

and we dedicate ourselves to move forward in partnership with Indigenous communities in a spirit of Reconciliation and collaboration.

Network models

MATH 8xyz – Lecture 25

Julien Arino
Department of Mathematics @ University of Manitoba
Maud Menten Institute @ PIMS
julien.arino@umanitoba.ca

Winter 20XX

Outline



At the beginning of the COVID-19 crisis

- ▶ I was working under contract with the Public Health Agency of Canada on *COVID-19 importation risk assessment*
- ▶ Produced daily report with list of countries most likely to next report cases of COVID-19
- ▶ Used ensemble runs of a fitted global deterministic metapopulation model



- ▶ Very very long days (18-20 hours, 7 days a week)
 - ▶ including a lot of time waiting for the “cluster” to finish
- ⇒ PHAC gave me money for a cluster (yay Threadrippers!!!)
- ⇒ Also thought about whether my model was really adequate as our focus switched from thinking about movement on a planetary scale to movement within Canadian provinces

What is wrong with deterministic models?

- ▶ I pointed out yesterday that SARS-CoV-2 is one *single* realisation of a stochastic process
- ▶ Deterministic models “operate on averages” over a large ($\rightarrow \infty$) number of realisations
- ▶ If we want to get a better sense of what could happen, not only on average, then we need to see what can indeed happen

My new focus – Introductions

- ▶ I started thinking in particular about **introductions** (or importations) of pathogens into new populations
- ▶ Indeed, introductions are an obligatory step in spatial spread

First piece of evidence

In real life, introductions of pathogens does not always follow the pattern

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

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.

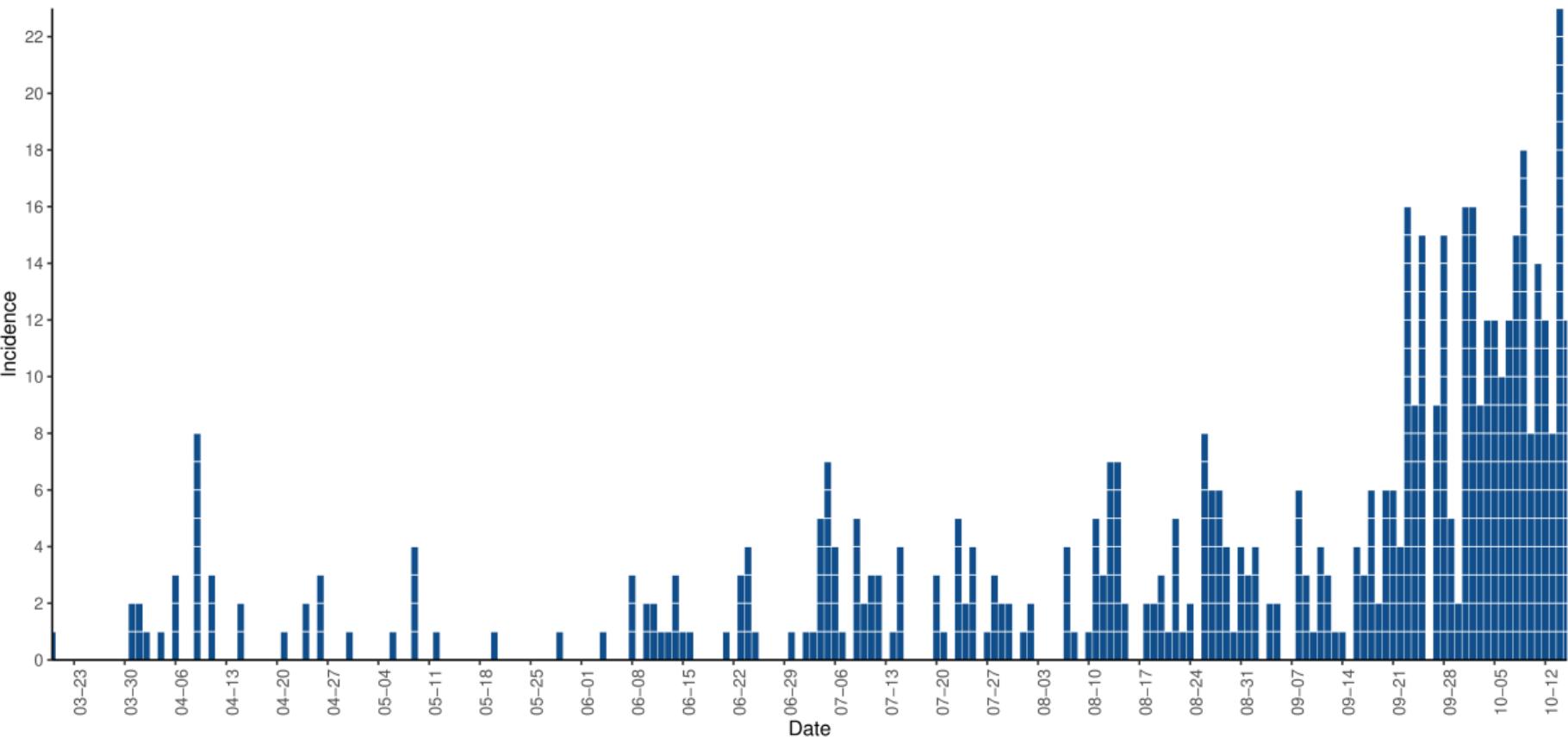
Table 1. Effect of vaccination scaling-up on the probability of successful viral introduction.

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.

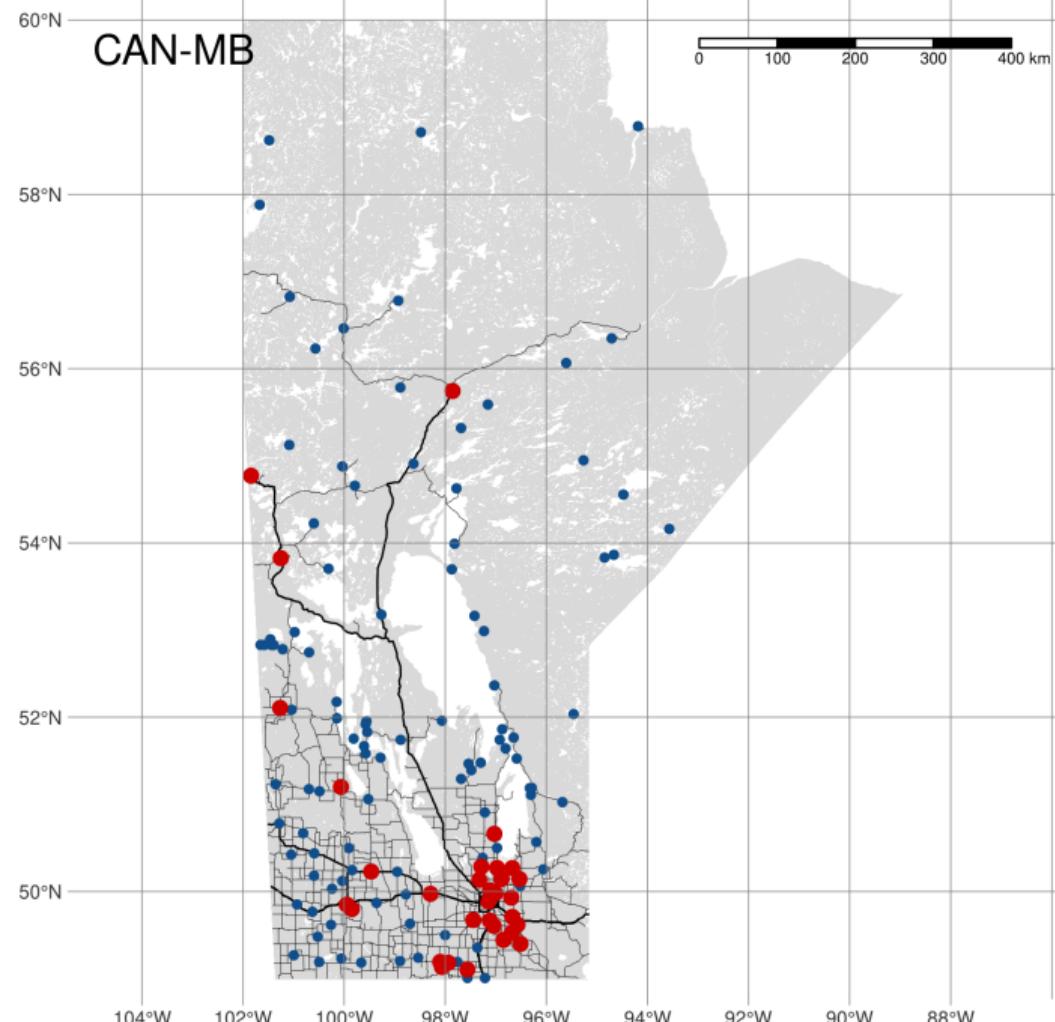
Second piece of evidence

The start of an outbreak can be extremely slow, with very few cases for quite a while



Why this is relevant

Far from the only reason, but as an example: Canada has remote/isolated communities that are vulnerable to introductions of pathogens



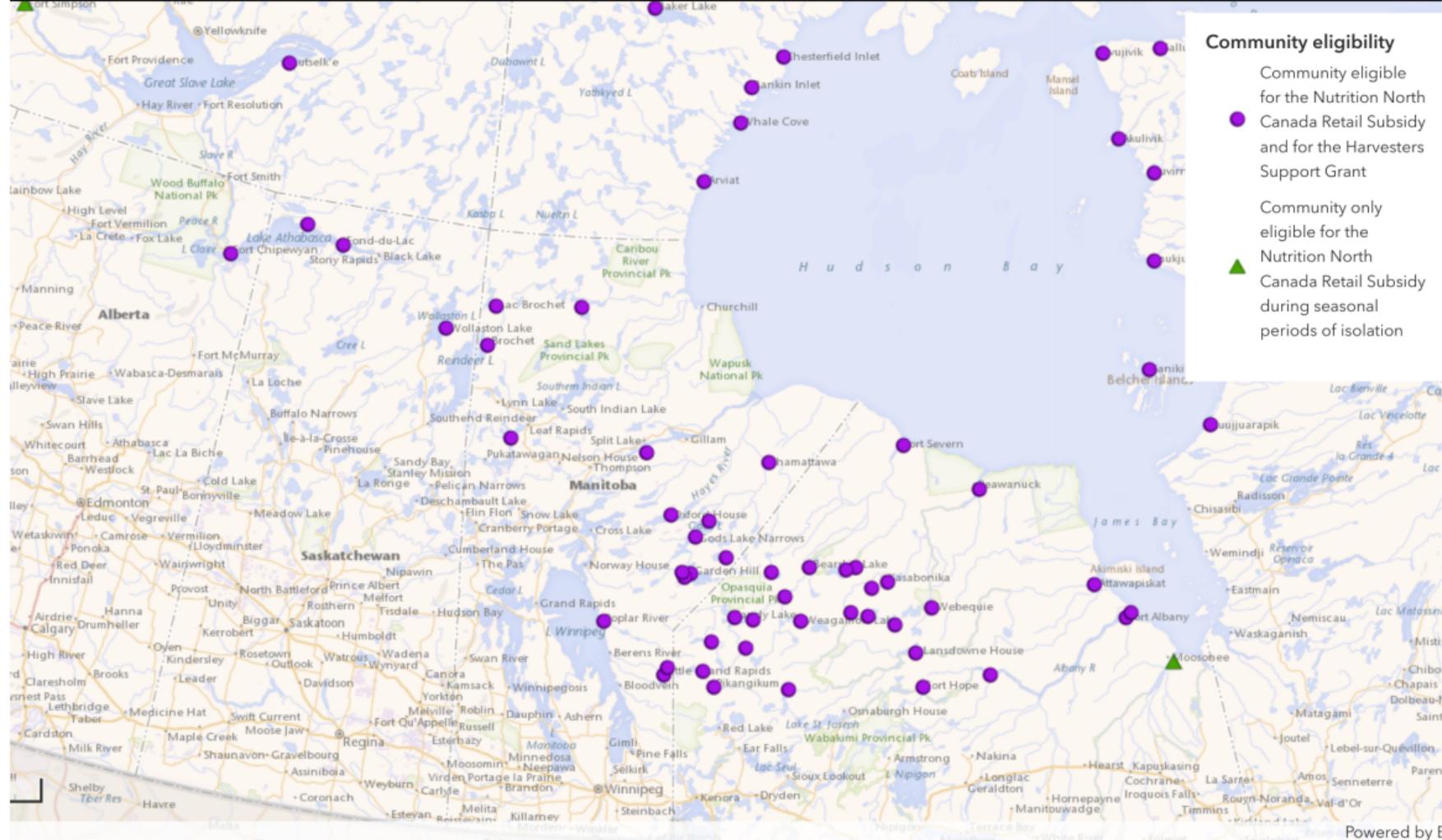
Community eligibility

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.

For First Nation and Métis Communities

Remote describes a geographical area where a community is located over 350 km from the nearest service centre having year-round access by land and/or water routes normally used in all weather conditions

Isolated means a geographical area that has scheduled flights and good telephone service, but is without year-round access by land and/or water normally used in all weather conditions

Remote-Isolated means a geographic area that has neither scheduled flights nor year-round access by land and/or water routes normally that can be used in all weather conditions, irrespective of the level of telephone and radio service available

For Inuit communities

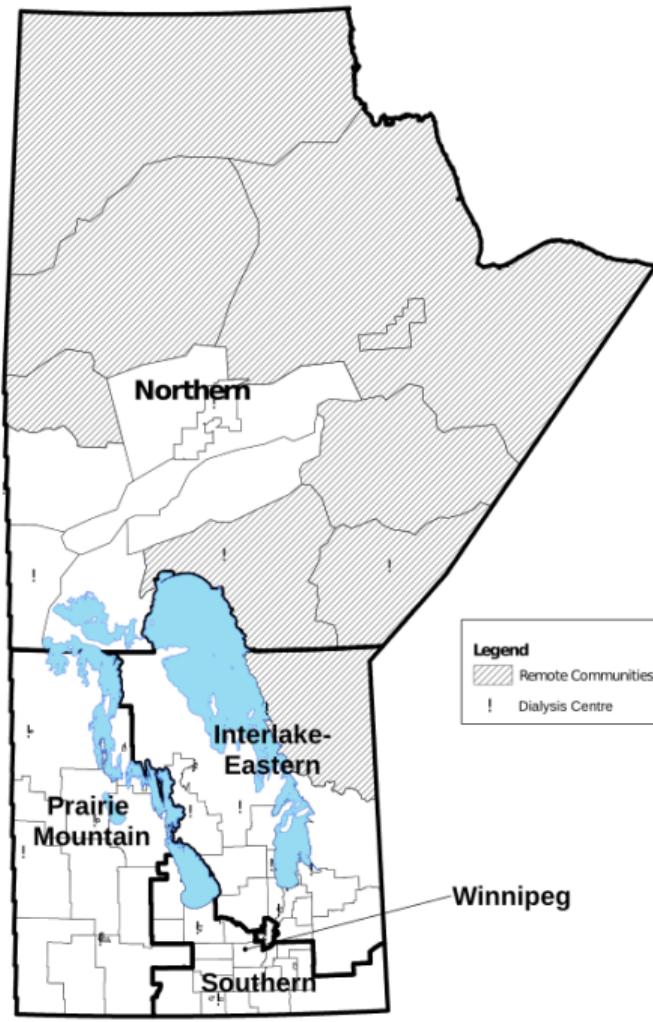
Inuit Communities to be referred to as **Inuit Nunangat**, not remote and isolated communities to respect the unique language and culture of Inuit regions, as well as the common challenges in social determinants of health, access to care, and infrastructure found across all Inuit communities

MB remote communities

Remote communities are communities in Manitoba that **do not have permanent road access** (i.e., no all-weather road), are **more than a four-hour drive from a major rural hospital (and a dialysis unit)**, or **have rail or fly-in access only**. This includes Norway House, Lynn Lake, Leaf Rapids, Gillam, and Cross Lake. If most communities in a health district are designated as "remote", the entire district is designated as "remote". In Manitoba, remote districts include:

- ▶ Northern Health Region: NO23, NO13, NO25, NO16, NO22, NO26, NO28, NO31, and
- ▶ Interlake-Eastern Health Region: 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



Travel to/from remote or isolated communities

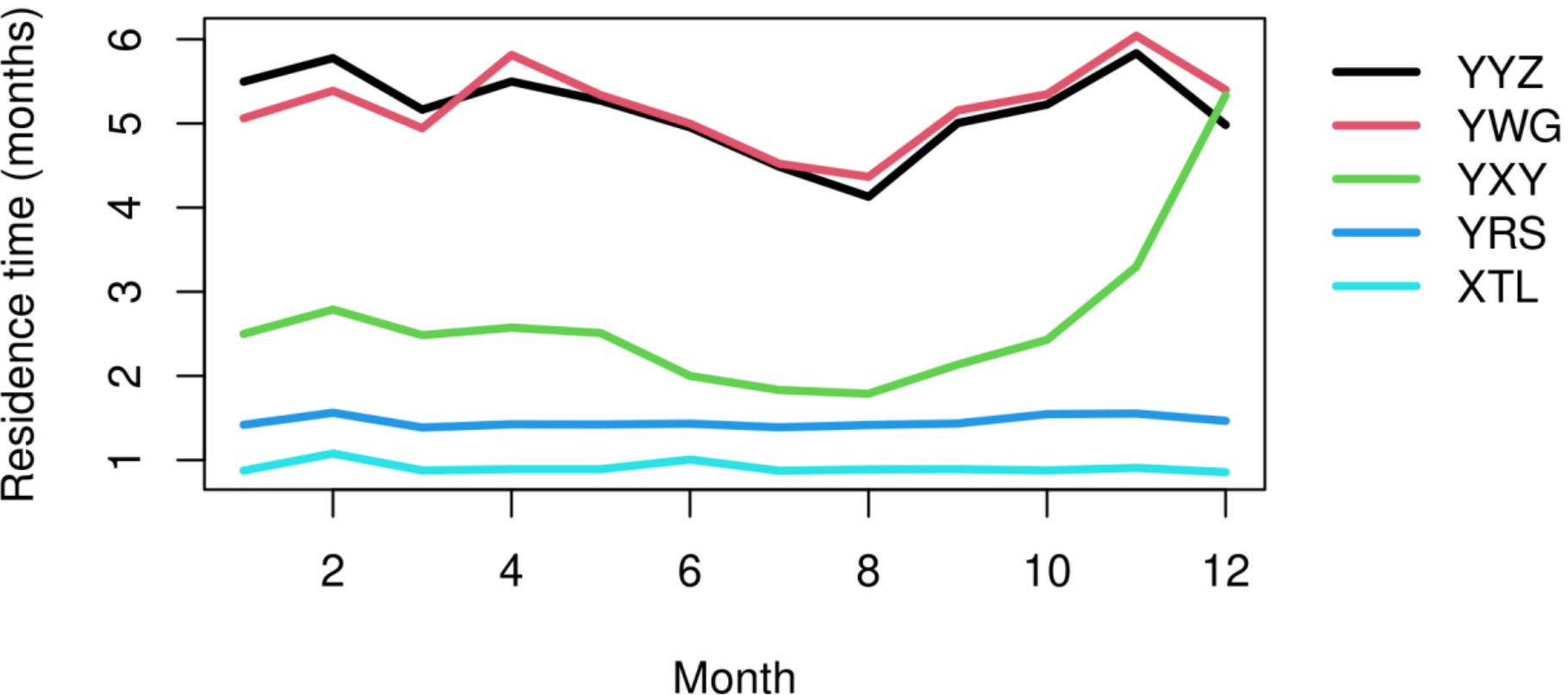
How do you think this compares to travel in non-remote/isolated communities ?

Residence time (the lake ecology version): theoretic time an average water or comparable molecule spends in a lake, considering inflow into and outflow from the lake

Think of residence times in these communities: what is the average time a person spends in a remote or isolated community before leaving it?

The **residence time in a location** is the total number of trips inbound into and outbound from location over a duration of time (1 month here) divided by the normal population in the location

Residence times in months



The paradox of travel to/from remote/isolated communities

Travel volumes small but movement rates high

ICs are highly connected to the urban centre(s) they are subordinated to

Further reinforced in Winnipeg by urban indigenous population (102,075 or 12.45% of metro population), meaning many family connections exist





See in particular the work of Horst Thieme

If one considers time of sojourn in compartments from a more detailed perspective, one obtains integro-differential models

We use here continuous random variables. See chapters 12 and 13 in Thieme's book for arbitrary distributions

Time to events

We suppose that a system can be in two states, A and B

- ▶ At time $t = 0$, the system is in state A
- ▶ An event happens at some time $t = \tau$, which triggers the switch from state A to state B

Let us call T the random variable

“time spent in state A before switching into state B”

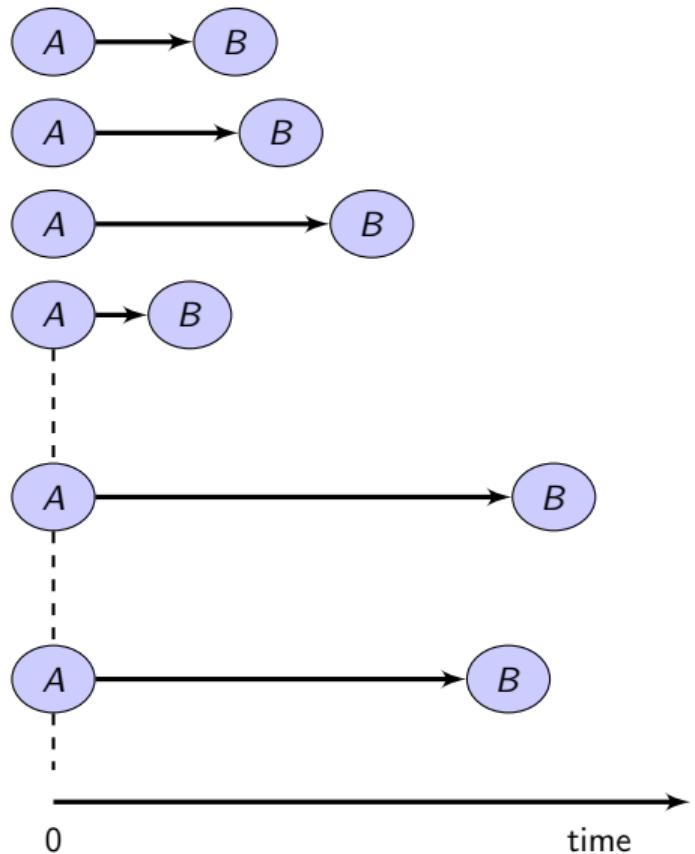
The states can be anything:

- ▶ A : working, B : broken
- ▶ A : infected, B : recovered
- ▶ A : alive, B : dead
- ▶ ...

We take a collection of objects or individuals that are in state A and want some law for the **distribution** of the times spent in A , i.e., a law for T

For example, we make light bulbs and would like to tell our customers that on average, our light bulbs last 200 years...

We conduct an **infinite** number of experiments, and observe the time that it takes, in every experiment, to switch from A to B



A distribution of probability is a model

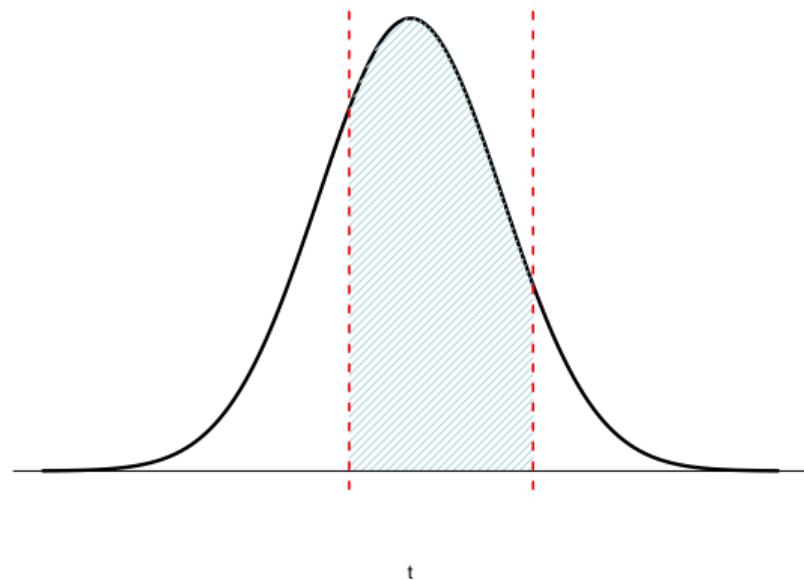
From the sequence of experiments, we deduce a model, which in this context is called a **probability distribution**

We assume that T is a **continuous** random variable

Probability density function

Since T is continuous, it has a continuous **probability density function** f

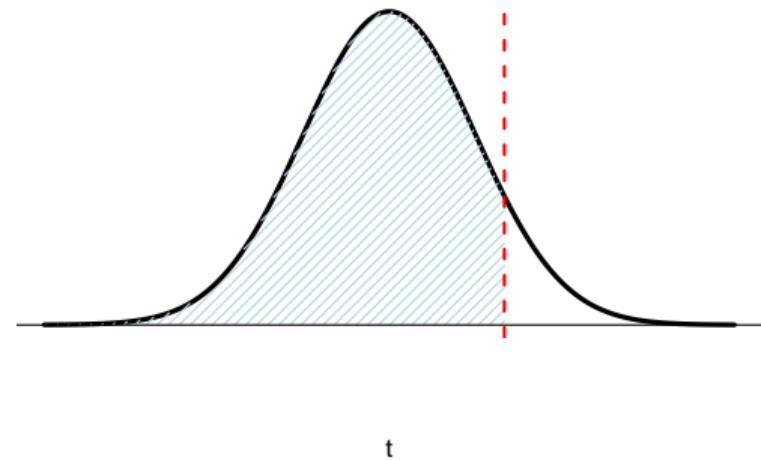
- ▶ $f \geq 0$
- ▶ $\int_{-\infty}^{+\infty} f(s)ds = 1$
- ▶ $\mathbb{P}(a \leq T \leq b) = \int_a^b f(t)dt$



Cumulative distribution function (c.d.f.)

The **cumulative distribution function** is a function $F(t)$ that characterizes the distribution of T , and defined by

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



Survival function

Another characterization of the distribution of the random variable T is through the **survival** (or **sojourn**) function

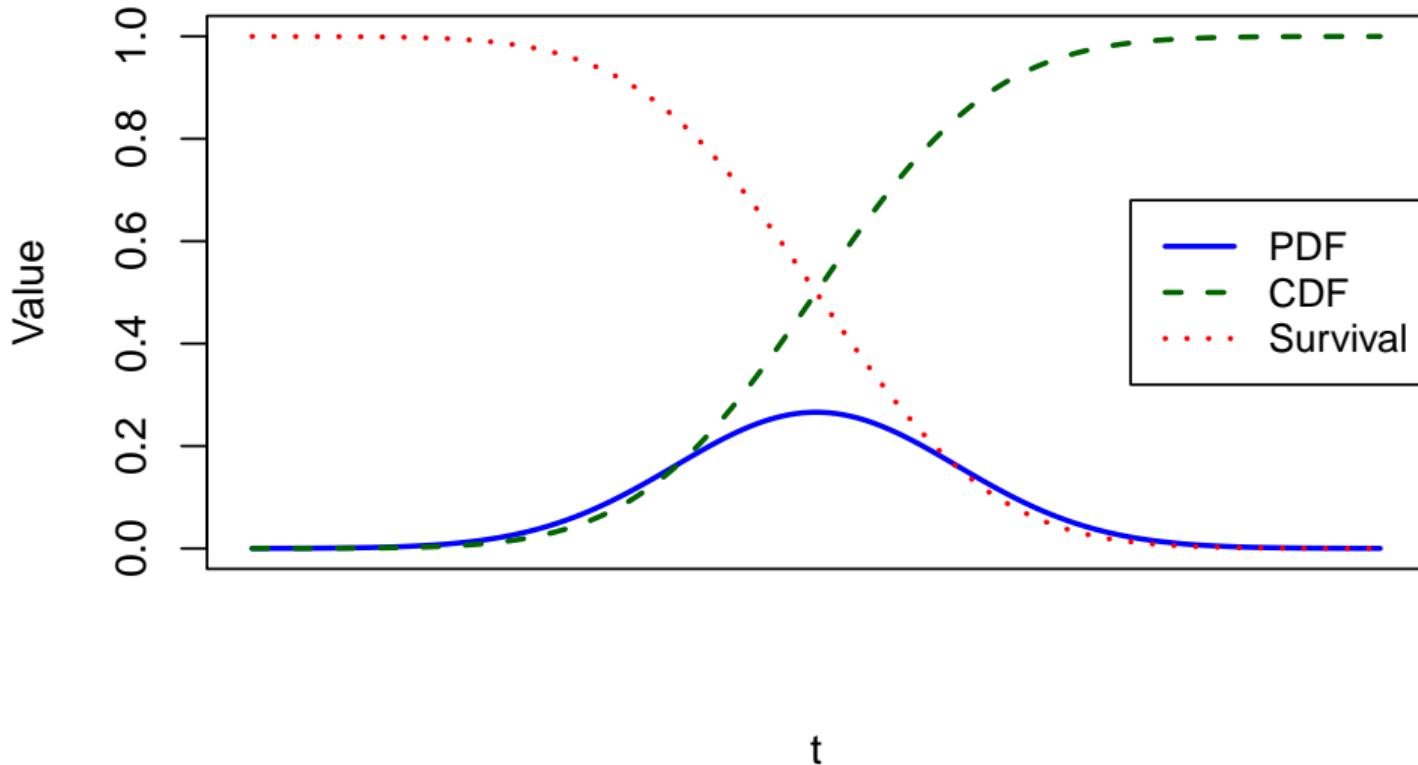
The survival function of state A is given by

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

This gives a description of the **sojourn time** of a system in a particular state (the time spent in the state)

S is a nonincreasing function (since $S = 1 - F$ with F a c.d.f.), and $S(0) = 1$ (since T is a nonnegative random variable)

PD, CD and Survival functions



The **average sojourn time** τ in state A is given by

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

Since $\lim_{t \rightarrow \infty} t S(t) = 0$, it follows that

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

Expected future lifetime:

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

$$\begin{aligned} S(t) - S(a) &= \mathbb{P}\{\text{survive during } (a, t) \text{ having survived until } a\} \\ &= \exp\left(-\int_a^t h(u) du\right) \end{aligned}$$

Hazard rate

The **hazard rate** (or **failure rate**) is

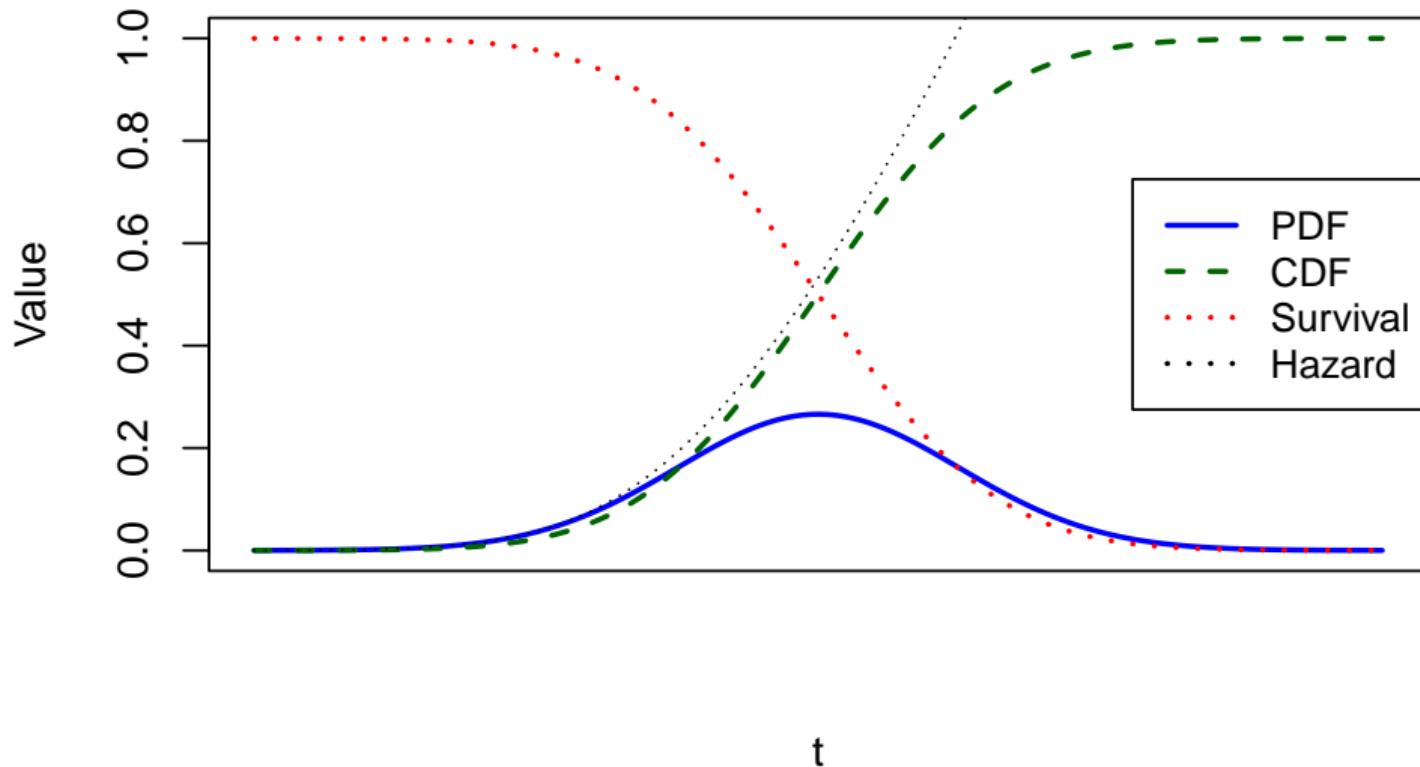
$$\begin{aligned} h(t) &= \lim_{\Delta t \rightarrow 0} \frac{\mathcal{S}(t) - \mathcal{S}(t + \Delta t)}{\Delta t} \\ &= \lim_{\Delta t \rightarrow 0} \frac{\mathbb{P}(T < t + \Delta t | T \geq t)}{\Delta t} \\ &= \frac{f(t)}{\mathcal{S}(t)} \end{aligned}$$

It gives probability of failure between t and Δt , given survival to t .

We have

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

PD, CD and Survival functions & Hazard rate



Competing risks

Suppose now that the system starts in state A at time $t = 0$ and that depending on which of the two events \mathcal{E}_1 or \mathcal{E}_2 takes place first, it switches to state B_1 or B_2 , respectively

Consider the random variables T_A , *time spent in state A* (or sojourn time in A), T_{AB_1} , *time before switch to B_1* and T_{AB_2} , *time before switch to B_2*

If we consider state A , we cannot observe the variables T_{AB_1} or T_{AB_2} . What is observable is the sojourn time in A

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

(where * indicates that a quantity is observable)

Failure rate by type of event

We have two (or more) types of events whose individual failure rates have to be accounted for

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

where $\mathbb{P}(T < t + \Delta t, S = S_j | T \geq t)$ is the probability of failure due to cause S_j ($j = 1, 2$ ici), i.e., S is a discrete r.v. representing the event that is taking place

By the law of total probability, since only one of the event can take place, if there are n risks, then

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

or, identically,

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

As a consequence, suppose a process is subject to two competing exponential risks with respective distributions with parameters θ_1 and θ_2

Then the mean sojourn time in the initial state before being affected by one of the two risks is

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



The exponential distribution

The random variable T has an **exponential** distribution if its probability density function takes the form

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

with $\theta > 0$. Then the survival function for state A is of the form $S(t) = e^{-\theta t}$, for $t \geq 0$, and the average sojourn time in state A is

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

Particularities of the exponential distribution

The standard deviation of an exponential distribution is also $1/\theta$. When estimating θ , it is impossible to distinguish the mean and the standard deviation

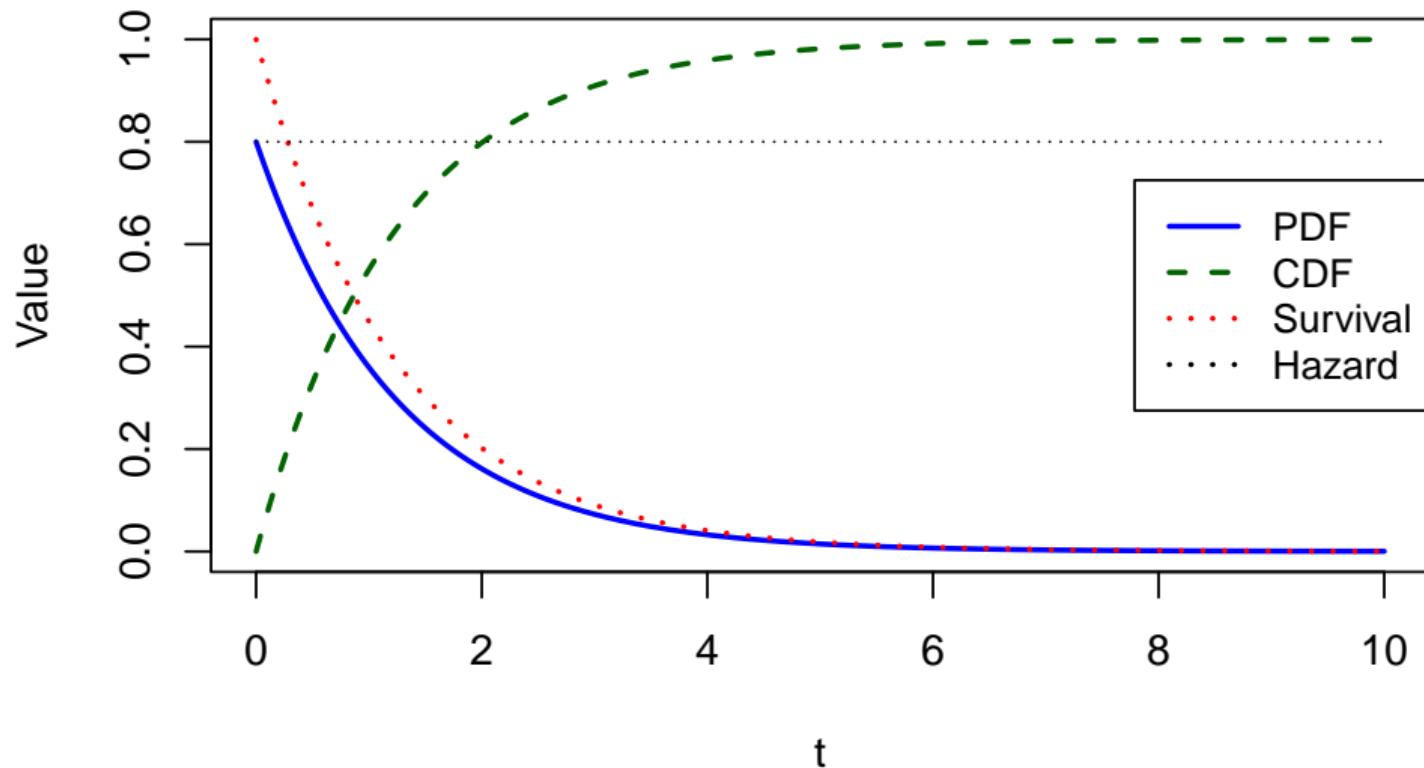
The exponential distribution is **memoryless**: its conditional probability obeys

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

The exponential and geometric distributions are the only memoryless probability distributions

The exponential distribution has a constant hazard function $h(t) \equiv \theta$

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



The Dirac delta distribution

If for some constant $\omega > 0$,

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

meaning that T has a Dirac delta distribution $\delta_\omega(t)$, then the average sojourn time is

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

with standard deviation $\sigma = 0$

The Gamma distribution

R.v. X is **Gamma** distributed ($X \sim \Gamma(k, \theta)$) with **shape parameter** k and **scale parameter** θ (or **rate** $\beta = 1/\theta$) (all positive) if its probability density function takes the form

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

where $x > 0$ and Γ is the Euler Gamma function, defined for all $z \in \mathbb{C}$ s.t. $\operatorname{Re}(z) > 0$ by

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

Properties of the Gamma distribution

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

Survival function

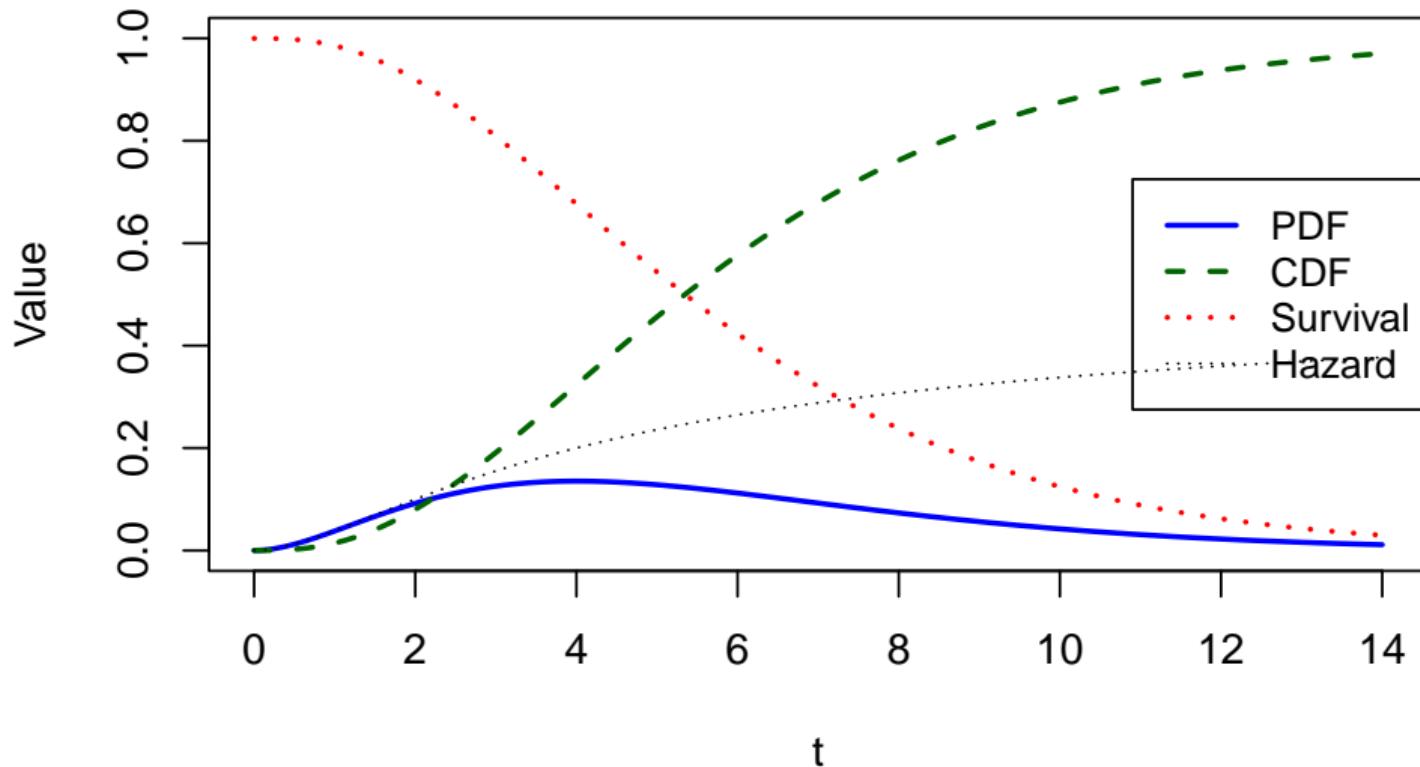
$$S(t) = 1 - \frac{1}{\Gamma(k)} \gamma\left(k, \frac{t}{\theta}\right) = 1 - \frac{1}{\Gamma(k)} \gamma(k, \beta t)$$

where

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

is an incomplete Gamma function

PDF, CDF, Survival & Hazard of Gamma Distribution





A model for a cohort with one cause of death

Consider a **cohort** of individuals born at the same time, e.g., the same year

- ▶ At time $t = 0$, there are initially $N_0 > 0$ individuals
- ▶ All causes of death are compounded together
- ▶ The time until death, for a given individual, is a random variable T , with continuous probability density distribution $f(t)$ and survival function $S(t)$

$N(t)$ the cohort population at time $t \geq 0$

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

$S(t)$ proportion of initial population still alive at time t , so $N_0 S(t)$ number in the cohort still alive at time t

Case where T is exponentially distributed

Suppose that T has an exponential distribution with mean $1/d$ (or parameter d), $f(t) = de^{-dt}$. Then the survival function is $S(t) = e^{-dt}$, and (??) takes the form

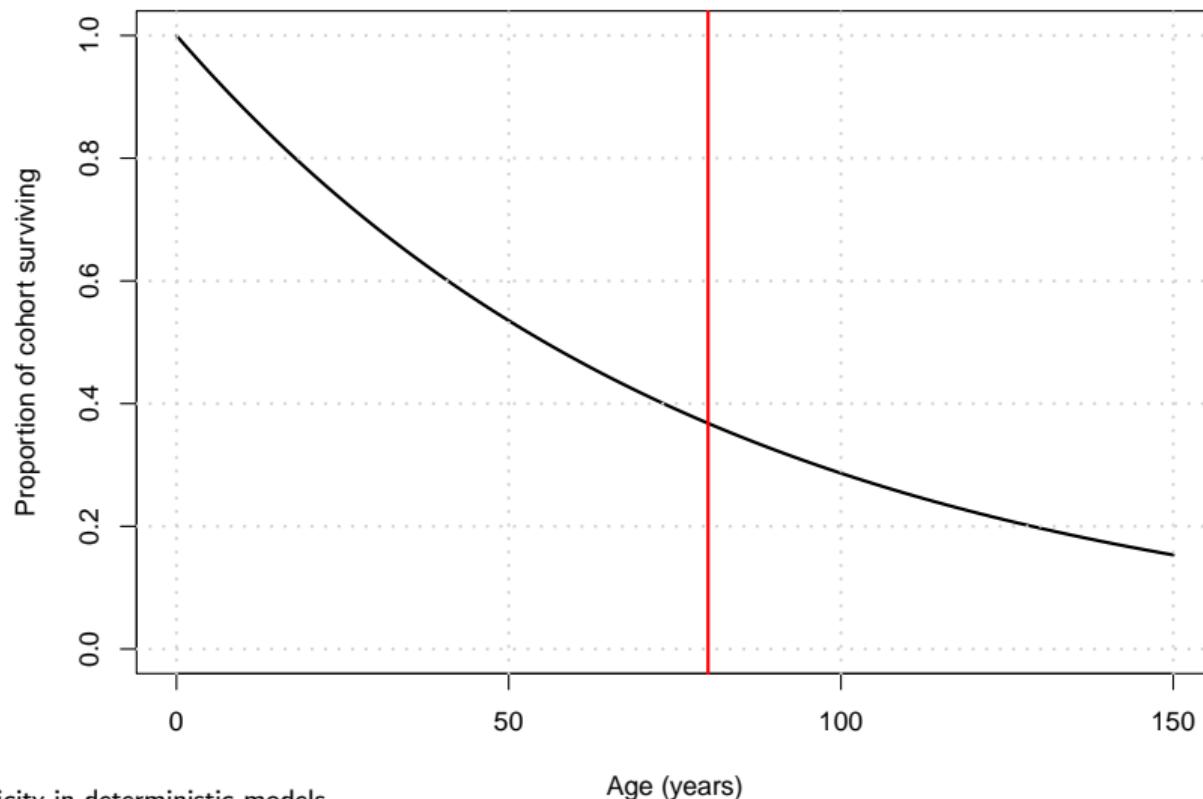
$$N(t) = N_0 e^{-dt} \tag{5}$$

Now note that

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

⇒ The ODE $N' = -dN$ makes the assumption that the life expectancy at birth is exponentially distributed

Survival function, $\mathcal{S}(t) = \mathbb{P}(T > t)$, for an exponential distribution with mean 80 years



Case where T has a Dirac delta distribution

Suppose that T has a Dirac delta distribution at $t = \omega$, giving the survival function

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

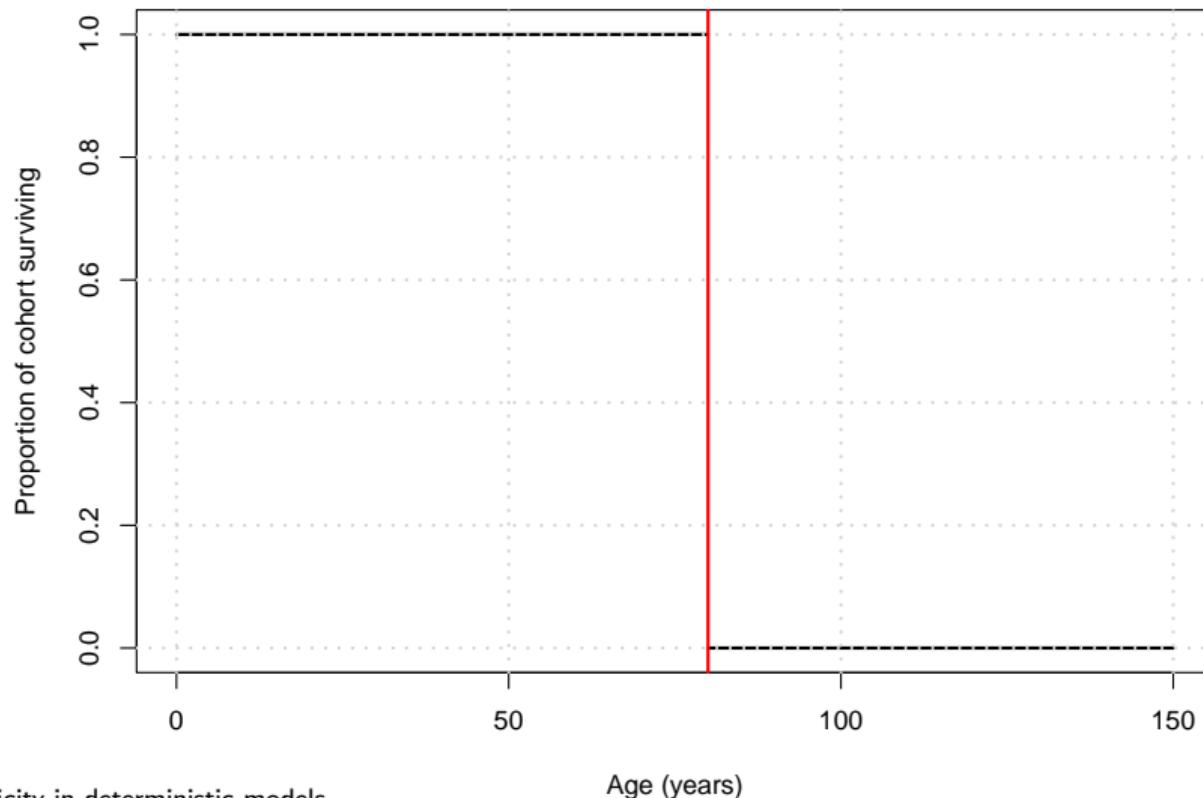
Then (??) takes the form

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

All individuals survive until time ω , then they all die at time ω

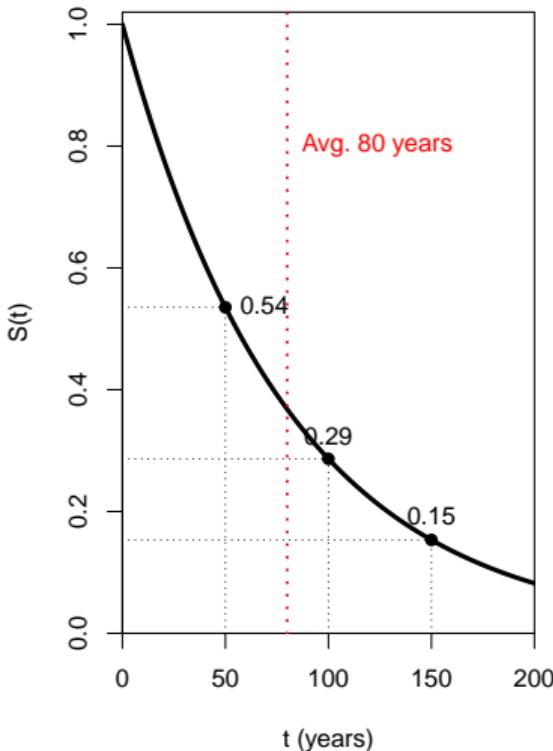
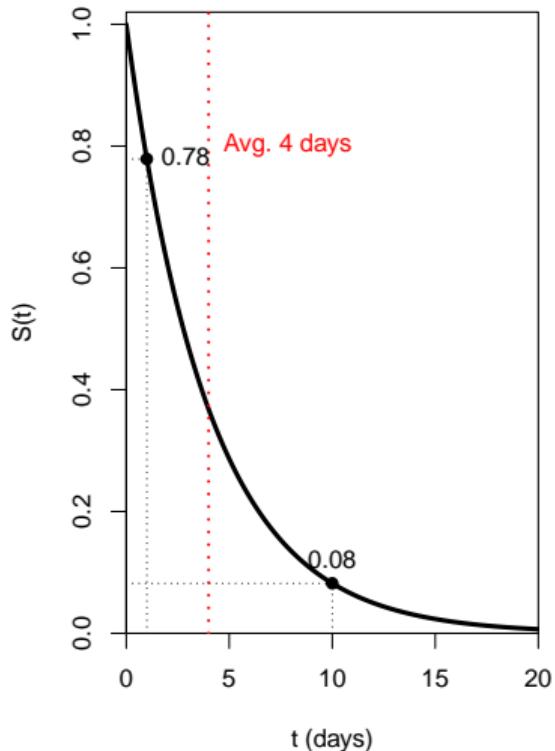
Here, $N' = 0$ everywhere except at $t = \omega$, where it is undefined

Survival function, $\mathcal{S}(t) = \mathbb{P}(T > t)$, for a Dirac distribution with mean 80 years





Survival for the exponential distribution



Issues with the exponential distribution

- ▶ Survival drops quickly
- ▶ Survival continues way beyond the mean

Acceptable if what matters is the average duration of sojourn in a compartment (e.g., long term dynamics)

More iffy if one is interested in short-term dynamics

- ▶ Exponential distribution with parameter θ has same mean and standard deviation $1/\theta$, i.e., a single parameter controls mean and dispersion about the mean

Exponential distributions are “bad” but also cool

X_1 and X_2 2 i.i.d. (independent and identically distributed) r.v. with parameters θ_1 and θ_2 . Then the probability density function of the r.v. $Z = X_1 + X_2$ is given by the convolution

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

The tool we use

Theorem 1

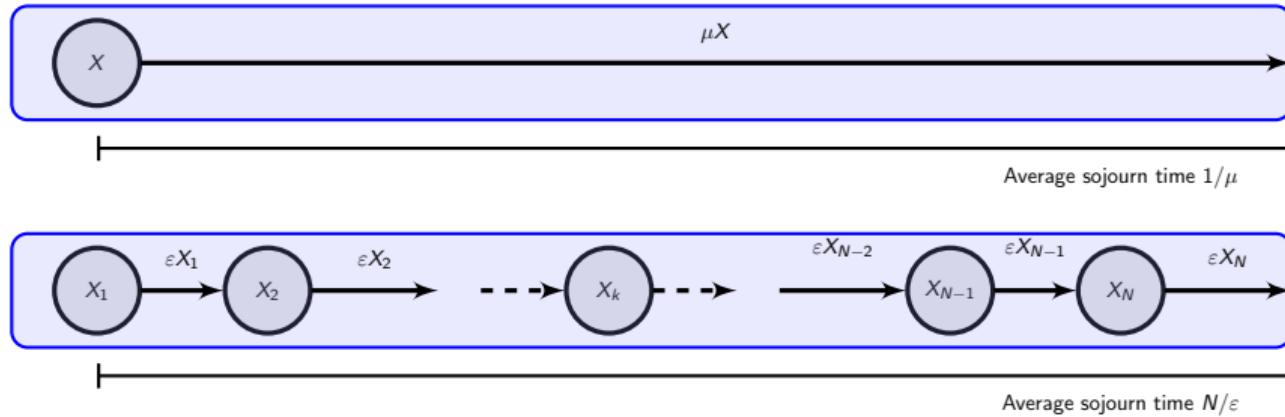
Let X_i be independent exponentially distributed random variables with parameter ξ and $Y = \sum_{i=1}^n X_i$

Then the random variable $YE(n, \xi)$, an Erlang distribution with shape parameter n and scale parameter ξ

(Erlang distribution: Gamma distribution with integer shape parameter)

Consequences for compartmental models

If n compartments are traversed successively by individuals, with each compartment having an outflow rate of $1/\xi$ (or a mean sojourn time of ξ), then the time of sojourn from entry into the first compartment to exit from the last is Erlang distributed with mean $E(Y) = n\xi$ and variance $\text{Var}(Y) = n\xi^2$



I have a Shiny app for this :)

Example: EVD incubation periods

Consider the incubation period for Ebola Virus Disease. During the 2014 EVD crisis in Western Africa, the WHO Ebola Response Team estimated incubation periods in a 2015 paper

Table S2 in the Supplementary Information in that paper gives the best fit for the distribution of incubation periods for EVD as a Gamma distribution with mean 10.3 days and standard deviation 8.2, i.e., $n\varepsilon = 10.3$ and $\varepsilon\sqrt{n} = 8.2$

From this, $\varepsilon = 8.2^2/10.3 \simeq 6.53$ and $n = 10.3^2/8.2^2 \simeq 1.57$. However, that is a Gamma distribution

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

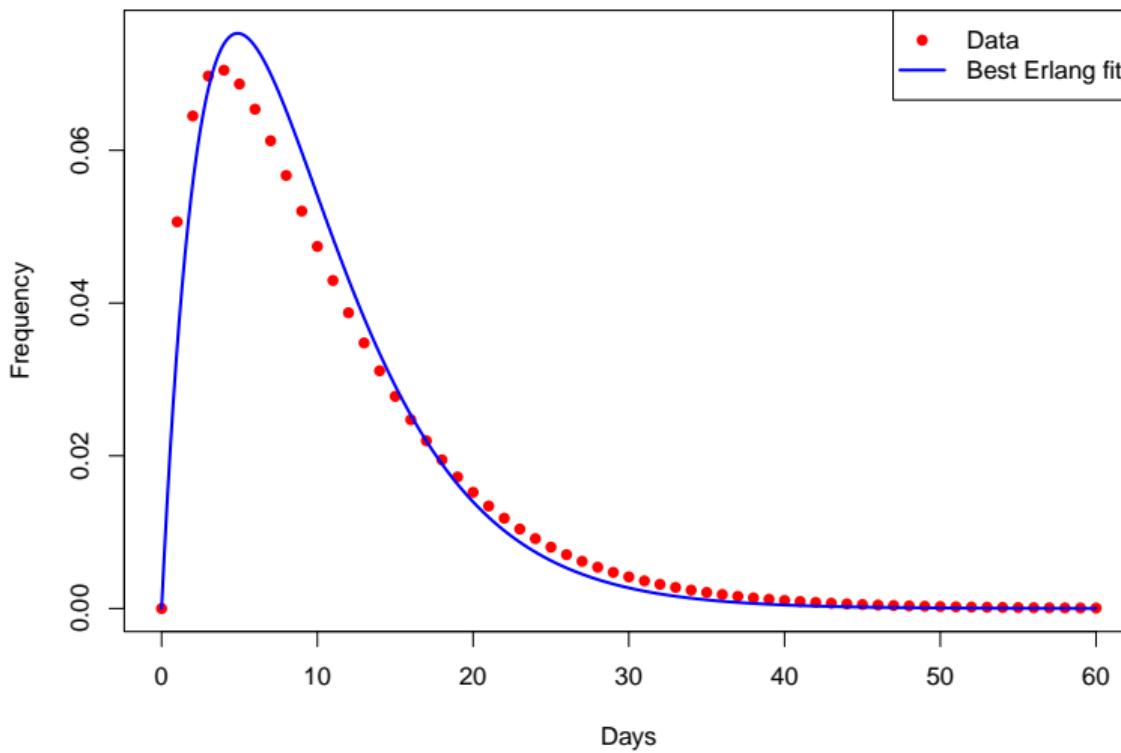
```
error_Gamma <- function(theta,shape,t,d) {  
  test_points <- dgamma(t, shape = shape, scale = theta)  
  ls_error <- sum((d-test_points)^2)  
  return(ls_error)  
}
```

```
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)
# Run the minimization
optim_fits <- optimize_gamma(time_points,data_points)
# Which is the best Erlang to fit the data
idx_best <- which.min(optim_fits$error)
```

We find the best fit below, which is obtained using 2 compartments





An SIS model

Hypotheses

- ▶ Individuals typically recover from the disease
- ▶ The disease does not confer immunity
- ▶ There is no birth or death (from the disease or natural)
⇒ Constant total population $N \equiv N(t) = S(t) + I(t)$
- ▶ Infection is of **standard incidence** type

Recovery

- ▶ Traditional models suppose that recovery occurs with rate constant γ
- ▶ Here, of the individuals that become infective at time t_0 , a fraction $S(t - t_0)$ remain infective at time $t \geq t_0$
- ▶ \Rightarrow For $t \geq 0$, $S(t)$ is a survival function. As such, it verifies $S(0) = 1$ and S is nonnegative and nonincreasing

Model for infectious individuals

Since N is constant, $S(t) = N - I(t)$ and we need only consider the following equation (where S is used for clarity)

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

- ▶ $I_0(t)$ number of individuals who were infective at time $t = 0$ and still are at time t
 - ▶ $I_0(t)$ is nonnegative, nonincreasing, and such that $\lim_{t \rightarrow \infty} I_0(t) = 0$
- ▶ $S(t-u)$ proportion of individuals who became infective at time u and who still are at time t

Expression under the integral

Integral equation for the number of infective individuals:

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

The term

$$\beta \frac{(N - I(u))I(u)}{N} e^{-\gamma(t-u)}$$

- ▶ $\beta(N - I(u))I(u)/N$ is the rate at which new infectives are created, at time u
- ▶ multiplying by $e^{-\gamma(t-u)}$ gives the proportion of those who became infectives at time u and who still are at time t

Summing over $[0, t]$ gives the number of infective individuals at time t

Case of an exponentially distributed time to recovery

Suppose $\mathcal{S}(t)$ such that sojourn time in the infective state has exponential distribution with mean $1/\gamma$, i.e., $\mathcal{S}(t) = e^{-\gamma t}$

Initial condition function $I_0(t)$ takes the form

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

with $I_0(0)$ the number of infective individuals at time $t = 0$. Obtained by considering the cohort of initially infectious individuals, giving a model such as (??)

Equation (??) becomes

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

Taking the time derivative of (??) yields

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

This is the classical logistic type ordinary differential equation (ODE) for I in an SIS model without vital dynamics (no birth or death)

Case of a step function survival function

Consider case where the time spent infected has survival function

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

i.e., the sojourn time in the infective state is a constant $\omega > 0$

In this case (??) becomes

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

Here, it is more difficult to obtain an expression for $I_0(t)$. It is however assumed that $I_0(t)$ vanishes for $t > \omega$

When differentiated, (??) gives, for $t \geq \omega$,

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

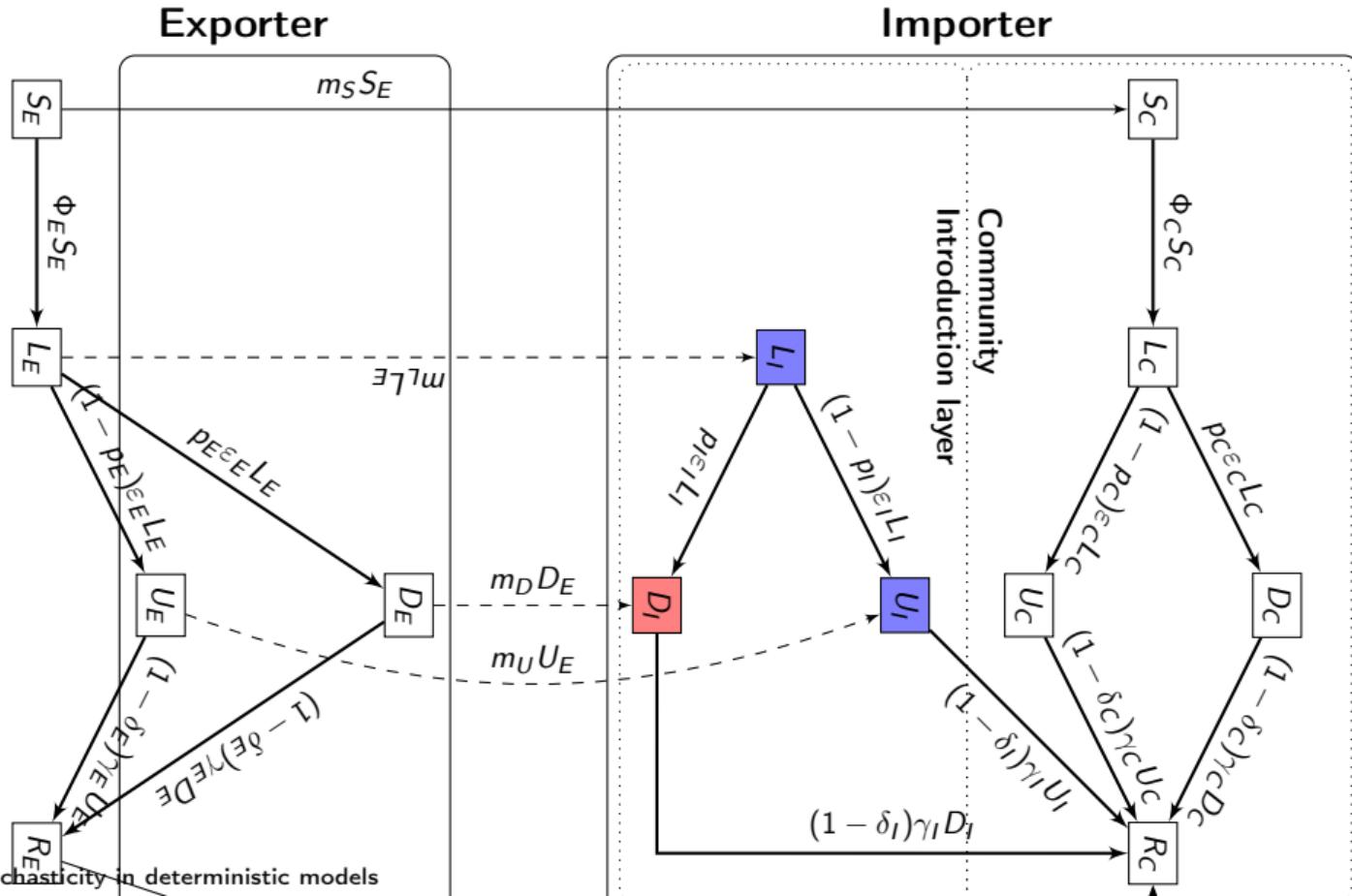
Since $I_0(t)$ vanishes for $t > \omega$, this gives the delay differential equation (DDE)

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



FIGS/ArinoCookePvdDVelasco.png

Model structure



FIGS/sim_2023_07_11-11_06_41_ID_9_run86.png

FIGS/sim_2023_07_11-11_06_41_ID_9_run2.png

FIGS/sim_2023_07_11-11_06_41_ID_9_run3.png

FIGS/sim_2023_07_11-11_06_41_ID_25_run5.png

FIGS/exporter_importer_panel_zoom.png

One last thought for the road

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.



What is a Branching Process?

The Core Idea

A branching process is a mathematical model for a population where individuals produce a random number of offspring and then die.

- ▶ Think of bacteria splitting, a virus spreading, or even the survival of family surnames.
- ▶ We start with an initial population, Z_0 .
- ▶ Each individual in generation n produces a number of offspring for generation $n+1$.
- ▶ This "number of offspring" is a random variable. All individuals produce offspring according to the same probability distribution, independently of each other.

The Galton-Watson Process

Let Z_n be the size of the population in generation n . We typically start with $Z_0 = 1$. The population evolves according to the rule:

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

- ▶ The term $X_{n,i}$ represents the number of offspring produced by the i -th individual in generation n .
- ▶ The variables $\{X_{n,i}\}$ are assumed to be **independent and identically distributed (i.i.d.)** integer-valued random variables.
- ▶ We call their common distribution $\{p_k\}_{k=0}^{\infty}$ the **offspring distribution**, where $p_k = P(X = k)$.

The Fundamental Questions

1. What is the long-term expected size of the population?
2. What is the probability that the population eventually dies out?

Mean Offspring

The fate of the population hinges on a single parameter: the mean of the offspring distribution

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

Expected Population Size

Using the law of total expectation, we find the expected size of the next generation:

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

Taking the expectation again, we get a simple recurrence:

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

The Three Regimes of Population Growth

The behavior of $E[Z_n] = Z_0\mu^n$ suggests three distinct cases:

Subcritical ($\mu < 1$)

$E[Z_n] \rightarrow 0$. The population is expected to shrink. It goes extinct with probability 1.

Critical ($\mu = 1$)

$E[Z_n] = Z_0$. The population is expected to remain stable. Surprisingly, it still goes extinct with probability 1.

Supercritical ($\mu > 1$)

$E[Z_n] \rightarrow \infty$. The population is expected to grow exponentially. It has a non-zero probability of surviving forever.

Tool: The Probability Generating Function

To find the extinction probability, we need a powerful tool: the **probability generating function (PGF)** of the offspring distribution X .

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

Key Properties

- ▶ $G(1) = \sum p_k = 1$
- ▶ The mean can be found from the derivative: $G'(1) = \sum kp_k = \mu$.
- ▶ The PGF of Z_n is the n -th iterate of $G(s)$ with itself. If $G_n(s)$ is the PGF of Z_n , then $G_{n+1}(s) = G(G_n(s))$.

The Extinction Probability Equation

Let π_0 be the probability of eventual extinction, starting with $Z_0 = 1$.

$$\pi_0 = P(\text{population dies out}) = \lim_{n \rightarrow \infty} P(Z_n = 0)$$

Since $P(Z_n = 0) = G_n(0)$, and $G_{n+1}(0) = G(G_n(0))$, in the limit the extinction probability π_0 must satisfy the equation:

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

Theorem 2

*The extinction probability π_0 is the **smallest non-negative solution** to the equation $s = G(s)$.*

- ▶ If $\mu \leq 1$, the only solution in $[0, 1]$ is $s = 1$. So $\pi_0 = 1$.
- ▶ If $\mu > 1$, there is a unique solution in $[0, 1)$, which is the extinction probability $\pi_0 < 1$.

From Discrete to Continuous Time

Limitation of Galton-Watson

Generations don't happen in synchronized steps in the real world. Individuals give birth and die at random times.

This leads us to **Continuous-Time Markov Chains (CTMCs)**.

- ▶ The state of the system is the population size, $k \in \{0, 1, 2, \dots\}$.
- ▶ Instead of generations, we have transition rates:
 - ▶ λ_k : rate of birth when population is size k (moves to $k + 1$).
 - ▶ δ_k : rate of death when population is size k (moves to $k - 1$).
- ▶ Often, we assume these rates are linear: $\lambda_k = k\lambda$ and $\delta_k = k\delta$. This means individuals act independently.

Branching Process Approximation of a CTMC

The Key Insight

At the beginning of an outbreak (or for a very large population), the dynamics caused by a single individual are largely independent of others.

This allows us to approximate the start of a CTMC population process with a branching process.

Example: A Simple Epidemic (SIR Model)

- ▶ S : Susceptible, I : Infected, R : Recovered.
- ▶ An infected person meets others at a certain rate. If they meet a susceptible, a new infection may occur (an "offspring").
- ▶ The infected person recovers (or dies) at another rate, ending their infectious period.
- ▶ **Question:** How many new infections does a single infected person cause on average?

Case Study: The Basic Reproduction Number \mathcal{R}_0

Consider a single infected individual in a large population of susceptibles.

- ▶ Let β be the infection rate (rate of producing "offspring").
- ▶ Let γ be the recovery rate (rate of "dying").

The individual's infectious lifetime is an exponential random variable with mean $1/\gamma$.
The average number of secondary infections they cause is:

$$\mathcal{R}_0 = (\text{rate of infection}) \times (\text{average infectious period}) = \beta \times \frac{1}{\gamma} = \frac{\beta}{\gamma}$$

The Connection

\mathcal{R}_0 is precisely the **mean offspring number** μ for the embedded branching process that approximates the start of the epidemic.

Applying Branching Theory to Epidemics

The fate of the epidemic's initial phase is determined by \mathcal{R}_0 :

- ▶ If $\mathcal{R}_0 \leq 1$ ($\mu \leq 1$): The number of infected individuals is a subcritical or critical process. The epidemic will die out with probability 1.
- ▶ If $\mathcal{R}_0 > 1$ ($\mu > 1$): The process is supercritical. There is a positive probability that the epidemic takes off and causes a major outbreak.

We can even calculate the probability of a major outbreak! It is $1 - \pi_0$, where π_0 is the extinction probability.

For this simple birth-death infection process, the PGF is $G(s) = \frac{\gamma}{\beta+\gamma} + \frac{\beta}{\beta+\gamma}s$. Solving $s = G(s)$ gives the extinction probability:

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

The probability of a major outbreak is $1 - 1/\mathcal{R}_0$.

Summary

- ▶ **Branching Processes** model populations with i.i.d. offspring generation.
- ▶ The fate of the population is determined by the **mean offspring number** μ . Extinction is certain if $\mu \leq 1$.
- ▶ The **extinction probability** π_0 can be calculated as the smallest non-negative fixed point of the probability generating function $G(s)$.
- ▶ The initial stages of many large-scale **Continuous-Time Markov Chains** can be approximated by a branching process.
- ▶ This allows us to apply the theory to real-world problems, like calculating an epidemic's **basic reproduction number** \mathcal{R}_0 and its probability of causing a major outbreak.

```
# From https://stackoverflow.com/questions/36868287/purl-within-knit-duplica
rmd_chunks_to_r_temp <- function(file){
  callr::r(function(file, temp){
    out_file = sprintf("../CODE/%s", gsub(".Rnw", ".R", file))
    knitr::purl(file, output = out_file, documentation = 1)
  }, args = list(file))
}
rmd_chunks_to_r_temp("course-03-stochastic-aspects.Rnw")
```

```
## Error: ! in callr subprocess.
## Caused by error in 'file(con, "r")':
## ! cannot open the connection
```

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.