model of the counting process type

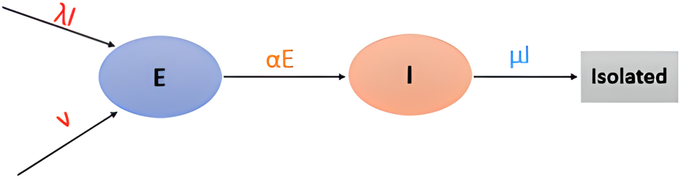
- **Disease dynamics:**



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

Model

- **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

$\Rightarrow$ Estimate the rates describing the dynamics of typhoid fever transmission (λ, ν, α and μ)

Observations

- The jump times: not observed
- The post jump locations: not 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

⇒ Difficulty: Classical estimation methods

- Continuous-time model with discrete observations
- No analytical / numerical expression for likelihood

Estimation Strategy

- Derive estimators of λ, ν, α and μ (moments estimators) for the complete observations framework (cumulative + number of exposed and infected)

$$(\lambda, \nu, \alpha, \mu) = f(E^*, I^*, R^*, N^\star),$$

where

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

and

$$N^\star = \lim_{t \to +\infty} \frac{\mathbb{E}_{(e_0, i_0, n_0)}[N_t]}{t}$$

$N_t := \sum_{k=0}^{[t]} O_k$ describing the cumulative number of new isolated cases from 0 to time t

$\Rightarrow$ Estimate E^*, I^*, R^* for our case: discrete time hidden observations framework

- Hidden multi-chain Markov model (HMCMM)

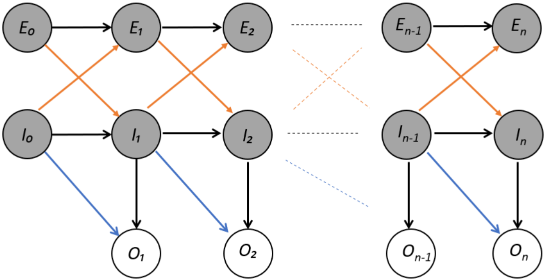


Figure: **HMCMM emission scheme.**

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

Estimation Steps

▶ Aim: Estimate the parameters of the HMM $M = (Q, \psi, \rho)$ by adapting the Baum-welch (EM) algorithm & derive an estimate of the transition matrix p of the chain (E_n, I_n)

• Rewrite the principal functions of the EM algorithm: Forward and Backward probabilities

• Adapted Baum-welch algorithm : Starting from an initial model $M^{(0)}$, the re-estimations are given by

$$Q^{n+1}_{(e,i,j),(e',i',j')} = \frac{\sum_{t=1}^{T-1} \xi^n_{(e,i,j),(e',i',j')}(t)}{\sum_{t=1}^{T} \gamma^n_{(e,i,j)}(t)} \delta_{i'=j}$$

$$\psi^{n+1}_{(e,i,j)}(o) = \frac{\sum_{t=1}^{T} \mathbb{1}_{o_t=o} \gamma^n_{(e,i,j)}(t)}{\sum_{t=1}^{T} \gamma^n_{(e,i,j)}(t)}, \quad \rho^{n+1}_{e,i,j} = \gamma^n_{(e,i,j)}(1)$$

$$p^{n+1}_{(e,i),(e',i')} = \frac{\sum_{j' \in \mathbb{N}} Q^{n+1}_{(e,i,j),(e',j,j')} \rho^{n+1}_{e,i,j}}{\sum_{j \in \mathbb{N}} \rho^{n+1}_{e,i,j}},$$

Estimation Steps

where

$$\xi^n_{(e,i,j),(e',i',j')}(t) = \frac{\alpha^n_{(e,i,j)}(t) p^n_{(e',i'),(.,j')} p^n_{(e,i),(e',i')} \psi^n_{(e',i',j')}(o_{t+1}) \beta^n_{(e',i',j')}(t+1)}{p^n_{(e,i),(.,j)} \sum_{(e,i,j)\in\mathbb{N}^3} \alpha^n_{(e,i,j)}(t)\beta^n_{(e,i,j)}(t)}$$

$$\gamma^n_{(e,i,j)}(t) = \frac{\alpha^n_{(e,i,j)}(t)\beta^n_{(e,i,j)}(t)}{\sum_{(e,i,j)\in\mathbb{N}^3} \alpha^n_{(e,i,j)}(t)\beta^n_{(e,i,j)}(t)}.$$

- Estimate the moments E^*, I^* and R^* of the chain (E_n, I_n) from the optimal transition matrix $\hat{p}$ obtained using Monte Carlo simulations

Numerical simulation

▶ Complete observations: We run a 10000 parallelized simulation of the process (E_t, I_t) starting from $(E_0, i_0) = (0, 0)$ with $\boldsymbol{\lambda = 0.05, \mu = 0.2, \alpha = 0.1}$ and $\boldsymbol{\nu = 0.015}$

- Fixed time step $\Delta t = 1 day \Rightarrow (E_n, I_n)$ and the observations O_n

▶ Number of observations $n = 10000$

Parameters	Estimates	Confidence intervals (95%)
$\lambda = 0.05$	0.050	$[0.049, 0.0506]$
$\mu = 0.2$	0.20	$[0.1998, 0.2001]$
$\alpha = 0.1$	0.100	$[0.0998, 0.1002]$
$\nu = 0.015$	0.015	$[0.0149, 0.01506]$

$\Rightarrow$ The accuracy of the estimates depends on the number of simulated trajectories and the number of observations

Numerical simulation

▶ Hidden framework: Parameters initialization: $\lambda^{(0)} \in [0.04, 0.07]$, $\mu^{(0)} \in [0.185, 0.25], \alpha^{(0)} \in [0.09, 0.130]$ et $\nu^{(0)} \in [0.013, 0.02]$

- For each quadruplet value $(\lambda_i^{(0)}, \mu_i^{(0)}, \alpha_i^{(0)}, \nu_i^{(0)})$ for $i = 1, \ldots, 15$:
  - Estimate $P^{(0)}$ and $\psi^{(0)}$ by Monte Carlo simulations
  - Executing the Baum-Welch algorithm adapted and choosing

$$\hat{p}^m = max_{\hat{p}^n} \mathbb{P}(O|M^n)$$

Number of observations	$n = 10000$		
Truncation of hidden state space	$N = 3$	$N = 4$	$N = 5$
$\hat{\lambda}^n$	0.050	0.049	0.05
$\hat{\mu}^n$	0.189	0.196	0.209
$\hat{\alpha}^n$	0.098	0.101	0.102
$\hat{\nu}^n$	0.0146	0.015	0.015

Numerical simulation

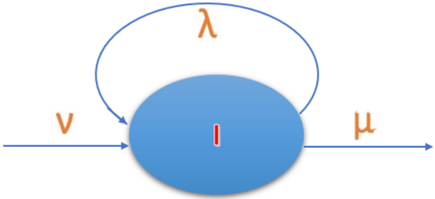
▶ Hidden framework:

- Truncation of the state space does not have a significant impact on the estimate
- The quality of parameter estimation depends essentially on the quantity of data available:
  - Improves the local convergence of the EM
  - Improves the estimate of $N^\star$ which depends essentially on the number of observations
  - It also leads to a significant increase in execution time.

$\Rightarrow$ The dataset provided by the Mayotte ARS (1580 observations) proved insufficient to obtain reliable and accurate estimates of the model parameters

Simplified model

- Linear birth-death-immigration (LBDI) model



$\Rightarrow$ 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

• Results based on hospitalisation data for Mayotte

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

$\Rightarrow$ The estimated parameters are

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

▶ $\hat{\lambda} < \hat{\mu}$: coherent with observations

• Person-to-person transmission plays a significant role in the spread of the disease in Mayotte

• The average duration of isolation is 7.5 days

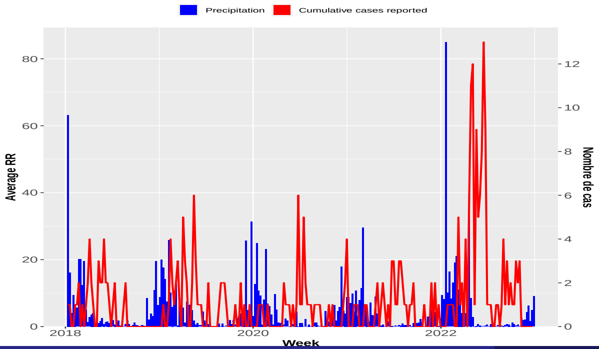
• Water quality plays an important role in the spread of typhoid fever in Mayotte

Conclusion & Perspective

⇒ Conclusion :

- Continuous-time model with discrete observations:
  - Derive estimators of parameters for the complete observations framework (in terms of transition probabilities or moments)
- No analytical / numerical expression for likelihood
  - The hidden framework: Hidden Multi-Chain Markov Model (HMCMM) or Hidden Markov Model (HMM)

⇒ Perspective:




thank you