

Aspects stochastiques

CIMAD N'Djaména 2025 – Cours 03

Julien Arino

Department of Mathematics @ University of Manitoba
Maud Menten Institute @ PIMS

julien.arino@umanitoba.ca

24 novembre 2025

The University of Manitoba campuses are located on original lands of Anishinaabeg, Ininew, Anisininew, Dakota and Dene peoples, and on the National Homeland of the Red River Métis. We respect the Treaties that were made on these territories, we acknowledge the harms and mistakes of the past, and we dedicate ourselves to move forward in partnership with Indigenous communities in a spirit of Reconciliation and collaboration.

Outline

Pourquoi incorporer la stochasticité ?

Stochasticité dans les modèles déterministes

Chaînes de Markov en temps continu

**Approximations des CMTC par processus de
branchement**

Pourquoi incorporer la stochasticité ?

Stochasticité dans les modèles déterministes

Chaînes de Markov en temps continu

Approximations des CMTC par processus de
branchement

Au début de la crise COVID-19

- ▶ Je travaillais sous contrat avec l'Agence de la santé publique du Canada sur l'évaluation du risque d'importation de COVID-19
- ▶ Je produisais un rapport quotidien avec la liste des pays les plus susceptibles de signaler des cas de COVID-19 dans les jours à venir
- ▶ J'utilisais des simulations ensemblistes d'un modèle de métapopulation déterministe global ajusté



- ▶ Jours très très longs (18-20 heures, 7 jours par semaine)
- ▶ Ceci incluait un bon moment à attendre que le “cluster” finisse de tourner
 - ⇒ PHAC/ACSP m'a donné des sous pour acheter un vrai cluster (Threadrippers !!!)
 - ⇒ J'ai aussi commencé à réfléchir à la validité de mon modèle alors que mon focus passait du mouvement au niveau planétaire à un mouvement plus local entre provinces du Canada

Qu'est-ce qui ne va pas avec les modèles déterministes ?

- ▶ J'ai souligné hier que le SARS-CoV-2 est une *unique* réalisation d'un processus stochastique
- ▶ Les modèles déterministes « opèrent sur des moyennes » sur un grand ($\rightarrow +\infty$) nombre de réalisations
- ▶ Si nous voulons avoir une meilleure idée de ce qui pourrait se passer, pas seulement en moyenne, alors nous devons voir ce qui peut effectivement se passer

Mon nouveau focus – Les introductions

- ▶ J'ai commencé à penser en particulier aux **introductions** (ou importations) de pathogènes dans de nouvelles populations
- ▶ En effet, les introductions sont une étape obligatoire dans la propagation spatiale

Premier élément de preuve

Dans la vraie vie, les introductions de pathogènes ne suivent pas toujours le schéma

$$\{\mathcal{R}_0 < 1 \implies \text{DFE} \mid \mathcal{R}_0 > 1 \implies \text{épidémie ou} \rightarrow \text{EEP}\}$$

Short Communication

SARS-CoV-2 in Nursing Homes: Analysis of Routine Surveillance Data in Four European Countries

Tristan Delory^{1,2*}, Julien Arino³, Paul-Emile Hay⁴, Vincent Klotz⁴, Pierre-Yves Boëlle¹

¹ Sorbonne Université, INSERM, Institut Pierre Louis d’Épidémiologie et de Santé Publique, IPLESP, F-75012, Paris, France. ²Centre Hospitalier Annecy Genevois, France. ³Department of Mathematics, University of Manitoba, Winnipeg, Manitoba, Canada. ⁴Groupe Colisee.

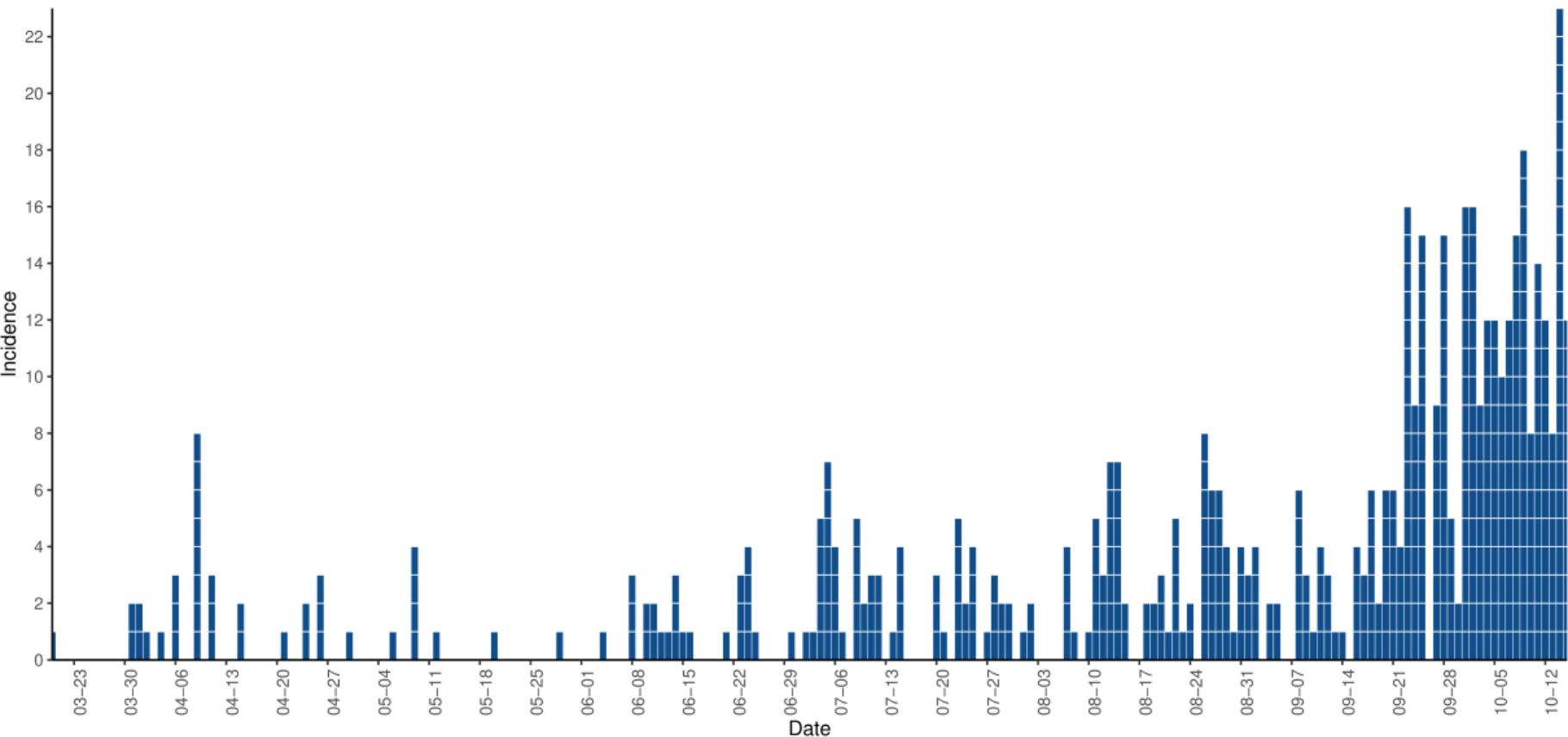
Table 1. Effect of vaccination scaling-up on the probability of successful viral introduction.

Period	Failed N = 136	Successful N = 366	aOR*	95%CI	P-value
Before vaccination	94 (69.1%)	311 (85.0%)	Ref		
January 15 to January 31	12 (8.8%)	37 (10.1%)	0.89	0.42 – 1.92	0.770
February 01 to February 15	17 (12.5%)	14 (3.8%)	0.23	0.10 – 0.52	<0.001
February 16 to February 28	13 (9.6%)	4 (1.1%)	0.08	0.02 - 0.29	<0.001

* Adjusted on study period, country, staffing ratio, cumulative attack rate at onset of introduction, and number of PCR per 1000-residents or 1000-staff members, at onset of introduction, and nursing home maximal capacity.

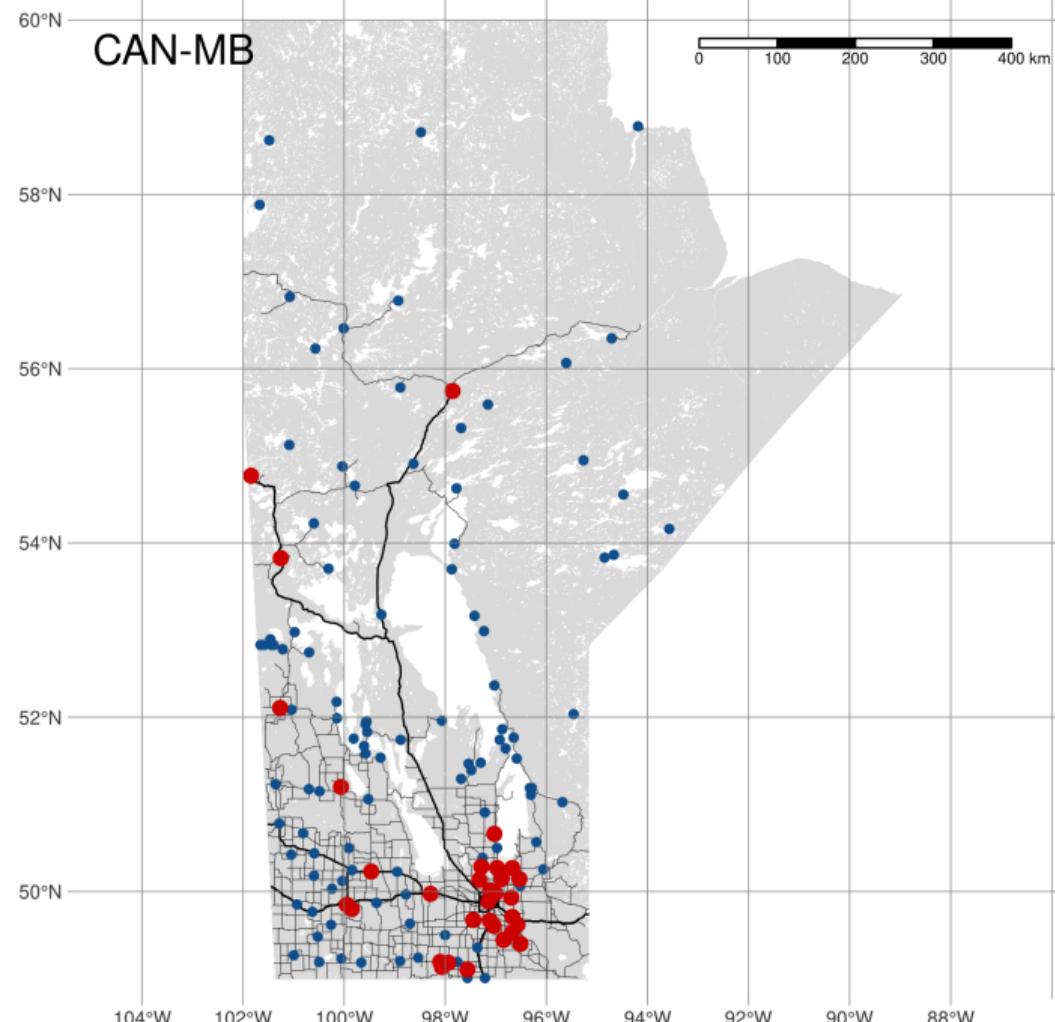
Deuxième élément de preuve

Le début d'une épidémie peut être extrêmement lent, avec très peu de cas pendant un certain temps



Pourquoi ceci est pertinent

Loin d'être la seule raison, mais à titre d'exemple : le Canada a des communautés éloignées/isolées qui sont vulnérables aux introductions de pathogènes



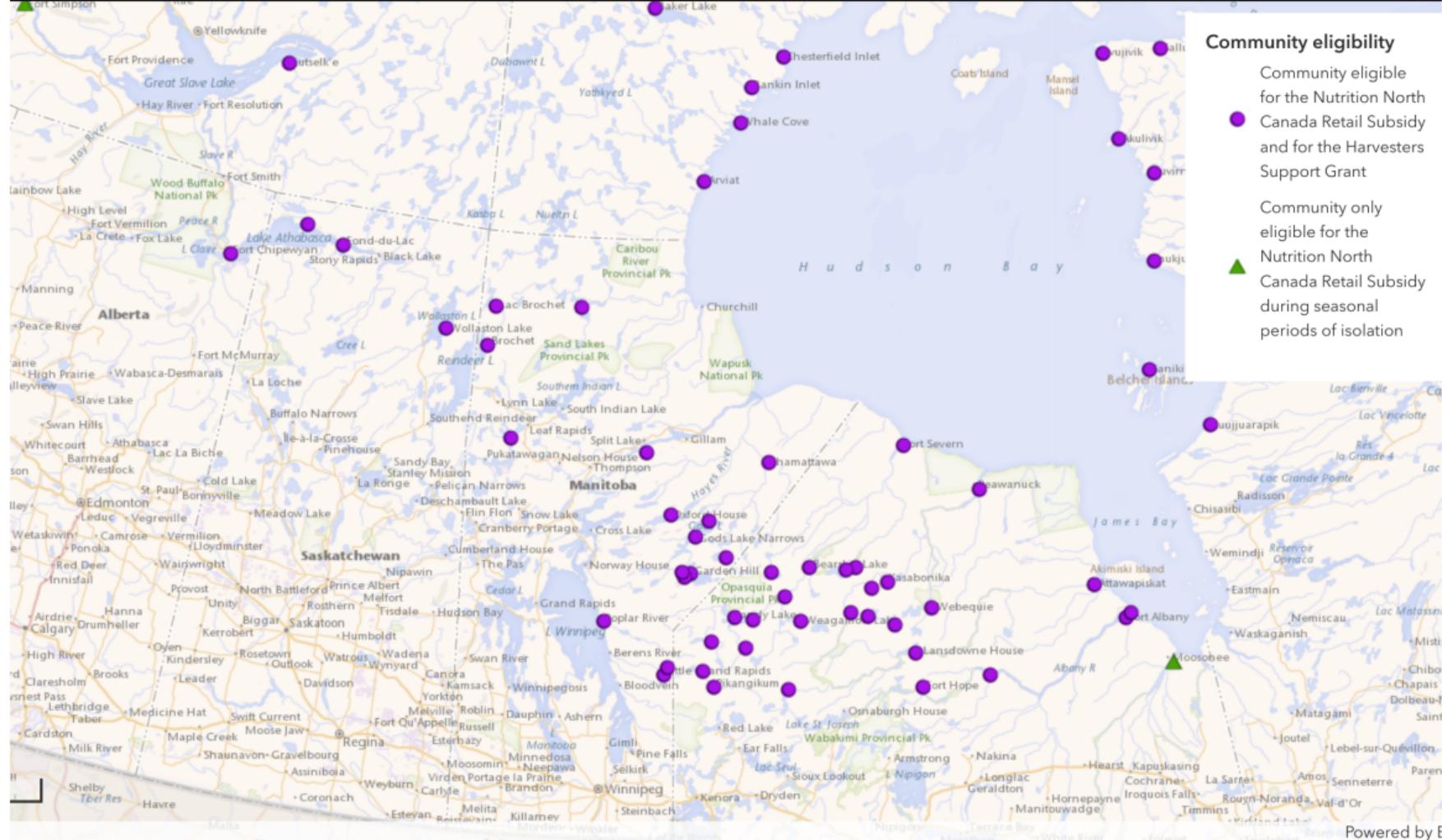
Community eligibility

Community eligible
for the Nutrition North

● Canada Retail Subsidy
and for the Harvesters
Support Grant

Community only
eligible for the
Nutrition North

▲ Canada Retail Subsidy
during seasonal
periods of isolation



Northern Manitoba chiefs call for immediate federal action on health-care crisis

Recent deaths linked to inadequate medical care include mother of 5 from Manto Sipi Cree Nation, chief says

CBC News · Posted: Apr 03, 2023 3:20 PM CDT | Last Updated: April 3, 2023



'A lengthy process to get help here'

Wasagamack is one of four First Nations communities that make up Island Lake, an area in northeastern Manitoba dotted with hundreds of small islands.

Island Lake has a population of at least 15,000, according to Scott Harper, the grand chief of Anisininew Okimawin, which represents the four communities.

Despite having a population roughly the size of Thompson, and having diabetes and hospitalization rates well above provincial averages, Island Lake has no hospital of its own. The region is accessible only by air, boat and an unreliable winter road.

The nursing station in Wasagamack First Nation, which has about 2,300 people, according to federal government data, typically operates short-staffed, with only two or three of five registered nurses working on any given rotation and a fly-in doctor who comes weekly.

Pour les communautés des Premières Nations et Métis

Éloignée décrit une **zone géographique** où une communauté est **située à plus de 350 km du centre de services le plus proche ayant un accès toute l'année** par voies terrestres et/ou maritimes normalement utilisées par tous les temps

Isolée signifie une **zone géographique** qui a des **vols réguliers** et un bon service téléphonique, mais est **sans accès toute l'année** par voies terrestres et/ou maritimes normalement utilisées par tous les temps

Éloignée-Isolée signifie une **zone géographique** qui n'a **ni vols réguliers ni accès toute l'année** par voies terrestres et/ou maritimes qui peuvent être utilisées par tous les temps, indépendamment du niveau de service téléphonique et radio disponible

Pour les communautés inuites

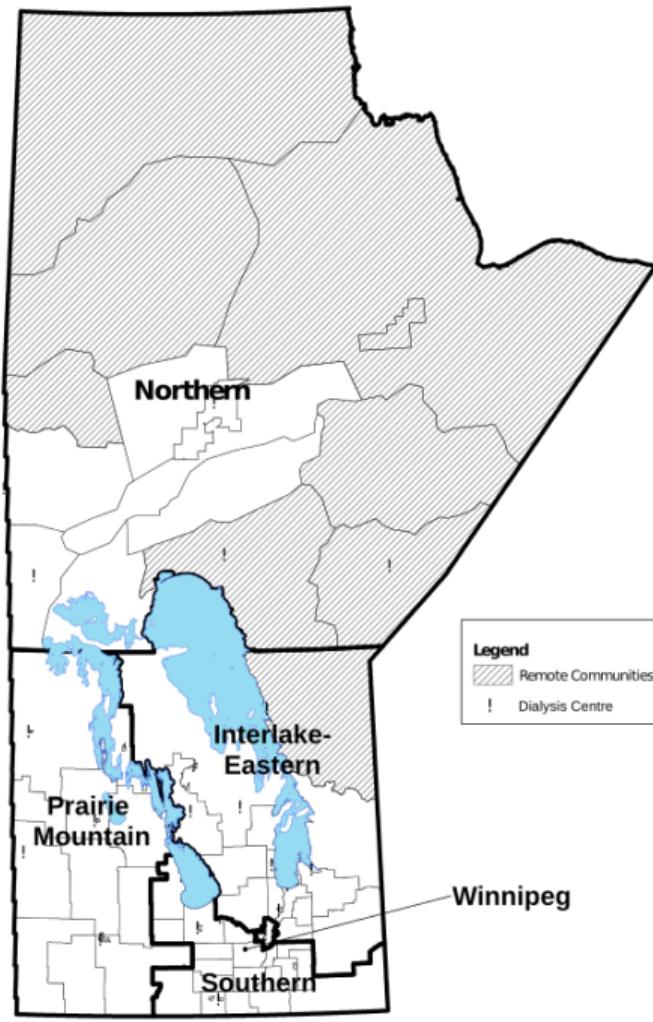
Les communautés inuites doivent être appelées **Inuit Nunangat**, et non communautés éloignées et isolées, afin de respecter la langue et la culture uniques des régions inuites, ainsi que les défis communs en matière de déterminants sociaux de la santé, d'accès aux soins et d'infrastructure qui se retrouvent dans toutes les communautés inuites

Communautés éloignées du MB

Les **communautés éloignées** sont des communautés du Manitoba qui **n'ont pas d'accès routier permanent** (c'est-à-dire, pas de route praticable en toutes saisons), sont **à plus de quatre heures de route** d'un hôpital rural important (et d'une unité de dialyse), **ou ont un accès uniquement par rail ou avion**. Cela inclut Norway House, Lynn Lake, Leaf Rapids, Gillam et Cross Lake. Si la plupart des communautés d'un district sanitaire sont désignées comme « éloignées », l'ensemble du district est désigné comme « éloigné ». Au Manitoba, les districts éloignés comprennent :

- ▶ Région sanitaire du Nord : NO23, NO13, NO25, NO16, NO22, NO26, NO28, NO31, et
- ▶ Région sanitaire Interlake-Est : IE61.

Chartier M, Dart A, Tangri N, Komenda P, Walld R, Bogdanovic B, Burchill C, Koseva I, McGowan K, Rajotte L. Care of Manitobans Living with Chronic Kidney Disease. Winnipeg, MB. Manitoba Centre for Health Policy, December 2015



Voyages vers/depuis des communautés éloignées ou isolées

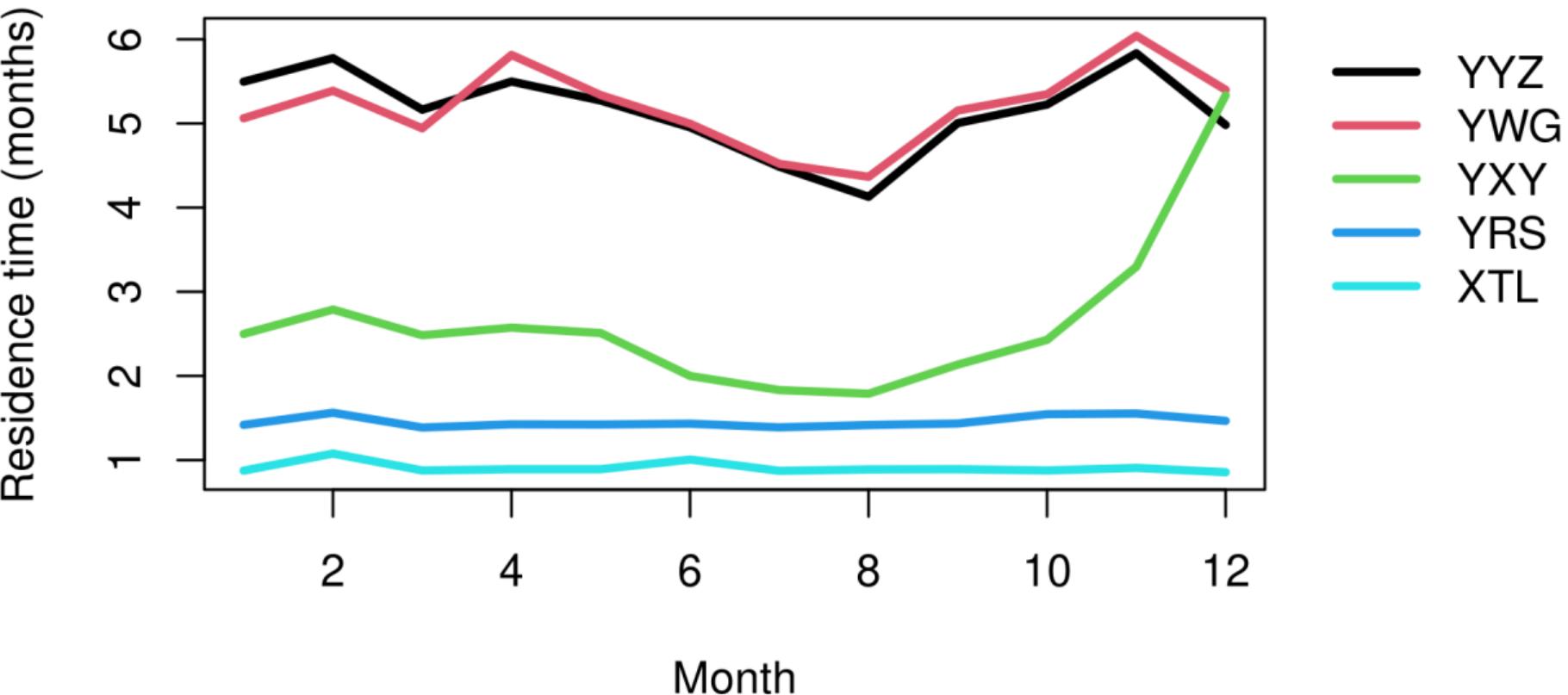
Comment pensez-vous que cela se compare aux voyages dans les communautés non éloignées/isolées ?

Temps de résidence (la version écologie lacustre) : temps théorique qu'une molécule d'eau moyenne ou comparable passe dans un lac, considérant l'afflux dans et l'écoulement hors du lac

Pensez aux temps de résidence dans ces communautés : quel est le temps moyen qu'une personne passe dans une communauté éloignée ou isolée avant de la quitter ?

Le **temps de résidence dans un lieu** est le nombre total de trajets entrants et sortants d'une localisation sur une durée (1 mois ici) divisé par la population normale dans la localisation

Residence times in months



Le paradoxe des voyages vers/depuis les communautés éloignées/isolées

Volumes de voyage faibles mais taux de mouvement élevés

Les CI sont fortement connectées au(x) centre(s) urbain(s) auxquels elles sont subordonnées

Encore renforcé à Winnipeg par la population autochtone urbaine (102 075 ou 12,45% de la population métropolitaine), ce qui signifie que de nombreux liens familiaux existent



Pourquoi incorporer la stochasticité ?

Stochasticité dans les modèles déterministes

Chaînes de Markov en temps continu

Approximations des CMTC par processus de
branchement

Stochasticité dans les modèles déterministes

Distributions des temps jusqu'aux événements

Deux distributions « extrêmes » et une plus agréable

Un modèle de cohorte simple avec décès

Une correction possible au problème de la distribution exponentielle

Temps de séjour dans un modèle de transmission de maladie SIS

Un modèle avec vaccination

Voir en particulier les travaux de Horst Thieme

Si l'on considère le temps de séjour dans les compartiments d'une perspective plus détaillée, on obtient des modèles intégro-différentiels

Nous utilisons ici des variables aléatoires continues. Voir les chapitres 12 et 13 du livre de Thieme pour des distributions arbitraires

Temps jusqu'aux événements

Nous supposons qu'un système peut être dans deux états, A et B

- ▶ Au temps $t = 0$, le système est dans l'état A
- ▶ Un événement se produit à un certain temps $t = \tau$, qui déclenche le passage de l'état A à l'état B

Appelons T la variable aléatoire

“temps passé dans l'état A avant de passer à l'état B”

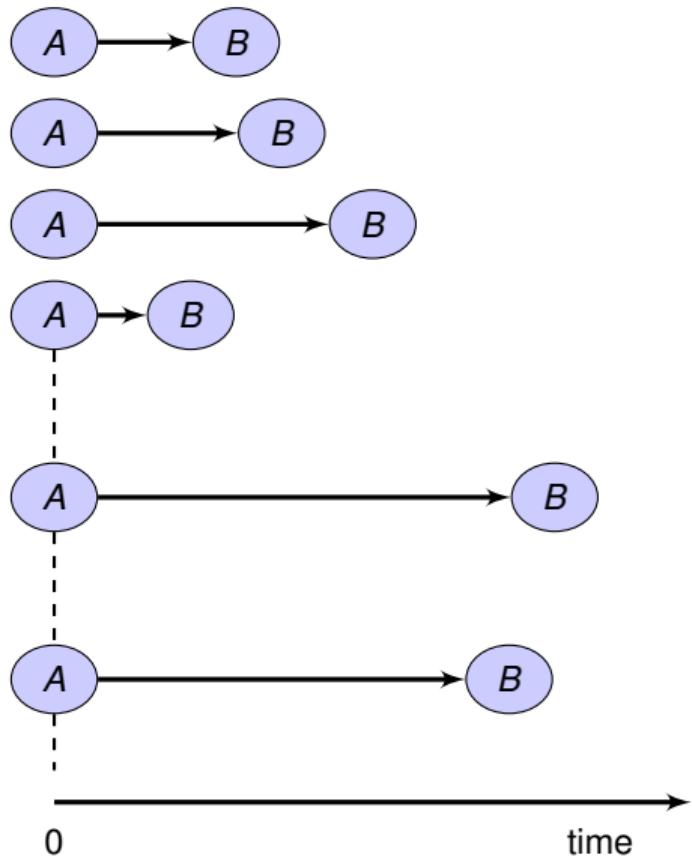
Les états peuvent être n'importe quoi :

- ▶ A : fonctionnel, B : en panne
- ▶ A : infecté, B : rétabli
- ▶ A : vivant, B : mort
- ▶ ...

Nous prenons une collection d'objets ou d'individus qui sont dans l'état A et voulons une loi pour la **distribution** des temps passés dans A , c'est-à-dire une loi pour T

Par exemple, nous fabriquons des ampoules et aimerais dire à nos clients qu'en moyenne, nos ampoules durent 200 ans...

Nous effectuons un nombre **infini** d'expériences, et observons le temps qu'il faut, dans chaque expérience, pour passer de A à B



Une distribution de probabilité est un modèle

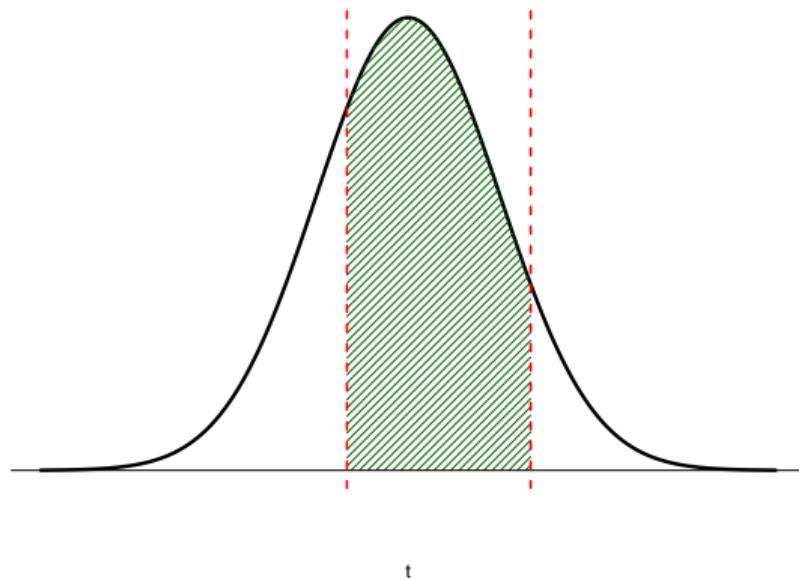
À partir de la séquence d'expériences, nous déduisons un modèle, qui dans ce contexte est appelé une **distribution de probabilité**

Nous supposons que T est une variable aléatoire **continue**

Fonction de densité de probabilité

Puisque T est continue, elle a une **fonction de densité de probabilité** continue f

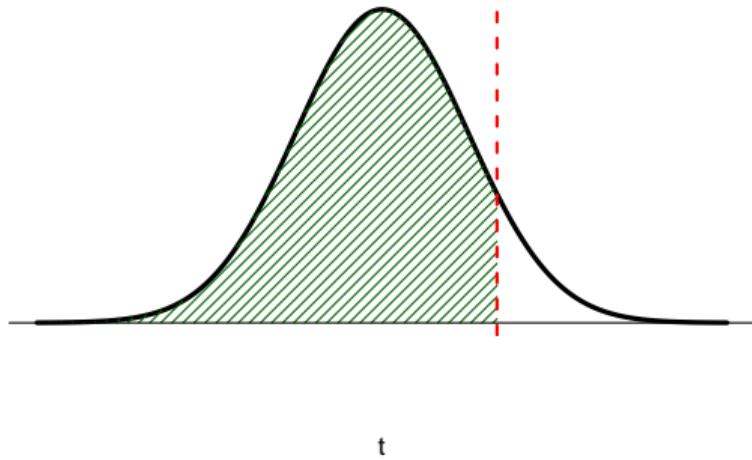
- ▶ $f \geq 0$
- ▶ $\int_{-\infty}^{+\infty} f(s)ds = 1$
- ▶ $\mathbb{P}(a \leq T \leq b) = \int_a^b f(t)dt$



Fonction de répartition (f.r.)

La **fonction de répartition** est une fonction $F(t)$ qui caractérise la distribution de T , et définie par \Leftrightarrow

$$F(s) = \mathbb{P}(T \leq s) = \int_{-\infty}^s f(x)dx$$



Fonction de survie

Une autre caractérisation de la distribution de la variable aléatoire T se fait par la fonction de **survie** (ou de **séjour**)

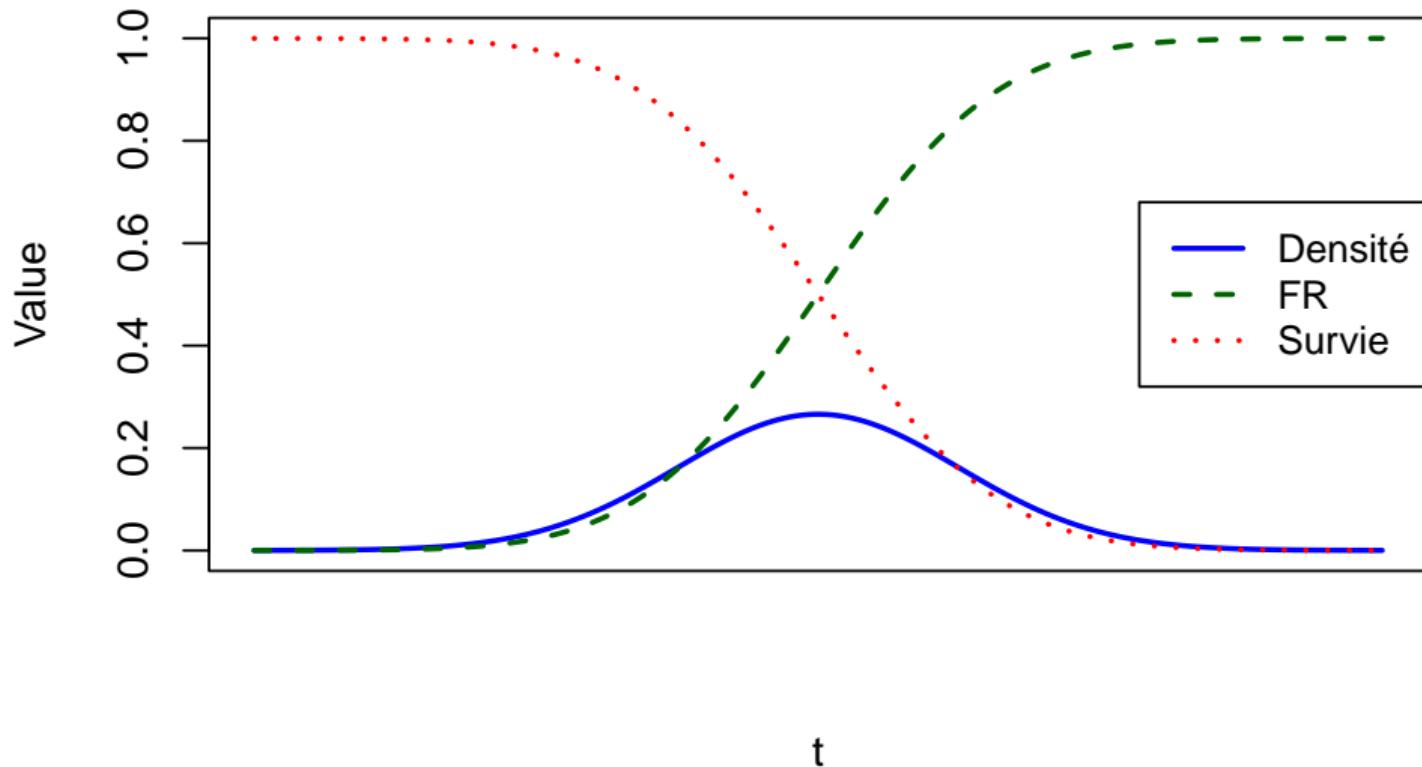
La fonction de survie de l'état A est donnée par

$$S(t) = 1 - F(t) = \mathbb{P}(T > t) \quad (1)$$

Ceci donne une description du **temps de séjour** d'un système dans un état particulier (le temps passé dans l'état)

S est une fonction non croissante (puisque $S = 1 - F$ avec F une f.r.), et $S(0) = 1$ (puisque T est une variable aléatoire non négative)

Distribution, fonctions de répartition et de survie



Le **temps de séjour moyen** τ dans l'état A est donné par

$$\tau = E(T) = \int_0^\infty t f(t) dt$$

Puisque $\lim_{t \rightarrow \infty} t S(t) = 0$, il s'ensuit que

$$\tau = \int_0^\infty S(t) dt$$

Durée de vie future espérée :

$$\frac{1}{S(t_0)} \int_0^\infty t f(t + t_0) dt$$

$$\begin{aligned} S(t) - S(a) &= \mathbb{P}\{\text{survivre pendant } (a, t) \text{ ayant survécu jusqu'à } a\} \\ &= \exp\left(-\int_a^t h(u) du\right) \end{aligned}$$

Taux de risque

Le **taux de risque** (ou **taux de défaillance**) est

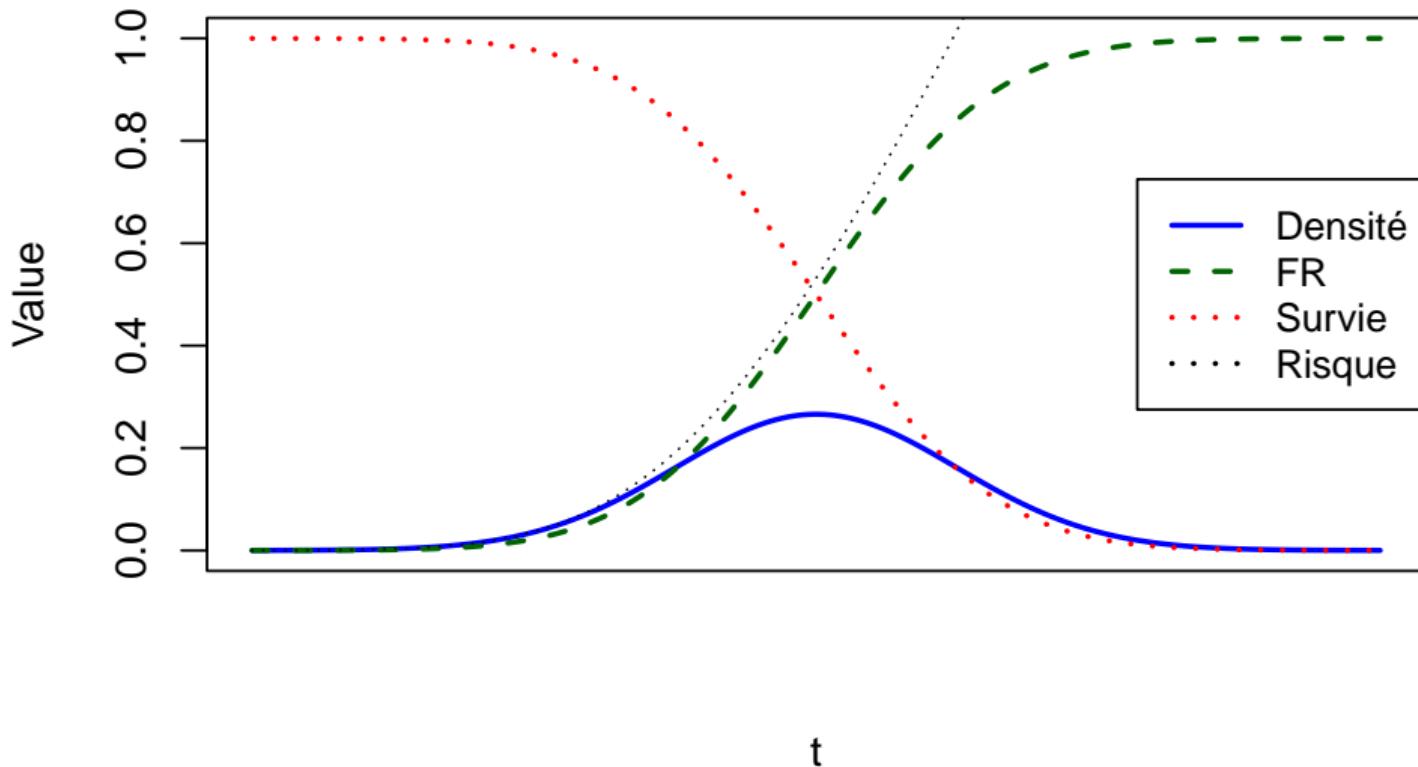
$$\begin{aligned} h(t) &= \lim_{\Delta t \rightarrow 0} \frac{\mathcal{S}(t) - \mathcal{S}(t + \Delta t)}{\Delta t} \\ &= \lim_{\Delta t \rightarrow 0} \frac{\mathbb{P}(T < t + \Delta t | T \geq t)}{\Delta t} \\ &= \frac{f(t)}{\mathcal{S}(t)} \end{aligned}$$

Il donne la probabilité de défaillance entre t et Δt , étant donné la survie jusqu'à t .

Nous avons

$$h(t) = -\frac{d}{dt} \ln \mathcal{S}(t)$$

Densité, f.r., survie & taux de risque



Risques concurrents

Supposons maintenant que le système commence dans l'état A au temps $t = 0$ et que selon lequel des deux événements \mathcal{E}_1 ou \mathcal{E}_2 a lieu en premier, il passe à l'état B_1 ou B_2 , respectivement

Considérons les variables aléatoires T_A , *temps passé dans l'état A (ou temps de séjour dans A)*, T_{AB_1} , *temps avant le passage à B_1* et T_{AB_2} , *temps avant le passage à B_2*

Si nous considérons l'état A , nous ne pouvons pas observer les variables T_{AB_1} ou T_{AB_2} . Ce qui est observable est le temps de séjour dans A

$$T_A^* = \min(T_{AB_1}, T_{AB_2})$$

(où * indique qu'une quantité est observable)

Taux de défaillance par type d'événement

Nous avons deux (ou plusieurs) types d'événements dont les taux de défaillance individuels doivent être pris en compte

$$h_j(t) = \lim_{\Delta t \rightarrow 0} \frac{\mathbb{P}(T < t + \Delta t, S = S_j | T \geq t)}{\Delta t}$$

où $\mathbb{P}(T < t + \Delta t, S = S_j | T \geq t)$ est la probabilité de défaillance due à la cause S_j ($j = 1, 2$ ici), c'est-à-dire S est une v.a. discrète représentant l'événement qui se produit

Par la loi de probabilité totale, puisqu'un seul des événements peut avoir lieu, s'il y a n risques, alors

$$h(t) = \sum_{i=1}^n h_i(t)$$

ou, de manière identique,

$$\mathcal{S}(t) = \exp \left(- \int_0^t \sum_{j=1}^n h_j(s) \, ds \right)$$

En conséquence, supposons qu'un processus est soumis à deux risques exponentiels concurrents avec des distributions respectives de paramètres θ_1 et θ_2

Alors le temps de séjour moyen dans l'état initial avant d'être affecté par l'un des deux risques est

$$\frac{1}{\theta_1 + \theta_2}$$

Stochasticité dans les modèles déterministes

Distributions des temps jusqu'aux événements

Deux distributions « extrêmes » et une plus agréable

Un modèle de cohorte simple avec décès

Une correction possible au problème de la distribution exponentielle

Temps de séjour dans un modèle de transmission de maladie SIS

Un modèle avec vaccination

La distribution exponentielle

La variable aléatoire T a une distribution **exponentielle** si sa fonction de densité de probabilité prend la forme

$$f(t) = \begin{cases} 0 & \text{si } t < 0, \\ \theta e^{-\theta t} & \text{si } t \geq 0, \end{cases} \quad (2)$$

avec $\theta > 0$. Alors la fonction de survie pour l'état A est de la forme $S(t) = e^{-\theta t}$, pour $t \geq 0$, et le temps de séjour moyen dans l'état A est

$$\tau = \int_0^\infty e^{-\theta t} dt = \frac{1}{\theta}$$

Particularités de la distribution exponentielle

L'écart-type d'une distribution exponentielle est également $1/\theta$. Lors de l'estimation de θ , il est impossible de distinguer la moyenne et l'écart-type

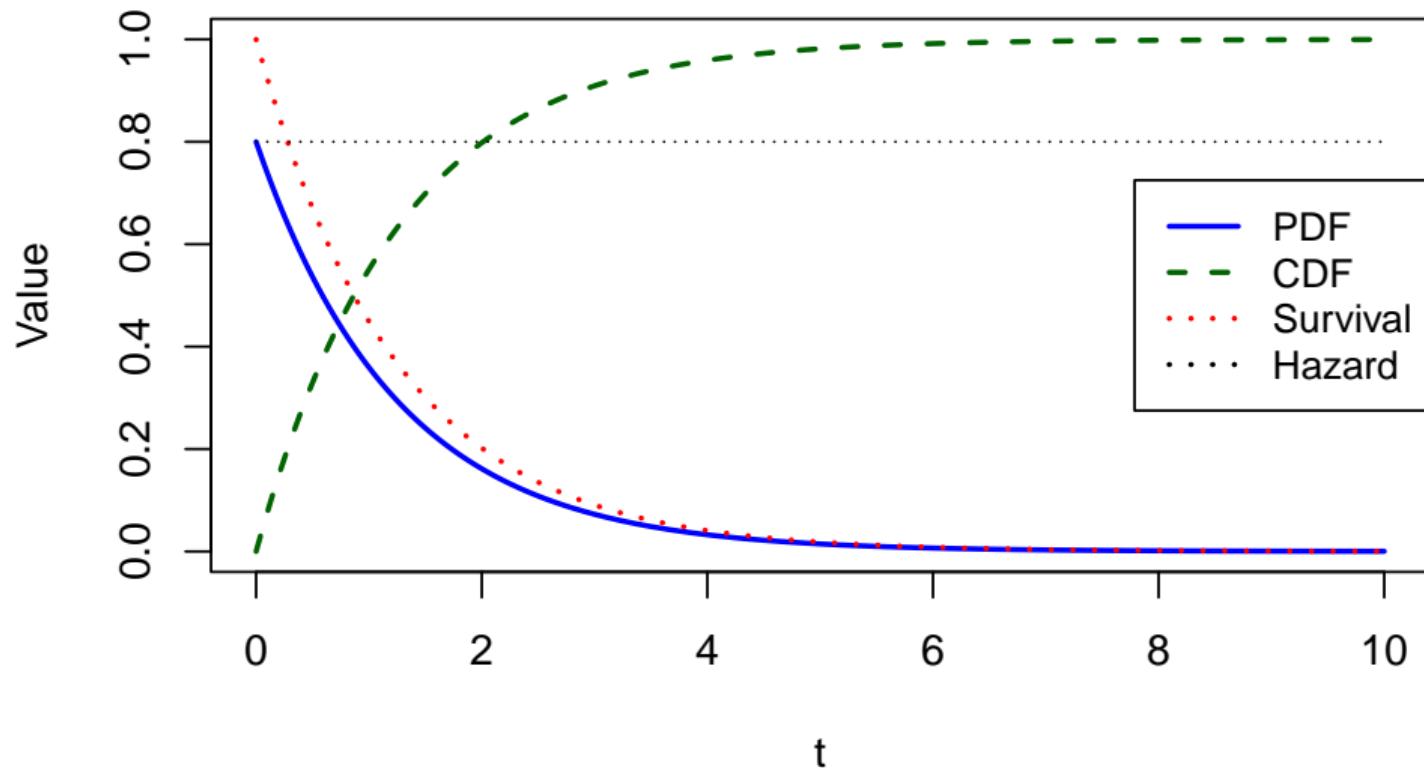
La distribution exponentielle est **sans mémoire** : sa probabilité conditionnelle obéit à

$$P(T > s + t \mid T > s) = P(T > t), \quad \forall s, t \geq 0$$

Les distributions exponentielle et géométrique sont les seules distributions de probabilité sans mémoire

La distribution exponentielle a une fonction de risque constante $h(t) \equiv \theta$

PD, CD and Surv. functions & Hazard rate of exponential



La distribution delta de Dirac

Si pour une constante $\omega > 0$,

$$S(t) = \begin{cases} 1, & 0 \leq t \leq \omega \\ 0, & \omega < t \end{cases}$$

ce qui signifie que T a une distribution delta de Dirac $\delta_\omega(t)$, alors le temps de séjour moyen est

$$\tau = \int_0^\omega dt = \omega$$

avec écart-type $\sigma = 0$

La distribution Gamma

La v.a. X est distribuée selon une loi **Gamma** ($X \sim \Gamma(k, \theta)$) avec **paramètre de forme** k et **paramètre d'échelle** θ (ou **taux** $\beta = 1/\theta$) (tous positifs) si sa fonction de densité de probabilité prend la forme

$$f(x; k, \theta) = \frac{x^{k-1} e^{-\frac{x}{\theta}}}{\Gamma(k)\theta^k} \quad (3)$$

où $x > 0$ et Γ est la fonction Gamma d'Euler, définie pour tout $z \in \mathbb{C}$ tel que $\text{Re}(z) > 0$ par

$$\Gamma : z \mapsto \int_0^{+\infty} t^{z-1} e^{-t} dt$$

Propriétés de la distribution Gamma

Moyenne $k\theta$, variance $k\theta^2$

Fonction de survie

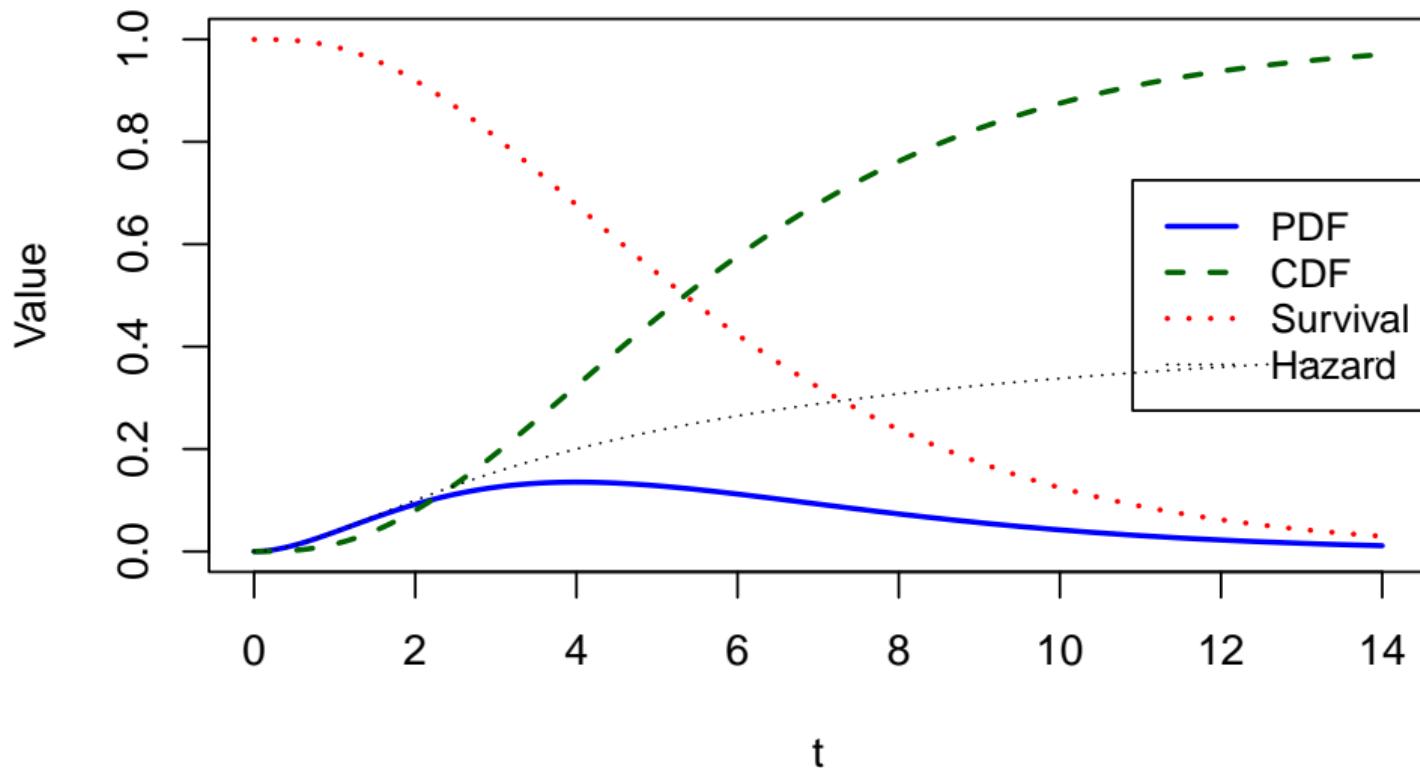
$$S(t) = 1 - \frac{1}{\Gamma(k)} \gamma \left(k, \frac{t}{\theta} \right) = 1 - \frac{1}{\Gamma(k)} \gamma(k, \beta t)$$

où

$$\gamma(a, x) = \int_0^x t^{a-1} e^{-t} dt$$

est une fonction Gamma incomplète

PDF, CDF, Survival & Hazard of Gamma Distribution



Stochasticité dans les modèles déterministes

Distributions des temps jusqu'aux événements

Deux distributions « extrêmes » et une plus agréable

Un modèle de cohorte simple avec décès

Une correction possible au problème de la distribution exponentielle

Temps de séjour dans un modèle de transmission de maladie SIS

Un modèle avec vaccination

Un modèle pour une cohorte avec une cause de décès

Considérons une **cohorte** d'individus nés au même moment, par exemple, la même année

- ▶ Au temps $t = 0$, il y a initialement $N_0 > 0$ individus
- ▶ Toutes les causes de décès sont combinées ensemble
- ▶ Le temps jusqu'au décès, pour un individu donné, est une variable aléatoire T , avec une distribution de densité de probabilité continue $f(t)$ et une fonction de survie $S(t)$

$N(t)$ la population de la cohorte au temps $t \geq 0$

$$N(t) = N_0 S(t) \tag{4}$$

$S(t)$ proportion de la population initiale encore en vie au temps t , donc $N_0 S(t)$ nombre dans la cohorte encore en vie au temps t

Cas où T est distribuée exponentiellement

Supposons que T ait une distribution exponentielle avec moyenne $1/d$ (ou paramètre d), $f(t) = de^{-dt}$. Alors la fonction de survie est $S(t) = e^{-dt}$, et (4) prend la forme

$$N(t) = N_0 e^{-dt} \tag{5}$$

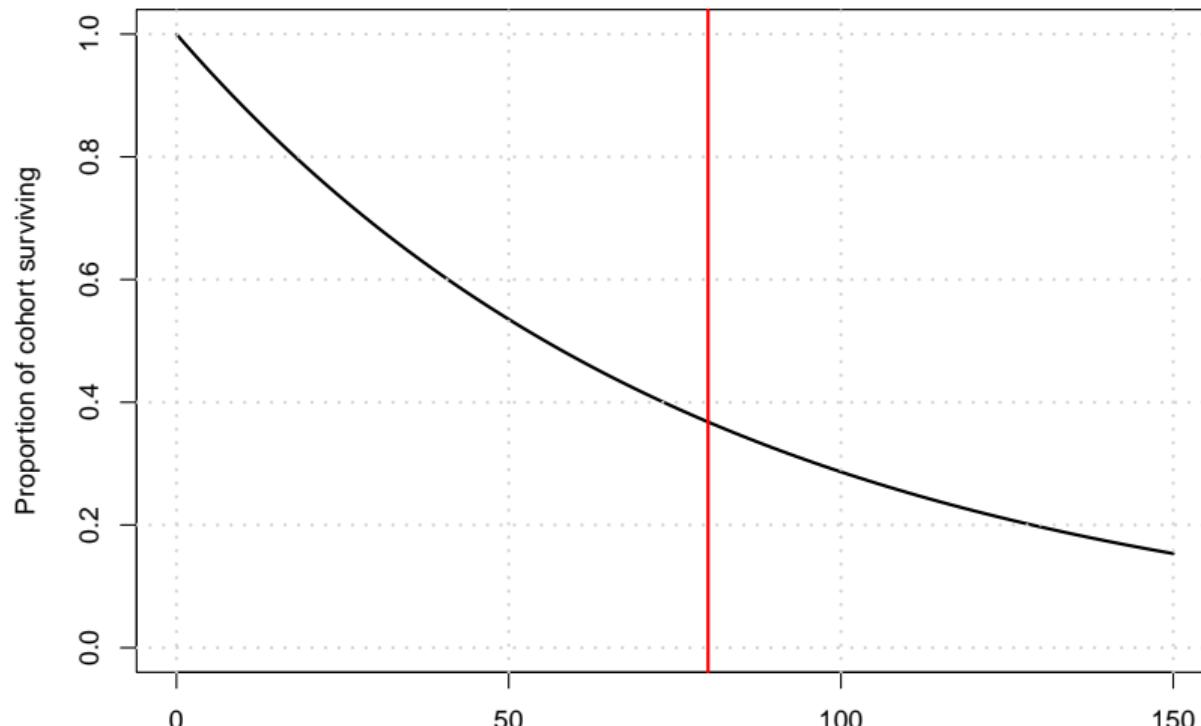
Notez maintenant que

$$\begin{aligned}\frac{d}{dt}N(t) &= -dN_0 e^{-dt} \\ &= -dN(t)\end{aligned}$$

avec $N(0) = N_0$.

⇒ L'EDO $N' = -dN$ fait l'hypothèse que l'espérance de vie à la naissance est distribuée exponentiellement

Fonction de survie, $S(t) = \mathbb{P}(T > t)$, pour une distribution exponentielle avec moyenne 80 ans



Cas où T a une distribution delta de Dirac

Supposons que T ait une distribution delta de Dirac en $t = \omega$, donnant la fonction de survie

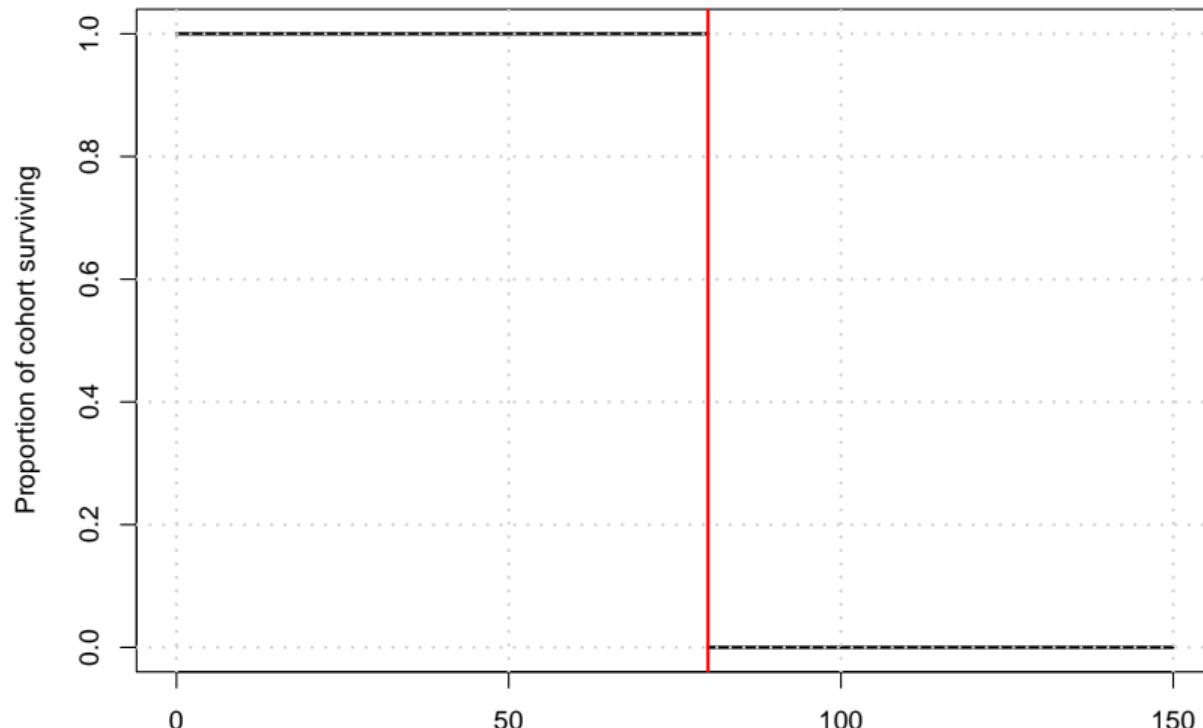
$$S(t) = \begin{cases} 1, & 0 \leq t \leq \omega \\ 0, & t > \omega \end{cases}$$

Alors (4) prend la forme

$$N(t) = \begin{cases} N_0, & 0 \leq t \leq \omega \\ 0, & t > \omega \end{cases} \quad (6)$$

Tous les individus survivent jusqu'au temps ω , puis ils meurent tous au temps ω

Fonction de survie, $S(t) = \mathbb{P}(T > t)$, pour une distribution de Dirac avec moyenne 80 ans



Stochasticité dans les modèles déterministes

Distributions des temps jusqu'aux événements

Deux distributions « extrêmes » et une plus agréable

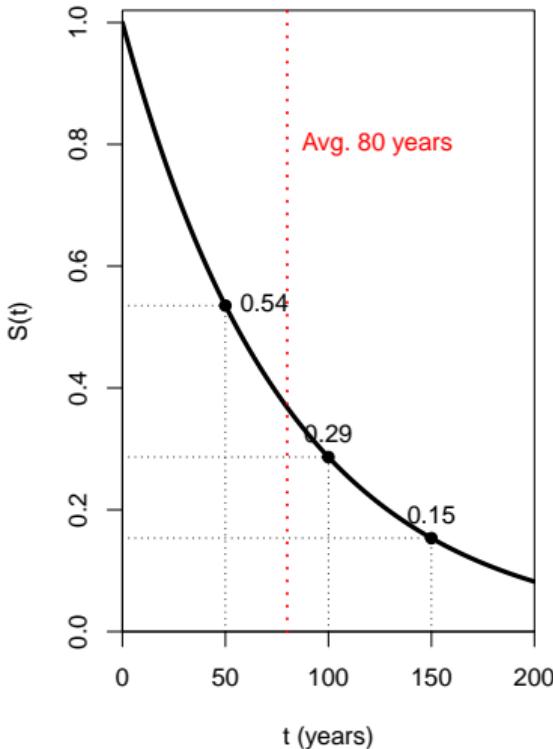
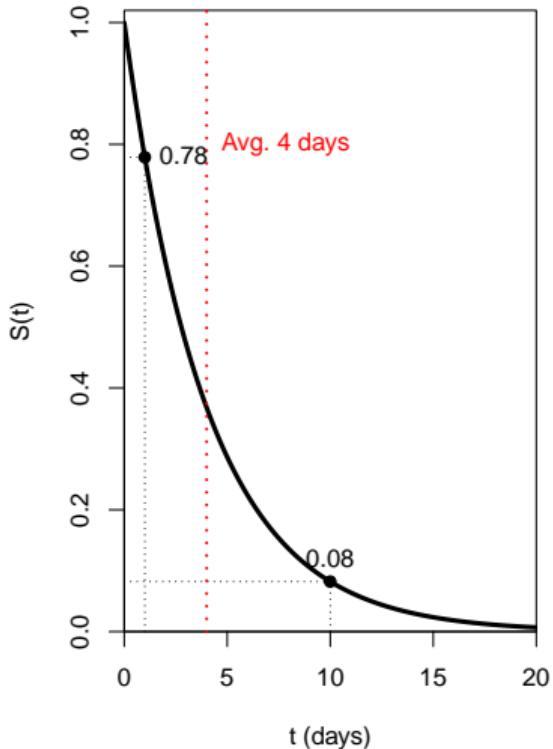
Un modèle de cohorte simple avec décès

Une correction possible au problème de la distribution exponentielle

Temps de séjour dans un modèle de transmission de maladie SIS

Un modèle avec vaccination

Survie pour la distribution exponentielle



Problèmes avec la distribution exponentielle

- ▶ La survie chute rapidement
- ▶ La survie continue bien au-delà de la moyenne

Acceptable si ce qui compte est la durée moyenne de séjour dans un compartiment (par ex., dynamique à long terme)

Plus délicat si l'on s'intéresse à la dynamique à court terme

- ▶ La distribution exponentielle avec paramètre θ a la même moyenne et écart-type $1/\theta$, c'est-à-dire qu'un seul paramètre contrôle la moyenne et la dispersion autour de la moyenne

Les distributions exponentielles sont « mauvaises » mais aussi cool

X_1 et X_2 2 v.a. i.i.d. (indépendantes et identiquement distribuées) avec paramètres θ_1 et θ_2 . Alors la fonction de densité de probabilité de la v.a. $Z = X_1 + X_2$ est donnée par la convolution

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{\infty} f_{X_1}(x_1) f_{X_2}(z - x_1) dx_1 \\ &= \int_0^z \theta_1 e^{-\theta_1 x_1} \theta_2 e^{-\theta_2(z-x_1)} dx_1 \\ &= \theta_1 \theta_2 e^{-\theta_2 z} \int_0^z e^{(\theta_2 - \theta_1)x_1} dx_1 \\ &= \begin{cases} \frac{\theta_1 \theta_2}{\theta_2 - \theta_1} (e^{-\theta_1 z} - e^{-\theta_2 z}) & \text{si } \theta_1 \neq \theta_2 \\ \theta^2 z e^{-\theta z} & \text{si } \theta_1 = \theta_2 =: \theta \end{cases} \end{aligned} \tag{7}$$

L'outil que nous utilisons

Theorem 1

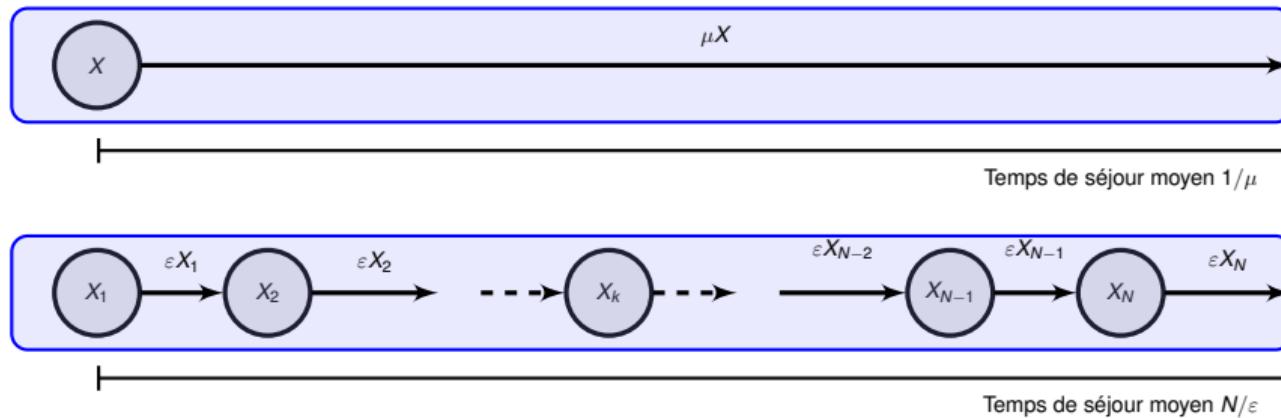
Soit X_i des variables aléatoires indépendantes distribuées exponentiellement avec paramètre ξ et $Y = \sum_{i=1}^n X_i$

Alors la variable aléatoire $Y \sim E(n, \xi)$, une distribution d'Erlang avec paramètre de forme n et paramètre d'échelle ξ

(Distribution d'Erlang : distribution Gamma avec paramètre de forme entier)

Conséquences pour les modèles compartimentaux

Si n compartiments sont traversés successivement par des individus, avec chaque compartiment ayant un taux de sortie de $1/\xi$ (ou un temps de séjour moyen de ξ), alors le temps de séjour depuis l'entrée dans le premier compartiment jusqu'à la sortie du dernier est distribué selon Erlang avec moyenne $E(Y) = n\xi$ et variance $\text{Var}(Y) = n\xi^2$



Exemple : périodes d'incubation de la MVE

Considérons la période d'incubation de la Maladie à Virus Ebola. Pendant la crise de MVE de 2014 en Afrique de l'Ouest, l'Équipe de réponse Ebola de l'OMS a estimé les périodes d'incubation dans un article de 2015

Le tableau S2 dans les Informations Supplémentaires de cet article donne le meilleur ajustement pour la distribution des périodes d'incubation de la MVE comme une distribution Gamma avec moyenne 10,3 jours et écart-type 8,2, c'est-à-dire $n\varepsilon = 10,3$ et $\varepsilon\sqrt{n} = 8,2$

De ceci, $\varepsilon = 8,2^2/10,3 \simeq 6,53$ et $n = 10,3^2/8,2^2 \simeq 1,57$. Cependant, c'est une distribution Gamma

Switching to a compartmental model approach

To use multiple compartments to better fit residence times, we need to find the closest possible Erlang distribution to this Gamma distribution

⇒ compute RSS errors between data points generated from the given Gamma distribution and an Erlang

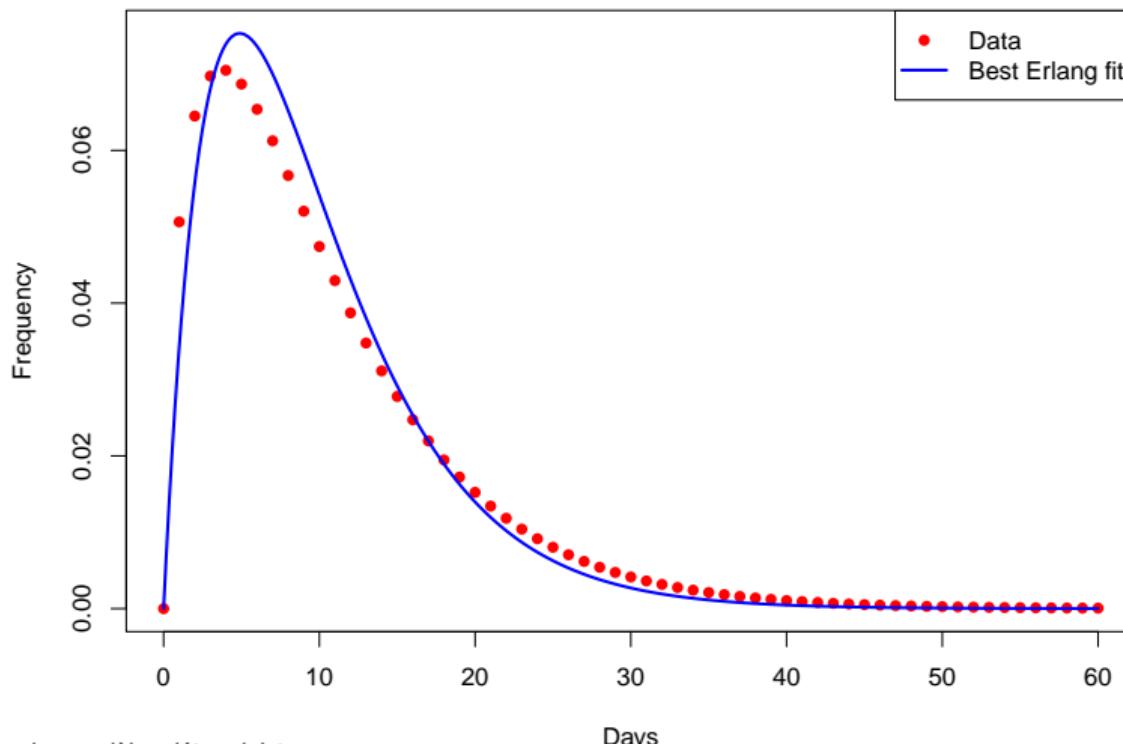
```
error_Gamma <- function(theta,shape,t,d) {  
  test_points <- dgamma(t, shape = shape, scale = theta)  
  ls_error <- sum((d-test_points)^2)  
  return(ls_error)  
}
```

```
optimize_gamma <- function(t,d) {  
  max_shape <- 10  
  error_vector <- mat.or.vec(max_shape,1)  
  scale_vector <- mat.or.vec(max_shape,1)  
  for (i in 1:max_shape) {  
    result_optim <- try(optim(par = 3,  
                               fn = error_Gamma,  
                               lower = 0,  
                               method = "L-BFGS-B",  
                               shape = i,  
                               t = t,  
                               d = d),  
                           TRUE)  
    if (!inherits(result_optim,"try-error")) {  
      error_vector[i] <- result_optim$value  
      scale_vector[i] <- result_optim$par
```

```
    } else {
      error_vector[i] <- NaN
      scale_vector[i] <- NaN
    }
  }
result_optim <- data.frame(seq(1,max_shape),
                           scale_vector,
                           error_vector)
colnames(result_optim) <- c("shape","scale","error")
result_optim <- result_optim[complete.cases(result_optim),]
return(result_optim)
}
```

```
time_points <- seq(0,60)
data_points <- dgamma(time_points, shape = 1.57,
                      scale = 6.53)
# Exécuter la minimisation
optim_fits <- optimize_gamma(time_points,data_points)
# Quelle est la meilleure Erlang pour ajuster les données
idx_best <- which.min(optim_fits$error)
```

Nous trouvons le meilleur ajustement ci-dessous, qui est obtenu en utilisant 2 compartiments



Stochasticité dans les modèles déterministes

Distributions des temps jusqu'aux événements

Deux distributions « extrêmes » et une plus agréable

Un modèle de cohorte simple avec décès

Une correction possible au problème de la distribution exponentielle

Temps de séjour dans un modèle de transmission de maladie SIS

Un modèle avec vaccination

Un modèle SIS

Hypothèses

- ▶ Les individus se rétablissent typiquement de la maladie
- ▶ La maladie ne confère pas d'immunité
- ▶ Il n'y a pas de naissance ou de décès (de la maladie ou naturel)
⇒ Population totale constante $N \equiv N(t) = S(t) + I(t)$
- ▶ L'infection est de type **incidence standard**

Rétablissement

- ▶ Les modèles traditionnels supposent que le rétablissement se produit avec un taux constant γ
- ▶ Ici, parmi les individus qui deviennent infectieux au temps t_0 , une fraction $S(t - t_0)$ reste infectieuse au temps $t \geq t_0$
- ▶ \Rightarrow Pour $t \geq 0$, $S(t)$ est une fonction de survie. En tant que telle, elle vérifie $S(0) = 1$ et S est non négative et non croissante

Modèle pour les individus infectieux

Puisque N est constant, $S(t) = N - I(t)$ et nous n'avons besoin de considérer que l'équation suivante (où S est utilisé pour plus de clarté)

$$I(t) = I_0(t) + \int_0^t \beta \frac{S(u)I(u)}{N} S(t-u) du \quad (8)$$

- ▶ $I_0(t)$ nombre d'individus qui étaient infectieux au temps $t = 0$ et le sont toujours au temps t
 - ▶ $I_0(t)$ est non négatif, non croissant, et tel que $\lim_{t \rightarrow \infty} I_0(t) = 0$
- ▶ $S(t-u)$ proportion d'individus qui sont devenus infectieux au temps u et qui le sont toujours au temps t

Expression sous l'intégrale

Équation intégrale pour le nombre d'individus infectieux :

$$I(t) = I_0(t) + \int_0^t \beta \frac{(N - I(u))I(u)}{N} S(t-u) du \quad (8)$$

Le terme

$$\beta \frac{(N - I(u))I(u)}{N} S(t-u)$$

- ▶ $\beta(N - I(u))I(u)/N$ est le taux auquel de nouveaux infectieux sont créés, au temps u
- ▶ multiplier par $S(t - u)$ donne la proportion de ceux qui sont devenus infectieux au temps u et qui le sont toujours au temps t

Sommer sur $[0, t]$ donne le nombre d'individus infectieux au temps t

Cas d'un temps de rétablissement distribué exponentiellement

Supposons $S(t)$ tel que le temps de séjour dans l'état infectieux ait une distribution exponentielle avec moyenne $1/\gamma$, c'est-à-dire, $S(t) = e^{-\gamma t}$

La fonction de condition initiale $I_0(t)$ prend la forme

$$I_0(t) = I_0(0)e^{-\gamma t}$$

avec $I_0(0)$ le nombre d'individus infectieux au temps $t = 0$. Obtenu en considérant la cohorte d'individus initialement infectieux, donnant un modèle tel que (4)

L'équation (8) devient

$$I(t) = I_0(0)e^{-\gamma t} + \int_0^t \beta \frac{(N - I(u))I(u)}{N} e^{-\gamma(t-u)} du \quad (9)$$

Prendre la dérivée temporelle de (9) donne

$$\begin{aligned}I'(t) &= -\gamma I_0(0)e^{-\gamma t} - \gamma \int_0^t \beta \frac{(N - I(u))I(u)}{N} e^{-\gamma(t-u)} du \\&\quad + \beta \frac{(N - I(t))I(t)}{N} \\&= -\gamma \left(I_0(0)e^{-\gamma t} + \int_0^t \beta \frac{(N - I(u))I(u)}{N} e^{-\gamma(t-u)} du \right) \\&\quad + \beta \frac{(N - I(t))I(t)}{N} \\&= \beta \frac{(N - I(t))I(t)}{N} - \gamma I(t)\end{aligned}$$

Ceci est l'équation différentielle ordinaire (EDO) classique de type logistique pour I dans un modèle SIS sans dynamique vitale (pas de naissance ou de décès)

Cas d'une fonction de survie en escalier

Considérons le cas où le temps passé infecté a une fonction de survie

$$S(t) = \begin{cases} 1, & 0 \leq t \leq \omega, \\ 0, & t > \omega. \end{cases}$$

c'est-à-dire, le temps de séjour dans l'état infectieux est une constante $\omega > 0$

Dans ce cas (8) devient

$$I(t) = I_0(t) + \int_{t-\omega}^t \beta \frac{(N - I(u))I(u)}{N} du. \quad (10)$$

Ici, il est plus difficile d'obtenir une expression pour $I_0(t)$. Il est cependant supposé que $I_0(t)$ s'annule pour $t > \omega$

Lorsque différentiée, (10) donne, pour $t \geq \omega$,

$$I'(t) = I'_0(t) + \beta \frac{(N - I(t))I(t)}{N} - \beta \frac{(N - I(t - \omega))I(t - \omega)}{N}.$$

Puisque $I_0(t)$ s'annule pour $t > \omega$, ceci donne l'équation différentielle à retard (EDR)

$$I'(t) = \beta \frac{(N - I(t))I(t)}{N} - \beta \frac{(N - I(t - \omega))I(t - \omega)}{N}.$$

Stochasticité dans les modèles déterministes

Distributions des temps jusqu'aux événements

Deux distributions « extrêmes » et une plus agréable

Un modèle de cohorte simple avec décès

Une correction possible au problème de la distribution exponentielle

Temps de séjour dans un modèle de transmission de maladie SIS

Un modèle avec vaccination

AN EPIDEMIOLOGY MODEL THAT INCLUDES A LEAKY VACCINE WITH A GENERAL WANING FUNCTION

JULIEN ARINO

Department of Mathematics and Statistics
University of Victoria, Victoria B.C., Canada V8W 3P4

K.L. COOKE

Department of Mathematics
Pomona College, Claremont, CA 91711-6348 USA

P. VAN DEN DRIESSCHE

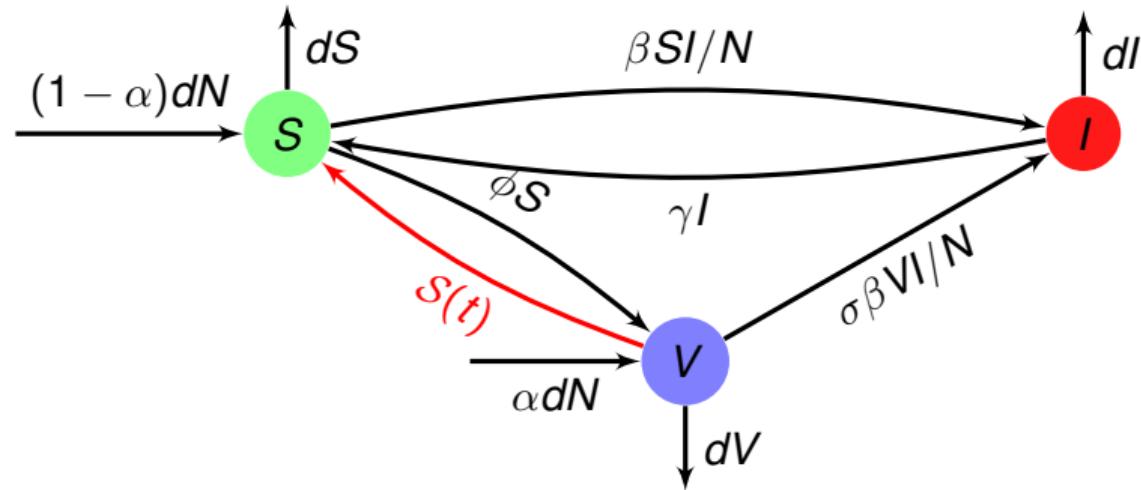
Department of Mathematics and Statistics
University of Victoria, Victoria B.C., Canada V8W 3P4

J. VELASCO-HERNÁNDEZ

Programa de Matemáticas Aplicadas y Computación
Instituto Mexicano del Petróleo, Eje Central Lázaro Cárdenas 152
San Bartolo Atepehuacan, D.F. 07730, Mexico

(Communicated by Linda Allen)

Structure du modèle



Hypothèses sur S

$S(t)$ est une fonction non négative et non croissante avec $S(0^+) = 1$, et telle que $\int_0^\infty S(u)du$ est positive et finie

Donc $S(t)$ est une fonction de survie

Le modèle SIS avec vaccination

$$\frac{dI(t)}{dt} = \beta(S(t) + \sigma V(t))I(t) - (d + \gamma)I(t) \quad (11a)$$

$$V(t) = V_0(t) + \int_0^t (\phi S(u) + \alpha d)S(t-u)e^{-d(t-u)}e^{-\sigma\beta\int_u^t I(x)dx}du \quad (11b)$$

- ▶ αd proportion de nouveau-nés vaccinés
- ▶ $\phi S(u)$ proportion de susceptibles vaccinés
- ▶ $S(t-u)$ fraction de la proportion vaccinée encore dans la classe V $t-u$ unités de temps après y être entrée
- ▶ $e^{-d(t-u)}$ fraction de la proportion vaccinée pas décédée de causes naturelles
- ▶ $e^{-\sigma\beta\int_u^t I(x)dx}$ fraction de la proportion vaccinée pas passée dans la classe infectieuse

Réduction du système en utilisant des fonctions $S(t)$ spécifiques

- ▶ La distribution des temps de déclin étant exponentielle conduit à un système d'EDO
- ▶ $S(t)$ provenant d'une distribution de Dirac conduit à un modèle EDR à retard discret

Soyez conscient (méfiez-vous ?) de ce qui se cache sous le capot

Les EDO et les EDR à retard discret vont bien, mais elles cachent des hypothèses assez fortes

Si vous utilisez une EDO pour un modèle *endémique*, bien : vous travaillez typiquement avec $t \rightarrow \infty$ et l'EDO « utilise » la moyenne de votre paramètre

Si, d'autre part, vous utilisez une EDO pour évaluer le comportement pour des temps qui ne sont pas trop différents de la moyenne de vos paramètres, alors méfiez-vous

Les retards discrets font l'hypothèse que chaque individu dans votre population a le même temps de séjour si un compartiment quelconque se voit appliquer un retard... Vous pourriez vouloir penser aux retards distribués ou à d'autres mécanismes pour compenser ceci

Vous pourriez vouloir utiliser un marteau différent



Pourquoi incorporer la stochasticité ?

Stochasticité dans les modèles déterministes

Chaînes de Markov en temps continu

**Approximations des CMTC par processus de
branchement**

Chaînes de Markov en temps continu

Chaînes de Markov en temps continu

EDO et CMTC

Simuler les CMTC (en théorie)

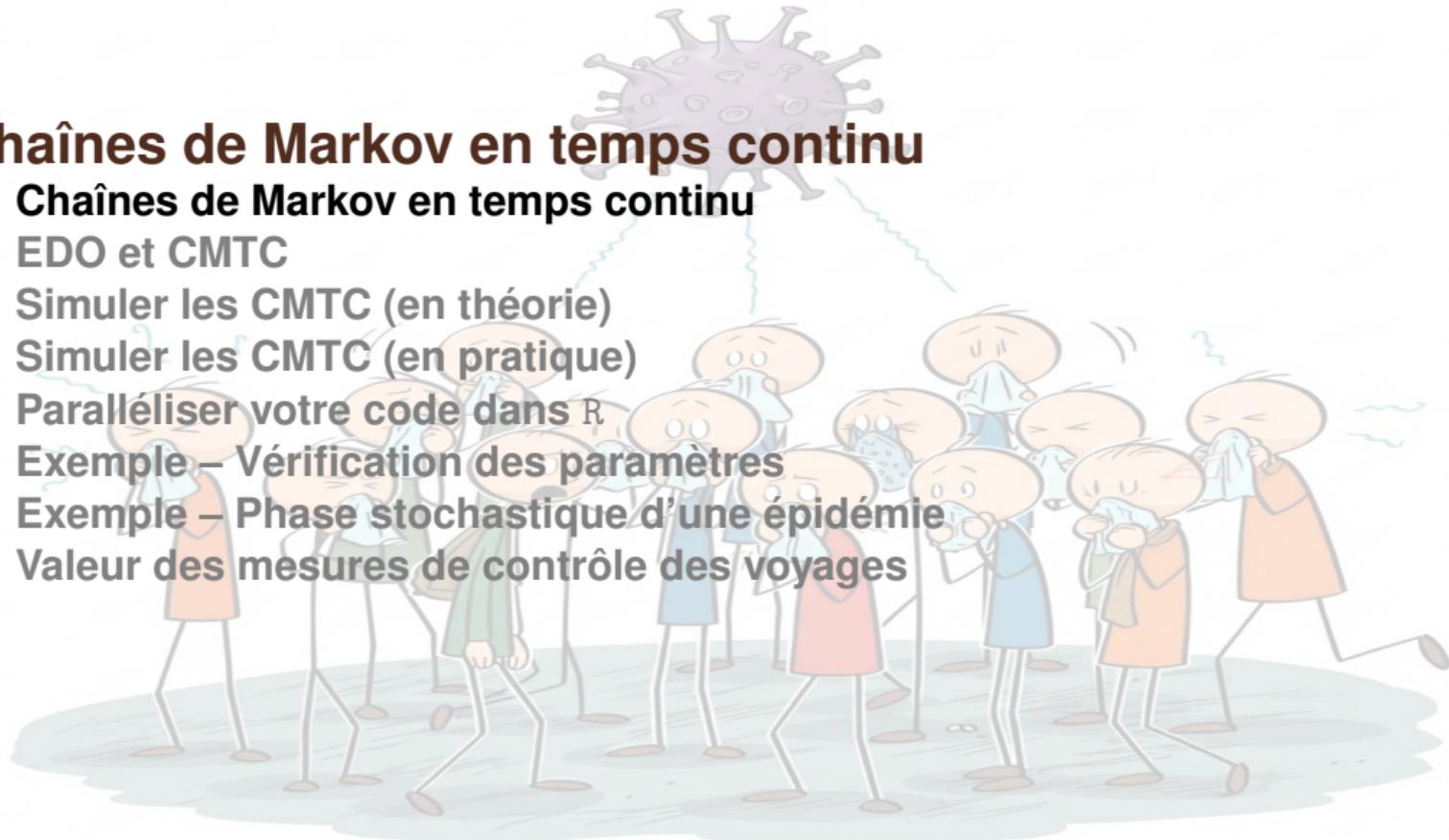
Simuler les CMTC (en pratique)

Paralléliser votre code dans R

Exemple – Vérification des paramètres

Exemple – Phase stochastique d'une épidémie

Valeur des mesures de contrôle des voyages



Du temps discret au temps continu

Chaînes de Markov en temps discret (CMTD)

Un système transite entre états à des pas de temps fixes et discrets ($n = 0, 1, 2, \dots$)

- ▶ Le futur dépend uniquement de l'état présent (Propriété de Markov)
- ▶ Régi par une **matrice de probabilité de transition P** , où P_{ij} est la probabilité de passer de l'état i à j en une étape

Chaînes de Markov en temps continu (CMTC)

Un système peut transiter entre états à **n'importe quel moment**

- ▶ Le temps passé dans un état est une **variable aléatoire continue**
- ▶ Le « temps de séjour » dans tout état i suit une **distribution exponentielle** paramétrée par un *taux de sortie q_i*
- ▶ Ceci est une conséquence directe de la Propriété de Markov appliquée au temps continu (l'exponentielle est la seule distribution continue qui est « sans mémoire »)

Taux de transition

Dynamique d'une CMTC définie par des **taux de transition**, pas des probabilités

Definition 2 (Taux de transition)

Pour deux états $i \neq j$, le taux $q_{ij} \geq 0$ est le taux instantané de transition de l'état i à l'état j

- ▶ Pour un petit intervalle de temps Δt , la probabilité de transition de i à j est approximativement $q_{ij}\Delta t$
- ▶ Le **taux de sortie** total de l'état i est $q_i = \sum_{j \neq i} q_{ij}$
- ▶ Le temps passé dans l'état i est une variable aléatoire exponentielle $T_i \sim E(q_i)$

La matrice génératrice

Matrice génératrice (Matrice Q)

Rassembler tous les taux de transition dans une seule matrice Q

- ▶ Hors diagonale : $Q_{ij} = q_{ij}$ pour $i \neq j$ (Le taux de passage de i à j)
- ▶ Diagonale : $Q_{ii} = -q_i = -\sum_{j \neq i} q_{ij}$. (Le négatif du taux de sortie total de i)

Une propriété clé est que toutes les lignes de Q somment à zéro : $\sum_j Q_{ij} = 0$.

Équations de Kolmogorov

Soit $P(t)$ la matrice où $P_{ij}(t) = \mathbb{P}(X(t) = j | X(0) = i)$. Comment $P(t)$ évolue-t-elle dans le temps ?

Équations de Kolmogorov progressives

Décrit le taux de changement de probabilité de finir dans un état cible j

$$\frac{d}{dt} P(t) = P(t)Q$$

Sous forme élémentaire :

$$P'_{ij}(t) = \sum_k P_{ik}(t)Q_{kj}$$

Solution de l'ÉKP

La solution est l'**exponentielle matricielle**

$$P(t) = e^{tQ} = \sum_{k=0}^{\infty} \frac{(tQ)^k}{k!}$$

La matrice génératrice Q « génère » l'évolution du processus

Chaînes de Markov en temps continu

Chaînes de Markov en temps continu

EDO et CMTC

Simuler les CMTC (en théorie)

Simuler les CMTC (en pratique)

Paralléliser votre code dans R

Exemple – Vérification des paramètres

Exemple – Phase stochastique d'une épidémie

Valeur des mesures de contrôle des voyages

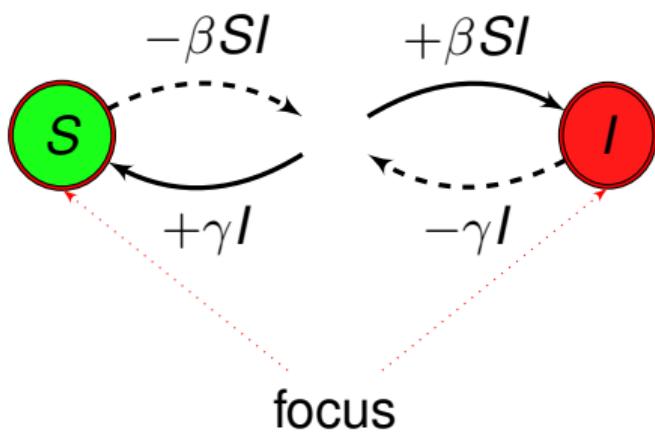
Convertir votre modèle compartimentale EDO en CMTC

Facile comme π :)

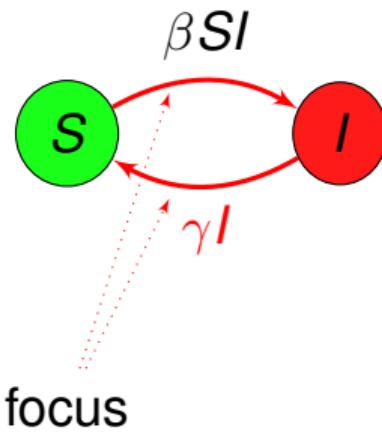
- ▶ Le modèle compartimentale EDO se concentre sur les flux entrants et sortants des compartiments
- ▶ Le modèle EDO a autant d'équations qu'il y a de compartiments
- ▶ Le modèle compartimentale CMTC se concentre sur les transitions
- ▶ Le modèle CMTC a autant de transitions qu'il y a de flèches entre (ou vers ou depuis) les compartiments

EDO vers CMTC : focus sur différents composants

EDO



CMTC



SIS sans démographie

Transition	Effet	Poids	Probabilité
$S \rightarrow S - 1, I \rightarrow I + 1$	nouvelle infection	βSI	$\frac{\beta SI}{\beta SI + \gamma I}$
$S \rightarrow S + 1, I \rightarrow I - 1$	rétablissement d'un infectieux	γI	$\frac{\gamma I}{\beta SI + \gamma I}$

Les états sont S, I

SIS avec démographie

Transition	Effet	Poids	Probabilité
$S \rightarrow S + 1$	naissance d'un susceptible	b	$\frac{b}{b+d(S+I)+\beta SI+\gamma I}$
$S \rightarrow S - 1$	décès d'un susceptible	dS	$\frac{dS}{b+d(S+I)+\beta SI+\gamma I}$
$S \rightarrow S - 1, I \rightarrow I + 1$	nouvelle infection	βSI	$\frac{\beta SI}{b+d(S+I)+\beta SI+\gamma I}$
$I \rightarrow I - 1$	décès d'un infectieux	dI	$\frac{dI}{b+d(S+I)+\beta SI+\gamma I}$
$S \rightarrow S + 1, I \rightarrow I - 1$	rétablissement d'un infectieux	γI	$\frac{\gamma I}{b+d(S+I)+\beta SI+\gamma I}$

Les états sont S, I

Modèle de Kermack & McKendrick

Transition	Effet	Poids	Probabilité
$S \rightarrow S - 1, I \rightarrow I + 1$	nouvelle infection	βSI	$\frac{\beta SI}{\beta SI + \gamma I}$
$I \rightarrow I - 1, R \rightarrow R + 1$	rétablissement d'un infectieux	γI	$\frac{\gamma I}{\beta SI + \gamma I}$

Les états sont S, I, R

Chaînes de Markov en temps continu

Chaînes de Markov en temps continu

EDO et CMTC

Simuler les CMTC (en théorie)

Simuler les CMTC (en pratique)

Paralléliser votre code dans R

Exemple – Vérification des paramètres

Exemple – Phase stochastique d'une épidémie

Valeur des mesures de contrôle des voyages

L'algorithme de Gillespie

- ▶ Aussi appelé l'algorithme de simulation stochastique (SSA)
- ▶ Dérivé en 1976 par Daniel Gillespie
- ▶ Génère des solutions possibles pour les CMTC
- ▶ Extrêmement simple, donc vaut la peine d'apprendre à implémenter ; il existe cependant des paquets que vous pouvez utiliser (voir plus tard)

L'algorithme de Gillespie

Supposons que le système a un état $\mathbf{x}(t)$ avec condition initiale $\mathbf{x}(t_0) = \mathbf{x}_0$ et fonctions de propension a_i des réactions élémentaires

définir $t \leftarrow t_0$ et $\mathbf{x}(t) \leftarrow \mathbf{x}_0$

tant que $t \leq t_f$

- $\xi_t \leftarrow \sum_j a_j(\mathbf{x}(t))$
- Tirer τ_t de $T \sim \mathcal{E}(\xi_t)$
- Tirer ζ_t de $\mathcal{U}([0, 1])$
- Trouver r , le plus petit entier tel que $\sum_{k=1}^j a_k(\mathbf{x}(t)) > \zeta_t \sum_j a_j(\mathbf{x}(t)) = \zeta_t \xi_t$
- Effectuer la prochaine réaction (celle indexée r)
- $t \leftarrow t + \tau_t$

Tirer aléatoirement d'une distribution exponentielle

Si vous n'avez pas de générateur de nombres aléatoires de distribution exponentielle.. Nous voulons τ_t de $T \sim \mathcal{E}(\xi_t)$, c'est-à-dire, T a une fonction de densité de probabilité

$$f(x, \xi_t) = \xi_t e^{-\xi_t x} \mathbf{1}_{x \geq 0}$$

Utiliser la fonction de distribution cumulative $F(x, \xi_t) = \int_{-\infty}^x f(s, \xi_t) ds$

$$F(x, \xi_t) = (1 - e^{-\xi_t x}) \mathbf{1}_{x \geq 0}$$

qui a des valeurs dans $[0, 1]$. Donc tirer ζ de $\mathcal{U}([0, 1])$ et résoudre $F(x, \xi_t) = \zeta$ pour x

$$\begin{aligned} F(x, \xi_t) = \zeta &\Leftrightarrow 1 - e^{-\xi_t x} = \zeta \\ &\Leftrightarrow e^{-\xi_t x} = 1 - \zeta \\ &\Leftrightarrow \xi_t x = -\ln(1 - \zeta) \\ &\Leftrightarrow x = \boxed{\frac{-\ln(1 - \zeta)}{\xi_t}} \end{aligned}$$

L'algorithme de Gillespie (modèle SIS avec seulement l'éq. I)

définir $t \leftarrow t_0$ et $I(t) \leftarrow I(t_0)$

tant que $t \leq t_f$

- $\xi_t \leftarrow \beta(P^* - i)i + \gamma i$
- Tirer τ_t de $T \sim \mathcal{E}(\xi_t)$
- $v \leftarrow [\beta(P^* - i)i, \xi_t] / \xi_t$
- Tirer ζ_t de $\mathcal{U}([0, 1])$
- Trouver pos tel que $v_{pos-1} \leq \zeta_t \leq v_{pos}$
- selon pos
 - 1 : Nouvelle infection, $I(t + \tau_t) = I(t) + 1$
 - 2 : Fin de la période infectieuse, $I(t + \tau_t) = I(t) - 1$
- $t \leftarrow t + \tau_t$

Parfois Gillespie tourne mal

- ▶ Rappelons que le temps inter-événements est distribué exponentiellement
- ▶ Étape critique de l'algorithme de Gillespie :
 - ▶ $\xi_t \leftarrow$ poids de tous les événements possibles (*propension*)
 - ▶ Tirer τ_t de $T \sim \mathcal{E}(\xi_t)$
- ▶ Donc le temps inter-événements $\tau_t \rightarrow 0$ si ξ_t devient très grand pour un certain t
- ▶ Ceci peut causer le ralentissement de la simulation jusqu'à l'arrêt

Exemple : un processus de naissance et de mort

- ▶ Individus nés à un taux *per capita* b
- ▶ Individus meurent à un taux *per capita* d
- ▶ Implémentons cela en utilisant le Gillespie classique

(Voir `simulate_birth_death_CTMC.R` sur le dépôt GitHub du cours)

L'algorithme de Gillespie (modèle naissance-mort)

définir $t \leftarrow t_0$ et $N(t) \leftarrow N(t_0)$

tant que $t \leq t_f$

- $\xi_t \leftarrow (b + d)N(t)$
- Tirer τ_t de $T \sim \mathcal{E}(\xi_t)$
- $v \leftarrow [bN(t), \xi_t] / \xi_t$
- Tirer ζ_t de $\mathcal{U}([0, 1])$
- Trouver pos tel que $v_{pos-1} \leq \zeta_t \leq v_{pos}$
- selon pos
 - 1 : Naissance, $N(t + \tau_t) = N(t) + 1$
 - 2 : Décès, $N(t + \tau_t) = N(t) - 1$
- $t \leftarrow t + \tau_t$

```
birth_death_CTMC = function(b = 0.01, d = 0.01) {  
    t_0 = 0      # Initial time  
    N_0 = 100    # Initial population  
  
    # Vectors to store time and state. Initialise with initial condition.  
    t = t_0  
    N = N_0  
  
    t_f = 1000  # Final time  
  
    # Track the current time and state (could just check last entry in t  
    # and N, but will take more operations)  
    t_curr = t_0  
    N_curr = N_0  
    while (t_curr<=t_f) {  
        xi_t = (b+d)*N_curr
```

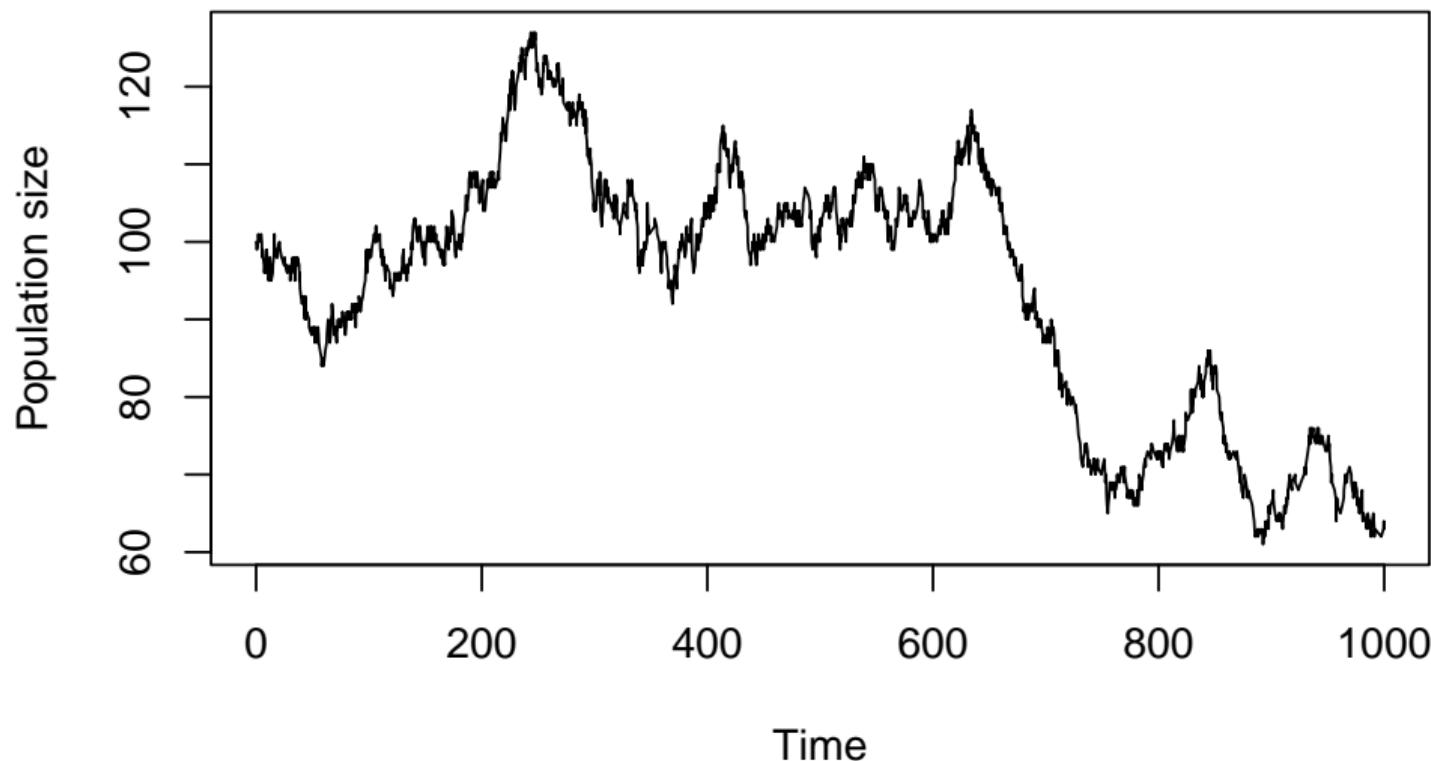
```

if (N_curr == 0) {
  break # Avoid error with rexp when xi_t = 0
}
tau_t = rexp(1, rate = xi_t)
t_curr = t_curr+tau_t
v = c(b*N_curr, xi_t)/xi_t
zeta_t = runif(n = 1)
pos = findInterval(zeta_t, v)+1
switch(pos,
  { N_curr = N_curr+1}, # Birth
  { N_curr = N_curr-1}) # Death
N = c(N, N_curr)
t = c(t, t_curr)
}
plot(t, N, type = "l",
      xlab = "Time", ylab = "Population size",
      main = paste("Birth-death CTMC with b =", b, "and d =", d))

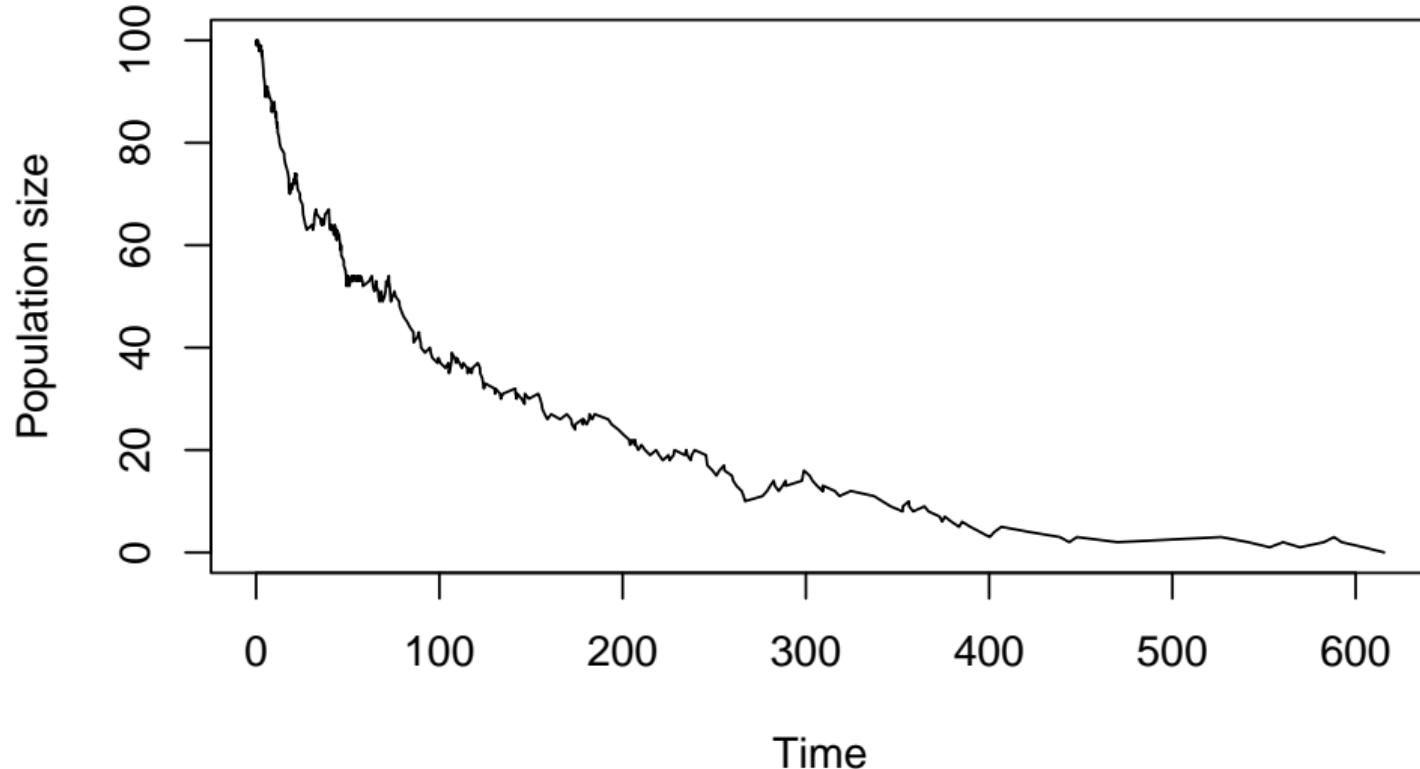
```

}

Birth-death CTMC with $b = 0.01$ and $d = 0.01$



Birth-death CTMC with $b = 0.01$ and $d = 0.02$

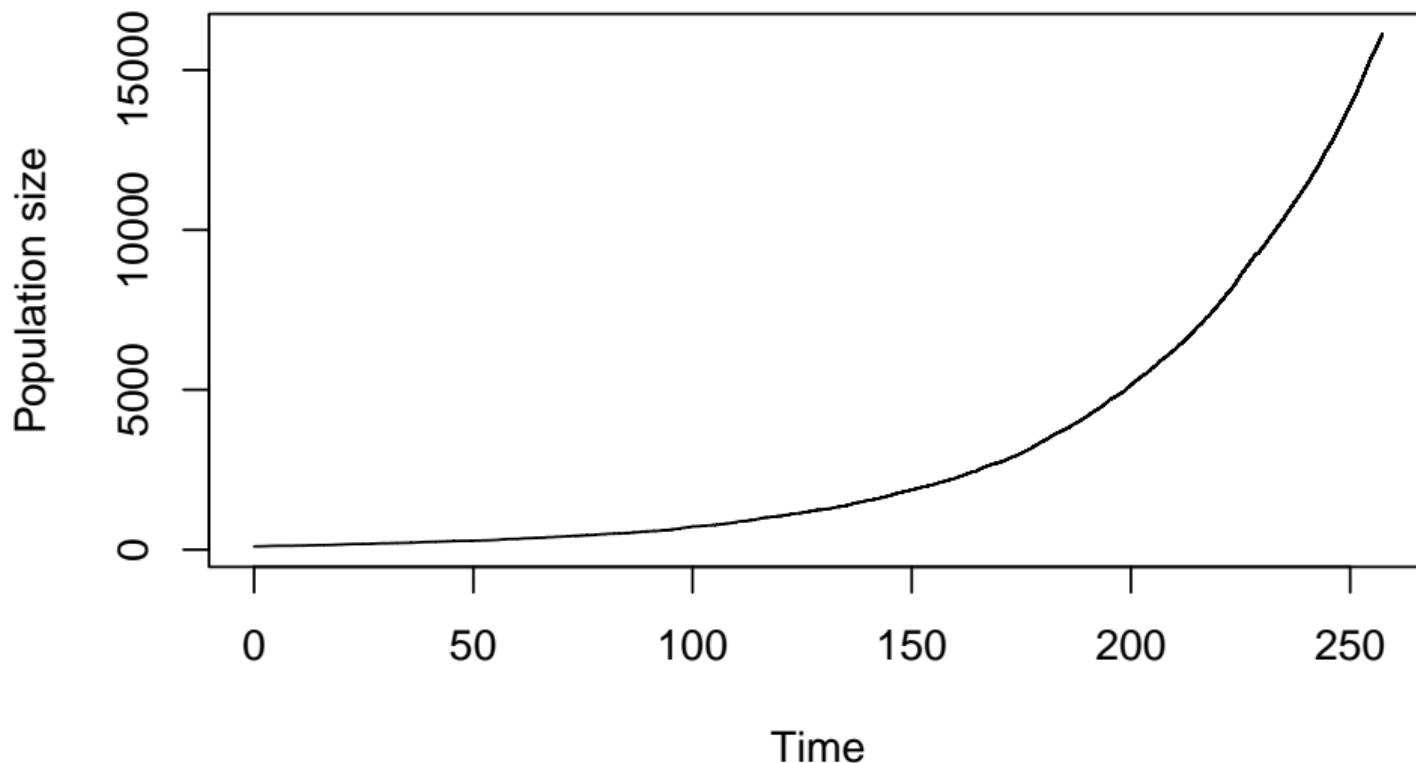


$$b = 0.03 \text{ & } d = 0.01\dots$$

Nous voulons exécuter la fonction avec ces valeurs de paramètres mais je sais à l'avance que cela ne fonctionnera pas bien, donc modifions un peu la fonction.
Nous ajoutons un test :

```
if (t[length(t)]-t[(length(t)-1)] < 1e-8) {  
    # If the time step is too small, stop the simulation  
    message("Stopping simulation because time step is too small")  
    break  
}
```

Birth-death CTMC with $b = 0.03$ and $d = 0.01$



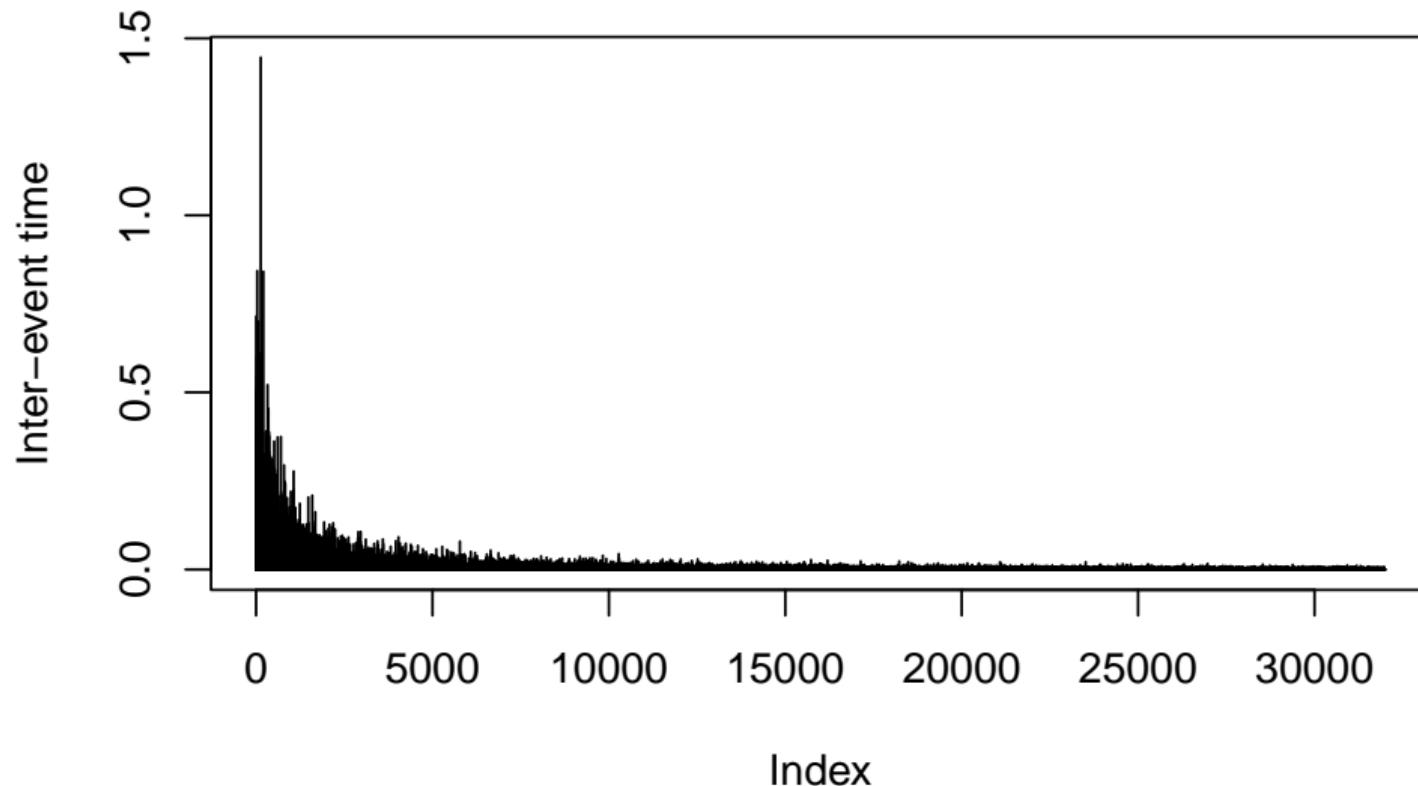
Last one did not go well

- ▶ Wanted 1000 time units (days ?)
- ▶ Interrupted at 257.3144961 because of the test
(Slide with $b < d$: sim stopped because the population went extinct, I did not stop it !)
- ▶ At stop time
 - ▶ $N = 1.6116 \times 10^4$
 - ▶ $|N| = 31977$ (et $|t|$ aussi, bien sûr !)
 - ▶ le temps avançait lentement

```
tail(diff(results$t))
```

```
## [1] 4.201952e-04 4.017531e-03 7.198722e-04 2.877816e-04 4.953929e-04
## [6] 2.904073e-09
```

Inter-event time for birth-death CTMC with $b=0.03$ and $d=0.01$



Chaînes de Markov en temps continu

Chaînes de Markov en temps continu

EDO et CMTC

Simuler les CMTC (en théorie)

Simuler les CMTC (en pratique)

Paralléliser votre code dans R

Exemple – Vérification des paramètres

Exemple – Phase stochastique d'une épidémie

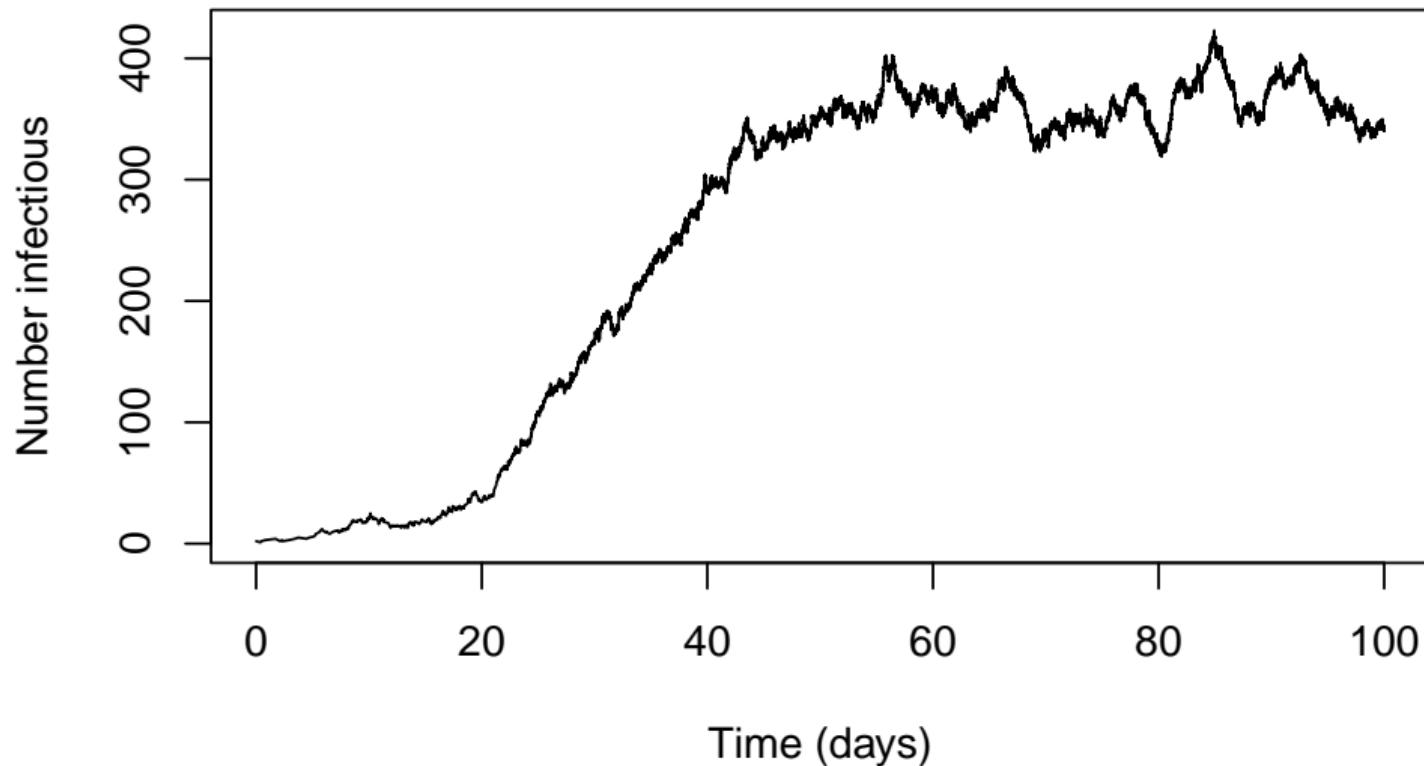
Valeur des mesures de contrôle des voyages

Tau-leaping (et les paquets) à la rescousse !

- ▶ Méthode d'*approximation* (comparée au Gillespie classique, qui est exact)
- ▶ Grosso modo : considérer des « groupes » d'événements au lieu d'événements individuels
- ▶ Bonne nouvelle : GillespieSSA2 et adaptivetau, deux paquets standard pour SSA dans R, implémentent le tau leaping

```
library(GillespieSSA2)
Pop <- 1000
I_0 <- 2
IC <- c(S = (Pop-I_0), I = I_0)
gamma = 1/3
#  $R_0 = \beta/\gamma * S_0$ , donc  $\beta = R_0 * \gamma / S_0$ 
beta = as.numeric(1.5*gamma/IC["S"])
params <- c(gamma = gamma, beta = beta)
t_f = 100
reactions <- list(
  reaction("beta*S*I", c(S=-1, I=+1), "new_infection"),
  reaction("gamma*I", c(S=+1, I=-1), "recovery")
)
set.seed(NULL)
```

```
sol <- ssa(  
    initial_state = IC,  
    reactions = reactions,  
    params = params,  
    method = ssa_exact(),  
    final_time = t_f,  
)  
plot(sol$time, sol$state[, "I"], type = "l",  
     xlab = "Time (days)", ylab = "Number infectious")
```



Chaînes de Markov en temps continu

Chaînes de Markov en temps continu

EDO et CMTC

Simuler les CMTC (en théorie)

Simuler les CMTC (en pratique)

Paralléliser votre code dans R

Exemple – Vérification des paramètres

Exemple – Phase stochastique d'une épidémie

Valeur des mesures de contrôle des voyages

Parallélisation

Pour voir plusieurs réalisations : bonne idée de paralléliser, puis interpoler les résultats. Écrire une fonction, par ex., `run_one_sim` qui.. exécute une simulation

Utiliser certains mécanismes de parallélisation pour exécuter `run_one_sim` en parallèle. Une façon facile de le faire est d'utiliser une version parallèle de `lapply`, qui applique une fonction à une liste

Ici, je montre la parallélisation en utilisant un paquet assez récent appelé `future` (et `future.apply`, qui contient l'équivalent pertinent de `lapply`)

J'illustre également une autre bibliothèque SSA que je trouve moins délicate sous Windows car les réactions ne sont pas précompilées : `adaptivetau`

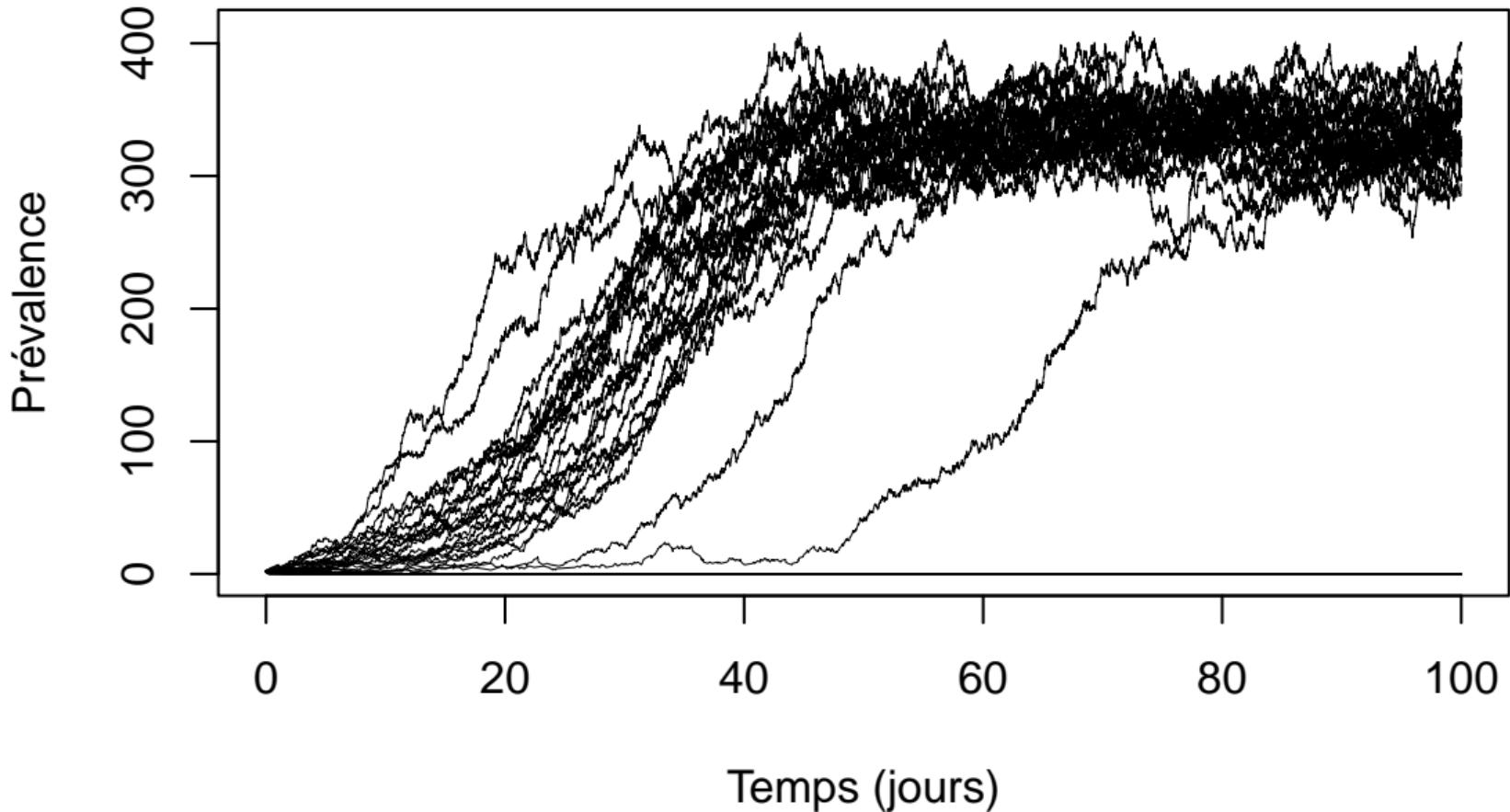
Partie commune – la fonction utilisée I

```
run_one_sim = function(params) {
  IC <- c(S = (params$Pop-params$I_0), I = params$I_0)
  params_local <- c(gamma = params$gamma, beta = params$beta)
  reactions <- list(
    # propensity function effects name for reaction
    reaction("beta*S*I", c(S=-1,I=+1), "new_infection"),
    reaction("gamma*I", c(S=+1,I=-1), "recovery")
  )
  set.seed(NULL)
  sol <- ssa(
    initial_state = IC,
    reactions = reactions,
    params = params_local,
    method = ssa_exact(),
```

Partie commune – la fonction utilisée II

```
final_time = params$t_f,
log_firings = TRUE      # This way we keep track of events
)
# Interpolate result (just I will do)
wanted_t = seq(from = 0, to = params$t_f, by = 0.01)
sol$interp_I = approx(x = sol$time, y = sol$state[, "I"],
                      xout = wanted_t)
names(sol$interp_I) = c("time", "I")
# Return result
return(sol)
}
```

CMTC avec $R_0 = 1.5$



Bénéfices de la parallélisation

Faisons tourner 100 sims entre `tictoc::tic()` and `tictoc::toc()`, donnant 66.958 sec elapsed, puis la version séquentielle

```
tictoc::tic()  
SIMS = lapply(X = 1:params$number_sims,  
              FUN = function(x) run_one_sim(params))  
tictoc::toc()
```

ce qui donne 318.141 sec elapsed sur un Intel(R) Core(TM) i9-8950HK CPU @ 2.90GHz 6C/12T (4.75× plus rapide) ou 12.067 sec elapsed versus 258.985 sec elapsed sur un processeur AMD Ryzen Threadripper 3970X 32-Core/64-Threads (21.46× plus rapide !)

Chaînes de Markov en temps continu

Chaînes de Markov en temps continu

EDO et CMTC

Simuler les CMTC (en théorie)

Simuler les CMTC (en pratique)

Paralléliser votre code dans R

Exemple – Vérification des paramètres

Exemple – Phase stochastique d'une épidémie

Valeur des mesures de contrôle des voyages

Exemple simple – Naissance et mort

Nous considérons un modèle de démographie avec un compartiment N , avec flux entrant de naissance constant et décès *per capita*

$$\frac{dN}{dt} = b - dN, \quad b, d > 0$$

- ▶ Équilibre $N^* = b/d$
- ▶ Si d est connu, ce modèle est **identifiable** : il y a un unique $b = dN^*$ tel que $N(t) \rightarrow N^*$

La CMTC correspondante prend la forme $N(t) \in \mathbb{N}$, avec les transitions

- ▶ $N \rightarrow N + 1$ au taux b
- ▶ $N \rightarrow N - 1$ au taux dN

Une petite expérience

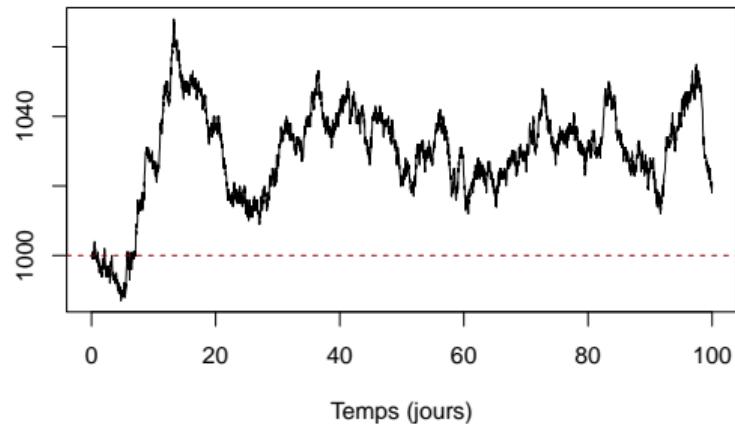
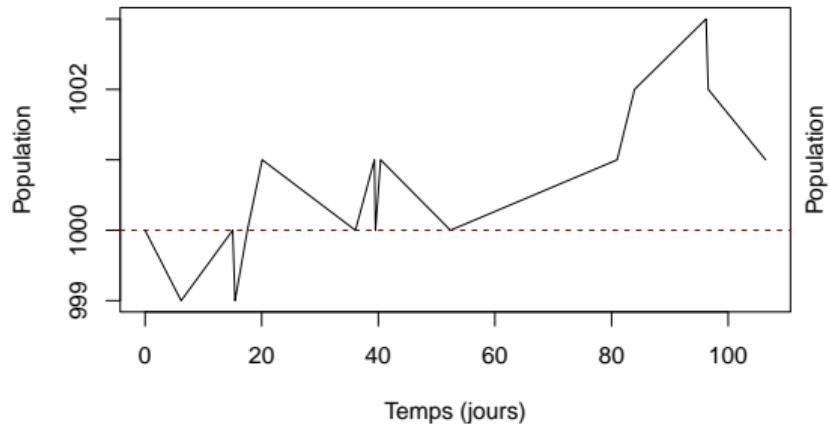
Supposons qu'on ait une population humaine de 1 000 personnes

Prétendons qu'on ne connaît pas vraiment l'ordre de grandeur de d ($1/d$ est la durée moyenne de vie). On veut garder une population autour de 1 000 individus

J'illustre 2 cas :

- ▶ mortalité $d = 1/45$ ans ($d = 1/(45 \cdot 365)$ avec unités de temps les jours)
- ▶ mortalité $d = 1/45$ jours

Pensez à une population de 1 000 personnes. Laquelle des figures ci-dessous vous semble convenir mieux en terme du nombre d'évènements ayant lieu tous les jours ?



Chaînes de Markov en temps continu

Chaînes de Markov en temps continu

EDO et CMTC

Simuler les CMTC (en théorie)

Simuler les CMTC (en pratique)

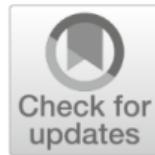
Paralléliser votre code dans R

Exemple – Vérification des paramètres

Exemple – Phase stochastique d'une épidémie

Valeur des mesures de contrôle des voyages

ORIGINAL ARTICLE

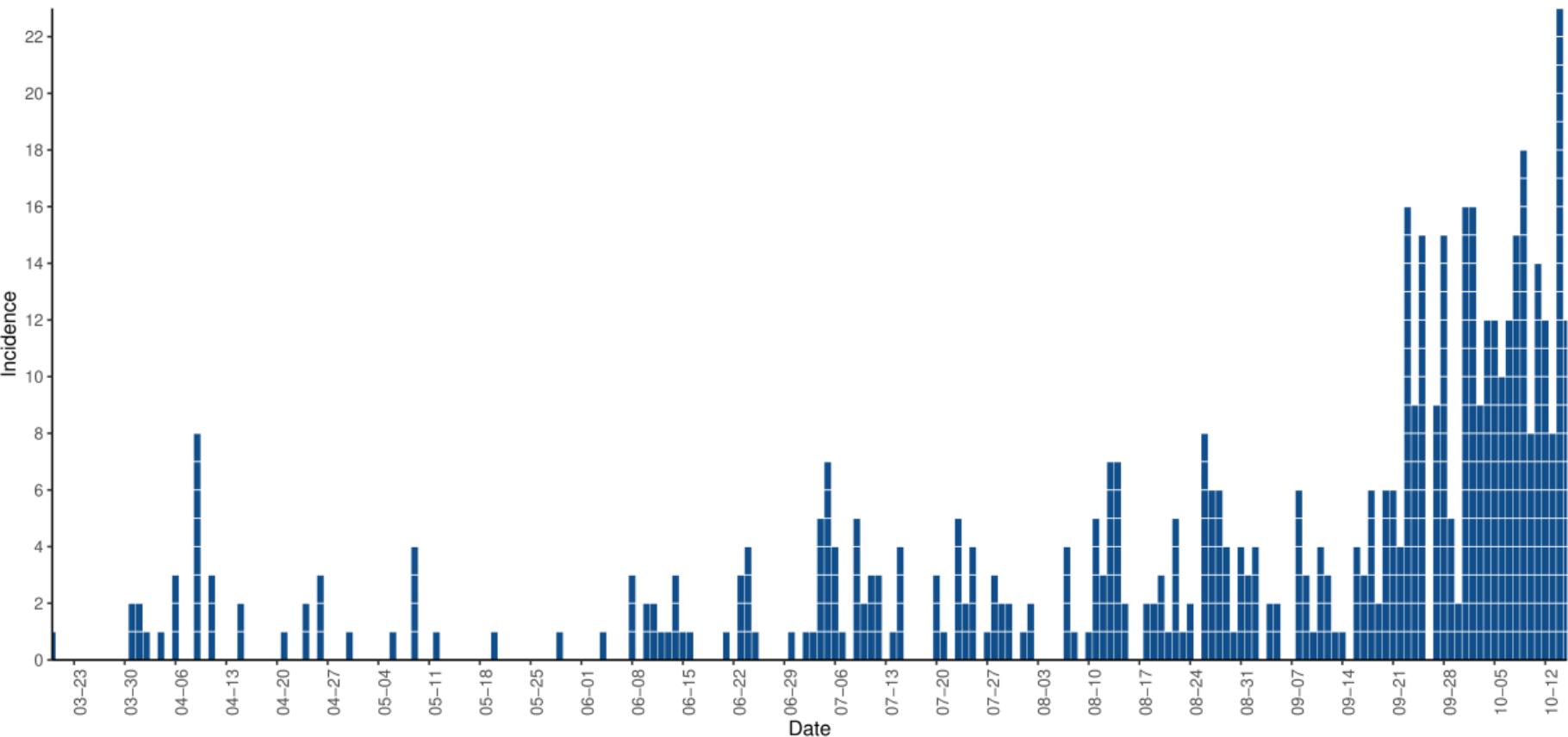


Effect of Movement on the Early Phase of an Epidemic

Julien Arino¹ · Evan Milliken² 

Received: 6 July 2021 / Accepted: 29 August 2022 / Published online: 23 September 2022

© The Author(s), under exclusive licence to Society for Mathematical Biology 2022



Enquête sur les types d'épidémie à l'aide d'un CMTC SIS simple

$$\mathbf{X}(t) = (S^A(t), I^A(t))$$

CMTC $\mathbf{X}(t)$ caractérisée par les transitions

Description	Transition	Taux
Infection	$(S^A, I^A) \rightarrow (S^A - 1, I^A + 1)$	$\beta^A S^A I^A$
Rétablissement	$(S^A, I^A) \rightarrow (S^A + 1, I^A - 1)$	$\gamma^A I^A$

Enquête sur les types d'épidémie à l'aide d'un CMTC SIS simple avec une variante

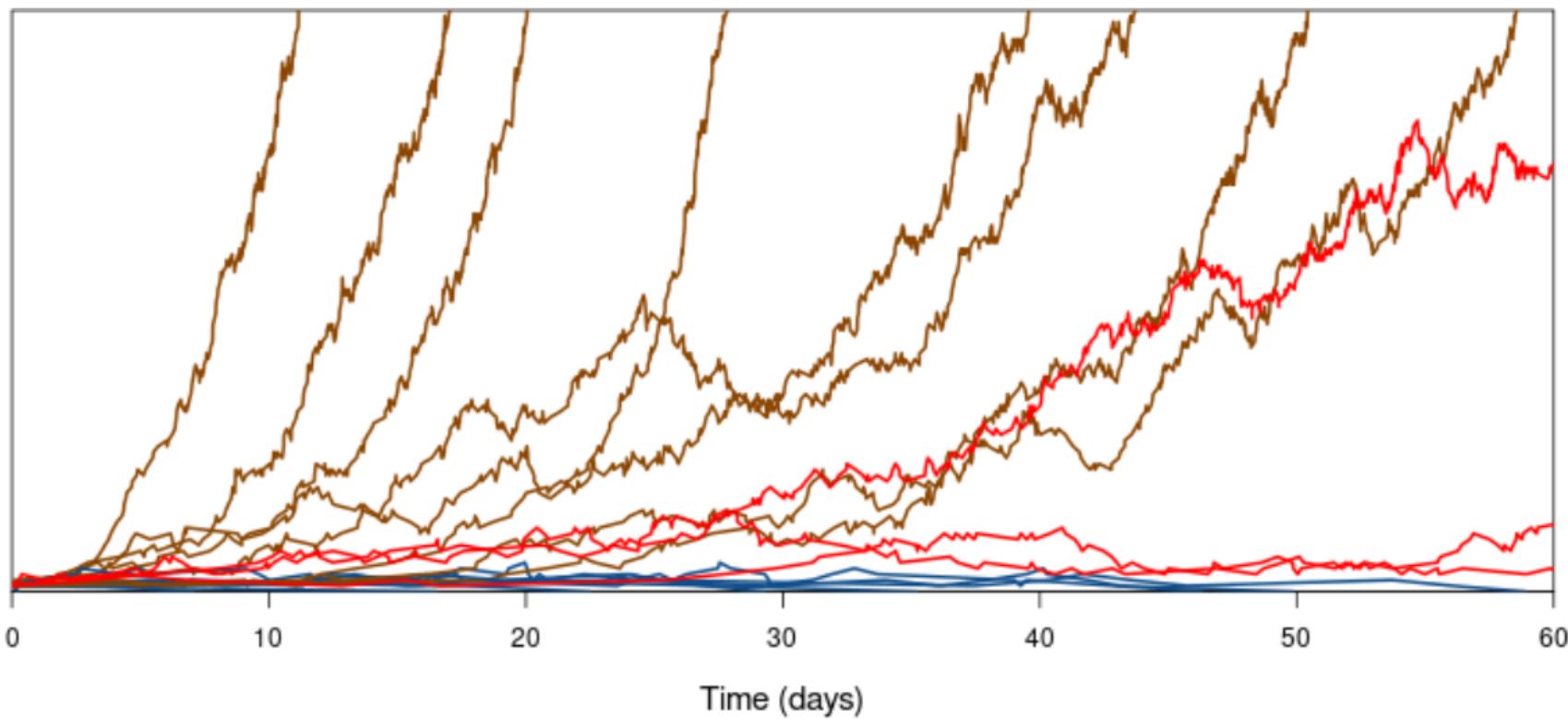
Une chaîne régulière de ce type a $I = 0$ comme seul état absorbant

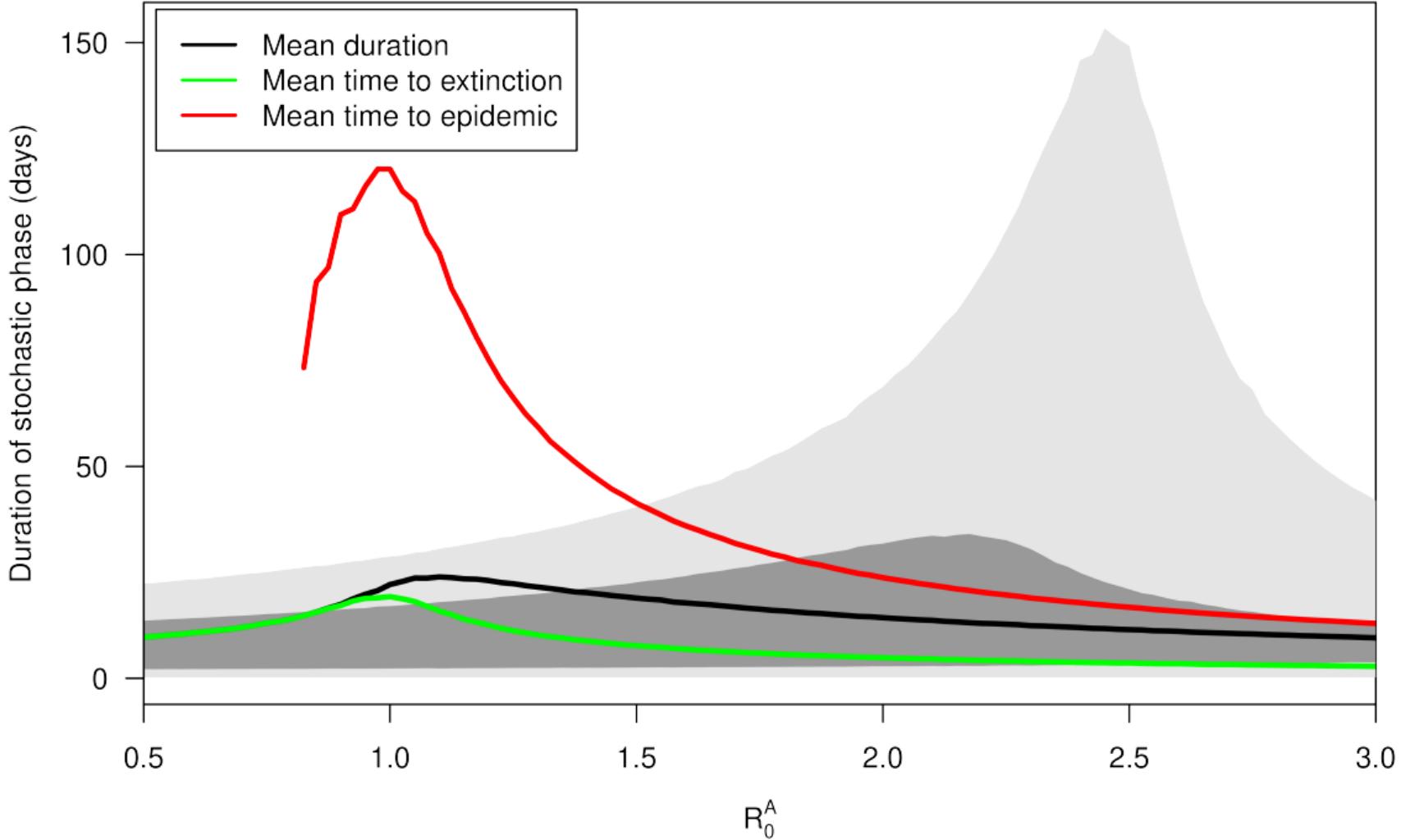
Nous ajoutons un autre état absorbant : si $I = \hat{I}$, alors la chaîne a *quitté* la phase stochastique et est dans une phase quasi-déterministe avec croissance exponentielle

En faisant cela, les mesures de temps d'absorption deviennent utilisables en plus des mesures de temps de premier passage

Et la question devient : combien de temps la chaîne « persiste » avant d'être absorbée ? Nous définissons la trajectoire inter-absorption comme la phase stochastique

Prevalence





Problème de la valeur de la borne supérieure \hat{I}

- ▶ Choisir \hat{I} trop petit et la phase stochastique ne durera pas longtemps
- ▶ Choisir \hat{I} trop grand et l'absorption ne se fera qu'au DFE
- ▶ Alors, comment choisir \hat{I} ?
 - ▶ Une formule de Whittle (1955)
 - ▶ Processus de branchement multitype (MTBP)

Chaînes de Markov en temps continu

Chaînes de Markov en temps continu

EDO et CMTC

Simuler les CMTC (en théorie)

Simuler les CMTC (en pratique)

Paralléliser votre code dans R

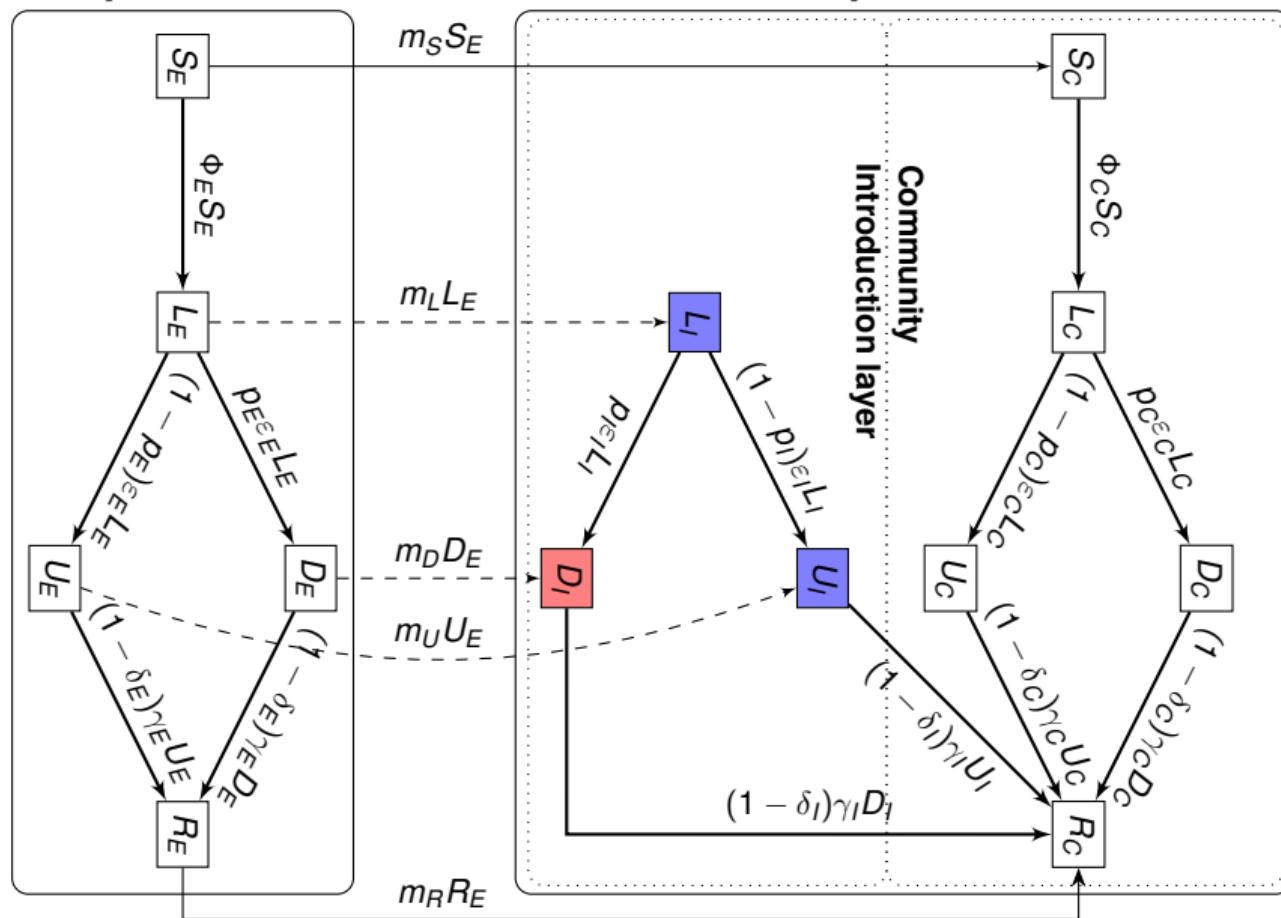
Exemple – Vérification des paramètres

Exemple – Phase stochastique d'une épidémie

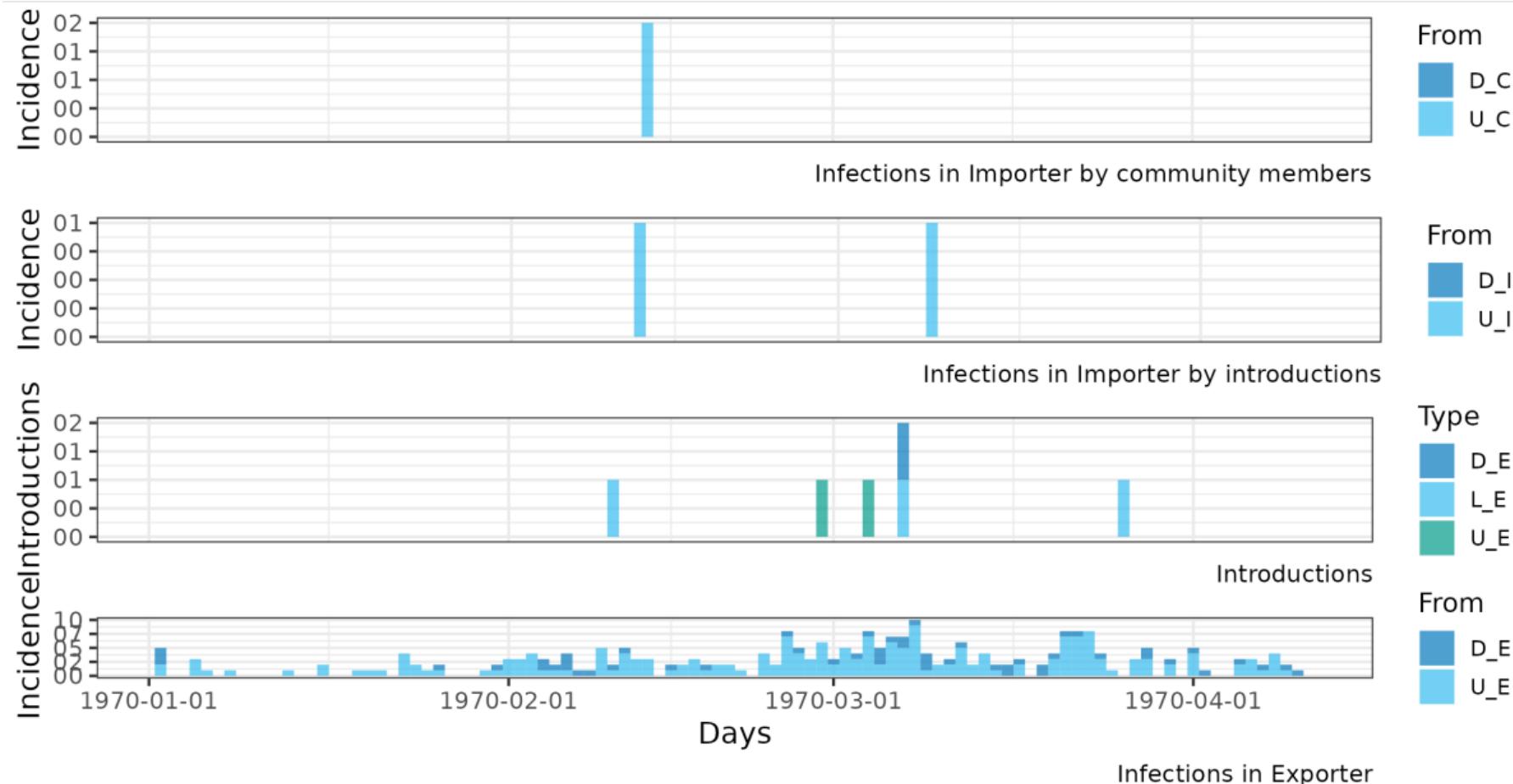
Valeur des mesures de contrôle des voyages

Exporter

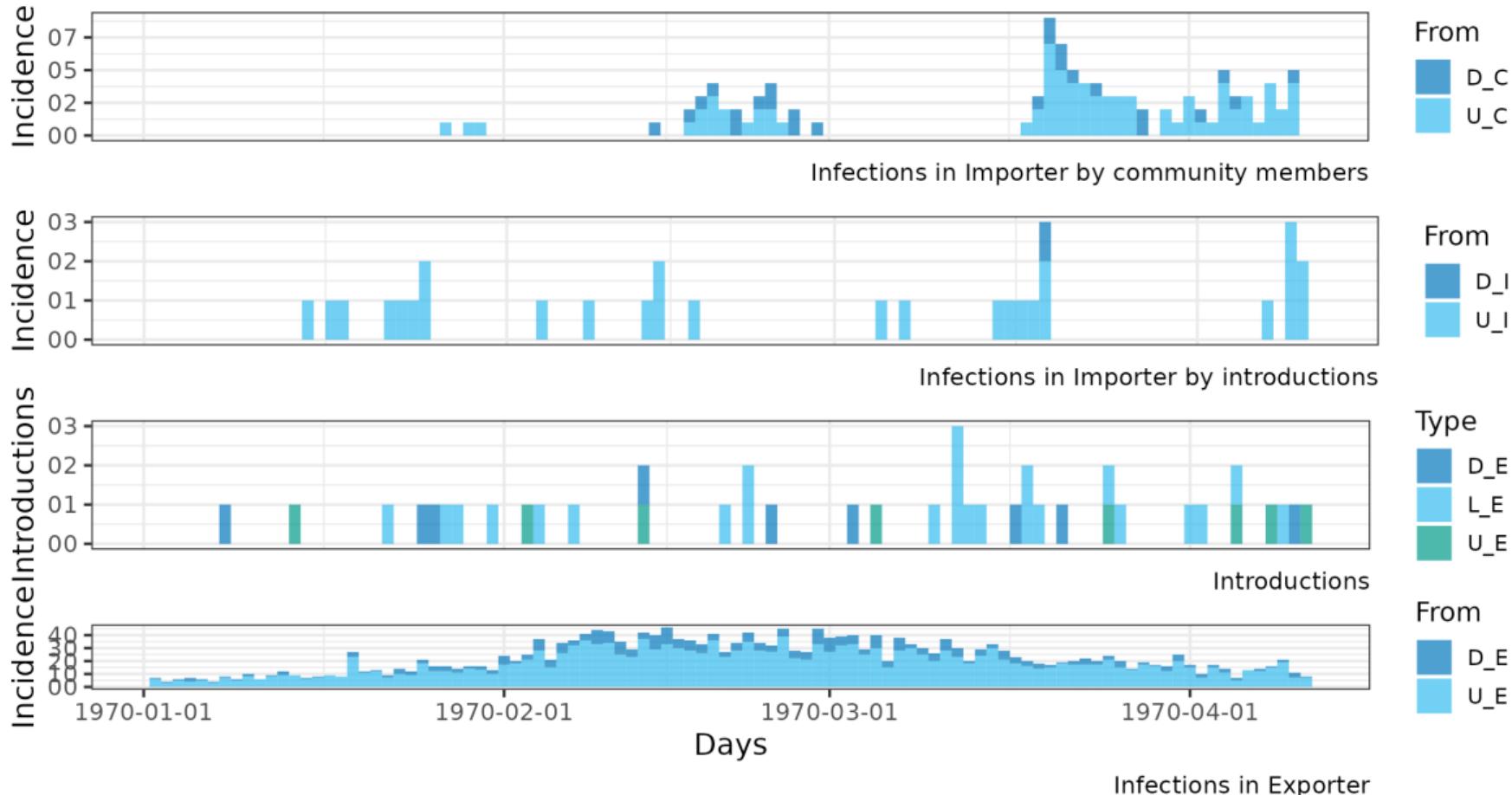
Importer



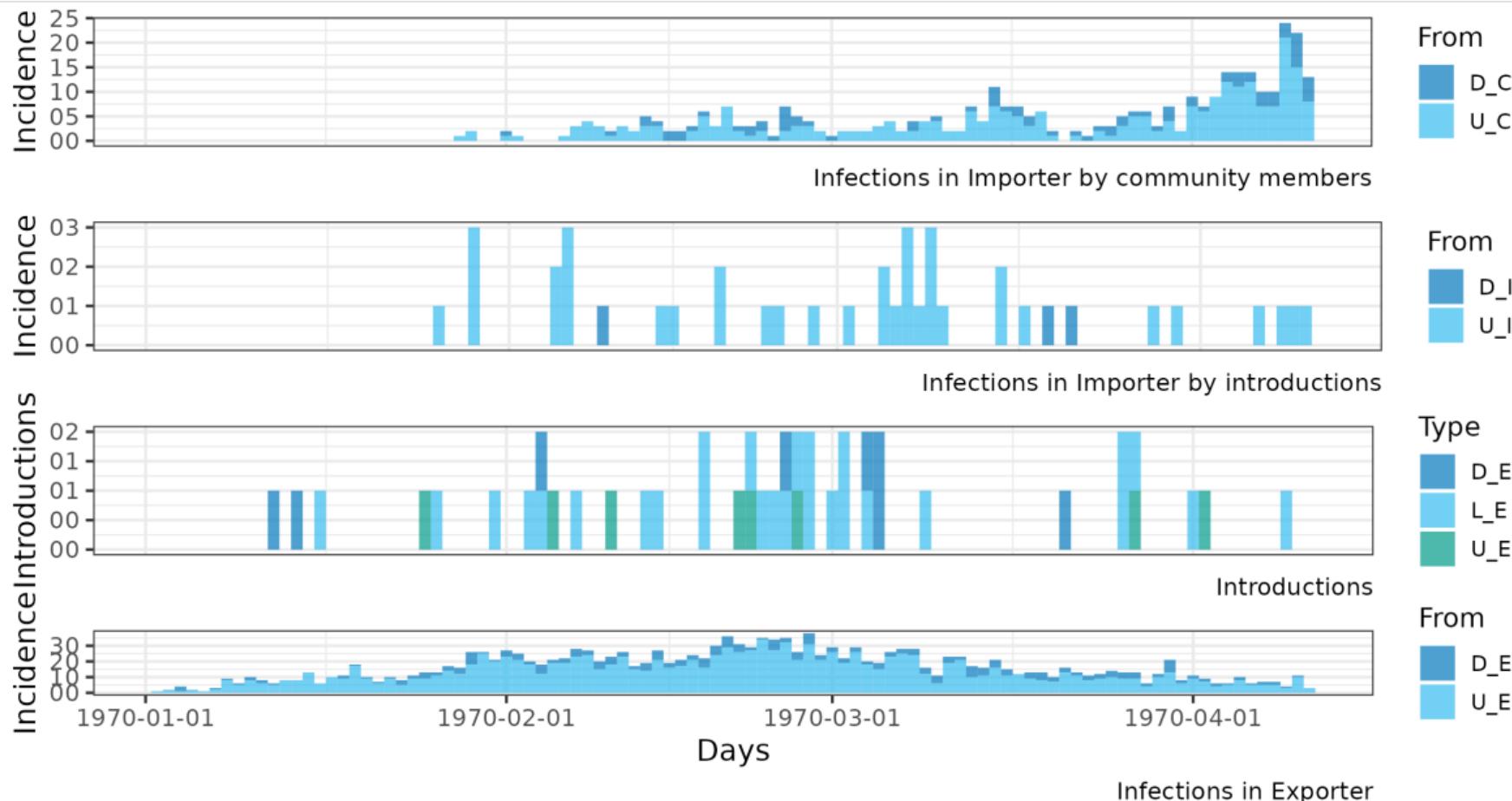
$$R_0^E = 1.5, R_0^C = 0.8, \text{pop}_E = 10000, \text{pop}_I = 10000$$



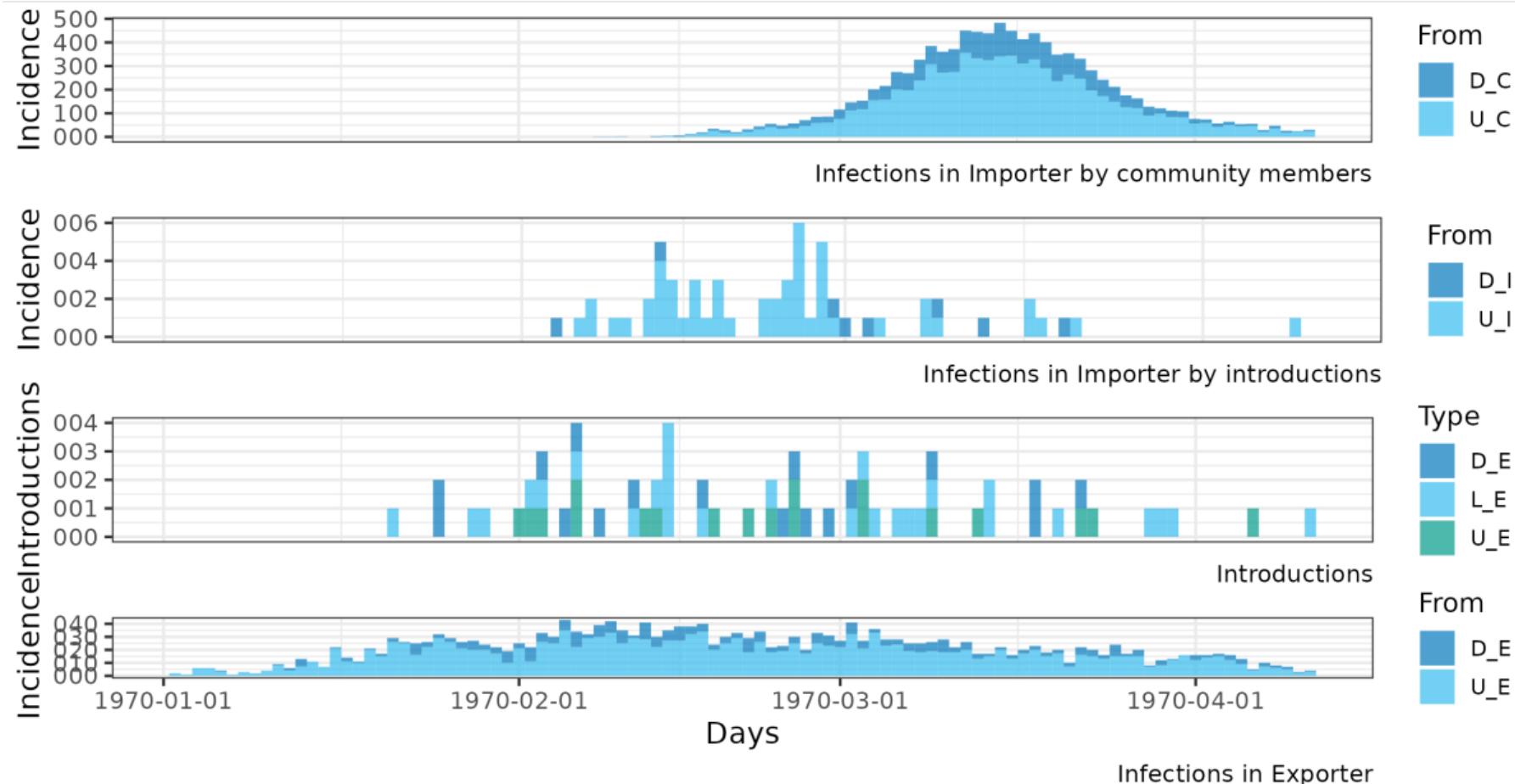
$$R_0^E = 1.5, R_0^C = 0.8, \text{pop}_E = 10000, \text{pop}_I = 10000$$

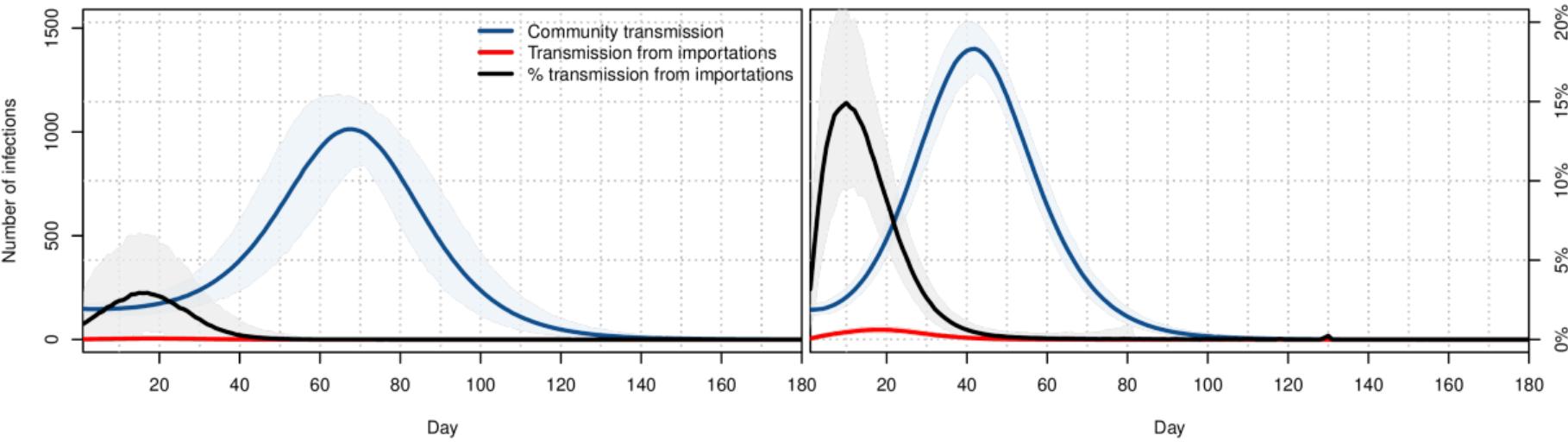


$$R_0^E = 1.5, R_0^C = 0.8, \text{pop}_E = 10000, \text{pop}_I = 10000$$



$$R_0^E = 1.5, R_0^C = 1.5, \text{pop}_E = 10000, \text{pop}_I = 10000$$





Une dernière pensée pour la route

V. Chetail. Crisis without borders: What does international law say about border closure in the context of Covid-19? *Frontiers in Political Science*, 2 (12) (2020)

[...] a powerful expression of state's sovereignty, immigration control provides a typical avenue for governments to reassure their citizens and bolster a national sense of belonging, while providing an ideal scapegoat for their own failure or negligence.



Pourquoi incorporer la stochasticité ?

Stochasticité dans les modèles déterministes

Chaînes de Markov en temps continu

Approximations des CMTC par processus de
branchement

Qu'est-ce qu'un processus de branchement ?

Un processus de branchement est un modèle mathématique pour une population où les individus produisent un nombre aléatoire de descendants puis meurent

Pensez aux bactéries qui se divisent, à un virus qui se propage, ou même à la survie des noms de famille

- ▶ Commençons avec une population initiale, Z_0
- ▶ Chaque individu de la génération n produit un nombre de descendants pour la génération $n + 1$
- ▶ Ce « nombre de descendants » est une variable aléatoire. Tous les individus produisent des descendants selon la même distribution de probabilité, indépendamment les uns des autres

Le processus de Galton-Watson

Soit Z_n la taille de la population à la génération n . Nous commençons typiquement avec $Z_0 = 1$

La population évolue selon la règle

$$Z_{n+1} = \sum_{i=1}^{Z_n} X_{n,i}$$

- ▶ $X_{n,i}$: nombre de descendants produits par le i -ème individu de la génération n
- ▶ Les variables $\{X_{n,i}\}$ sont des variables aléatoires **indépendantes et identiquement distribuées (i.i.d.)** à valeurs entières
- ▶ Nous appelons leur distribution commune $\{p_k\}_{k=0}^{\infty}$ la **distribution des descendants**, où $p_k = P(X = k)$

Questions fondamentales

1. Taille espérée à long terme de la population ?
2. Probabilité que la population finisse par s'éteindre ?

Moyenne de la distribution des descendants

Le destin de la population dépend d'un seul paramètre : la moyenne de la distribution des descendants

$$\mu = E[X] = \sum_{k=0}^{\infty} k \cdot p_k$$

Taille de population espérée

En utilisant la loi de l'espérance totale, nous trouvons la taille espérée de la prochaine génération

$$E[Z_{n+1}|Z_n] = E\left[\sum_{i=1}^{Z_n} X_{n,i} \middle| Z_n\right] = Z_n E[X] = Z_n \mu$$

En prenant l'espérance à nouveau, nous obtenons une récurrence simple

$$E[Z_{n+1}] = \mu E[Z_n]$$

Ceci implique

$$E[Z_n] = Z_0 \mu^n$$

Les trois régimes de croissance de population

Le comportement de $E[Z_n] = Z_0\mu^n$ suggère trois cas distincts

Sous-critique ($\mu < 1$)

$E[Z_n] \rightarrow 0$. La population devrait rétrécir. Elle s'éteint avec probabilité 1

Critique ($\mu = 1$)

$E[Z_n] = Z_0$. La population devrait rester stable. Curieusement, elle s'éteint quand même avec probabilité 1

Super-critique ($\mu > 1$)

$E[Z_n] \rightarrow \infty$. La population devrait croître exponentiellement. Elle a une probabilité non nulle de survivre à jamais

Fonction génératrice de probabilité

Pour trouver la probabilité d'extinction, nous utilisons la **fonction génératrice de probabilité** (FGP) de la distribution des descendants X .

$$G(s) = E[s^X] = \sum_{k=0}^{\infty} p_k s^k \quad \text{pour } |s| \leq 1$$

Propriétés clés

- ▶ $G(1) = \sum p_k = 1$
- ▶ La moyenne peut être trouvée à partir de la dérivée : $G'(1) = \sum kp_k = \mu$
- ▶ La FGP de Z_n est le n -ième itéré de $G(s)$ avec elle-même. Si $G_n(s)$ est la FGP de Z_n , alors $G_{n+1}(s) = G(G_n(s))$

Équation de la probabilité d'extinction

Soit π_0 la probabilité d'extinction éventuelle, en commençant avec $Z_0 = 1$

$$\pi_0 = P(\text{la population s'éteint}) = \lim_{n \rightarrow \infty} P(Z_n = 0)$$

Puisque $P(Z_n = 0) = G_n(0)$, et $G_{n+1}(0) = G(G_n(0))$, à la limite la probabilité d'extinction π_0 doit satisfaire l'équation

$$\pi_0 = G(\pi_0)$$

Theorem 3

*La probabilité d'extinction π_0 est la **plus petite solution non négative** de l'équation $s = G(s)$*

- ▶ Si $\mu \leq 1$, la seule solution dans $[0, 1]$ est $s = 1$. Donc $\pi_0 = 1$
- ▶ Si $\mu > 1$, il existe une solution unique dans $[0, 1)$, qui est la probabilité d'extinction $\pi_0 < 1$

Du temps discret au temps continu

Limitation de Galton-Watson

Les générations n'arrivent pas en étapes synchronisées dans le monde réel. Les individus donnent naissance et meurent à des moments aléatoires

Cela nous amène aux **Chaînes de Markov en temps continu (CMTC)**

- ▶ L'état du système est la taille de la population, $k \in \{0, 1, 2, \dots\}$
- ▶ Au lieu de générations, nous avons des taux de transition :
 - ▶ λ_k : taux de naissance lorsque la population est de taille k (passe à $k + 1$)
 - ▶ δ_k : taux de décès lorsque la population est de taille k (passe à $k - 1$).
- ▶ Souvent, nous supposons que ces taux sont linéaires : $\lambda_k = k\lambda$ et $\delta_k = k\delta$. Cela signifie que les individus agissent indépendamment

Approximation d'une CMTC par processus de branchement

L'idée clé

Au début d'une épidémie (ou pour une très grande population), la dynamique causée par un seul individu est largement indépendante des autres

Cela nous permet d'approximer le début d'un processus de population CMTC avec un processus de branchement

Exemple : Une épidémie simple (modèle SIR)

- ▶ S : Susceptible, I : Infecté, R : Rétabli
- ▶ Une personne infectée rencontre d'autres personnes à un certain taux. Si elle rencontre un susceptible, une nouvelle infection peut se produire (un « descendant »)
- ▶ La personne infectée se rétablit (ou meurt) à un autre taux, terminant sa période infectieuse
- ▶ **Question :** Combien de nouvelles infections une seule personne infectée cause-t-elle en moyenne ?

Étude de cas : Le nombre de reproduction de base \mathcal{R}_0

Considérons un seul individu infecté dans une grande population de susceptibles

- ▶ Soit β le taux d'infection (taux de production de « descendants »)
- ▶ Soit γ le taux de rétablissement (taux de « mort »)

La durée de vie infectieuse de l'individu est une variable aléatoire exponentielle avec moyenne $1/\gamma$

Le nombre moyen d'infections secondaires qu'il cause est

$$\mathcal{R}_0 = (\text{taux d'infection}) \times (\text{période infectieuse moyenne}) = \beta \times \frac{1}{\gamma} = \frac{\beta}{\gamma}$$

La connexion

\mathcal{R}_0 est précisément le **nombre moyen de descendants** μ pour le processus de branchement intégré qui approxime le début de l'épidémie

Application aux épidémies

Le destin de la phase initiale de l'épidémie est déterminé par \mathcal{R}_0

- ▶ Si $\mathcal{R}_0 \leq 1$ ($\mu \leq 1$) : Le nombre d'individus infectés est un processus sous-critique ou critique. L'épidémie s'éteint avec probabilité 1
- ▶ Si $\mathcal{R}_0 > 1$ ($\mu > 1$) : Le processus est super-critique. Il y a une probabilité positive que l'épidémie décolle et cause une épidémie majeure

Nous pouvons même calculer la probabilité d'une épidémie majeure ! C'est $1 - \pi_0$, où π_0 est la probabilité d'extinction

Pour ce processus d'infection naissance-mort simple, la FGP est

$G(s) = \frac{\gamma}{\beta+\gamma} + \frac{\beta}{\beta+\gamma}s$. Résoudre $s = G(s)$ donne la probabilité d'extinction

$$\pi_0 = \frac{\gamma}{\beta} = \frac{1}{\mathcal{R}_0}$$

La probabilité d'une épidémie majeure est $1 - 1/\mathcal{R}_0$

Résumé pour APB

- ▶ Les **processus de branchement** modélisent les populations avec une génération de descendants i.i.d.
- ▶ Le destin de la population est déterminé par le **nombre moyen de descendants** μ . L'extinction est certaine si $\mu \leq 1$
- ▶ La **probabilité d'extinction** π_0 peut être calculée comme le plus petit point fixe non négatif de la fonction génératrice de probabilité $G(s)$
- ▶ Les stades initiaux de nombreuses **chaînes de Markov en temps continu** à grande échelle peuvent être approximés par un processus de branchement
- ▶ Cela nous permet d'appliquer la théorie à des problèmes du monde réel, comme le calcul du **nombre de reproduction de base** \mathcal{R}_0 d'une épidémie et sa probabilité de causer une épidémie majeure

Bibliography I

-  Arino, J., Cooke, K., van den Driessche, P., and Velasco-Hernández, J. (2004). An epidemiology model that includes a leaky vaccine with a general waning function. *DCDS-B*, 4(2) :479–495.
-  Arino, J. and Milliken, E. M. (2022). Effect of mobility on the stochastic phase of an epidemic. *Bulletin of Mathematical Biology*, 84.
-  Chetail, V. (2020). Crisis without borders : What does international law say about border closure in the context of Covid-19 ? *Frontiers in Political Science*, 2 :606307.
-  Whittle, P. (1955). The outcome of a stochastic epidemic – A note on Bailey's paper. *Biometrika*, 42(1-2) :116–122.