Sojourn time in compartments

MATH 8xyz – Lecture 06

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

Winter 20XX

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

Distributions of times to events

Two “extreme” distributions and a nicer one

A simple cohort model with death

Possible fixes to the exponential distribution issue

Sojourn times in an SIS disease transmission model

Distributions of times to events

Two “extreme” distributions and a nicer one

A simple cohort model with death

Possible fixes to the exponential distribution issue

Sojourn times in an SIS disease transmission model

See some of the work of Horst Thieme on the subject

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 [?] (link) for arbitrary distributions

Time to events

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 T be the random variable "time spent in state A before switching into state B"

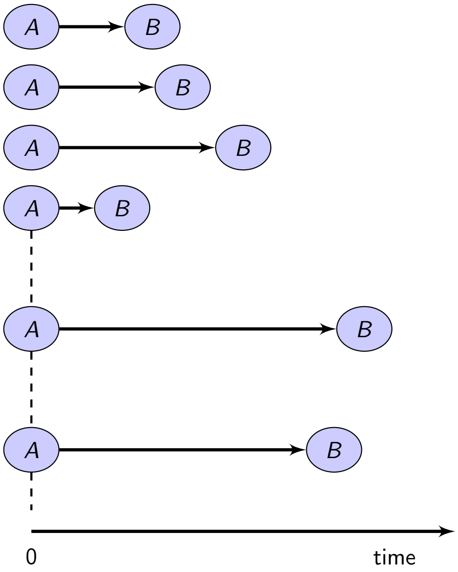
The states can be anything:

- 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
B
A
B
A
B
A
B
A
B
A
B
0
time

A distribution of probability is a model

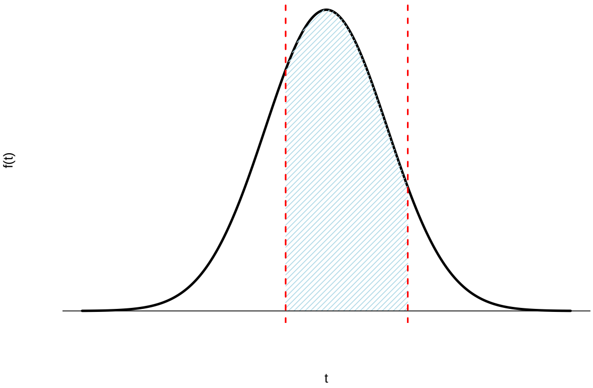
From the sequence of experiments, we deduce a model, which in this context is called a **probability distribution**

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

Probability density function

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

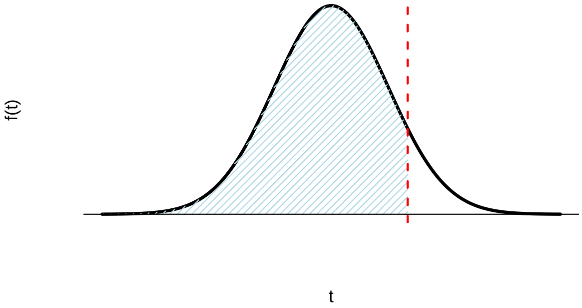
- $f \geq 0$
- $\int_{-\infty}^{+\infty} f(s)ds = 1$
- $\mathbb{P}(a \leq T \leq b) = \int_a^b f(t)dt$



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

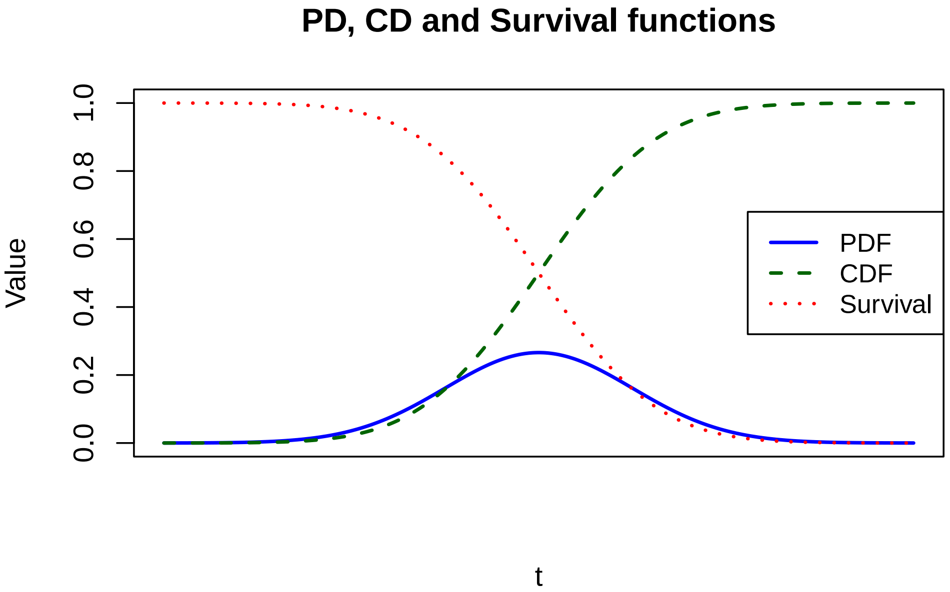
Another characterization of the distribution of the random variable T is through the **survival** (or **sojourn**) function

The survival function of state A is given by

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

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

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


PD, CD and Survival functions
Value
1.0
0.8
0.6
0.4
0.2
0.0
t
PDF
CDF
Survival

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

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

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

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

Expected future lifetime:

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

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

Hazard rate

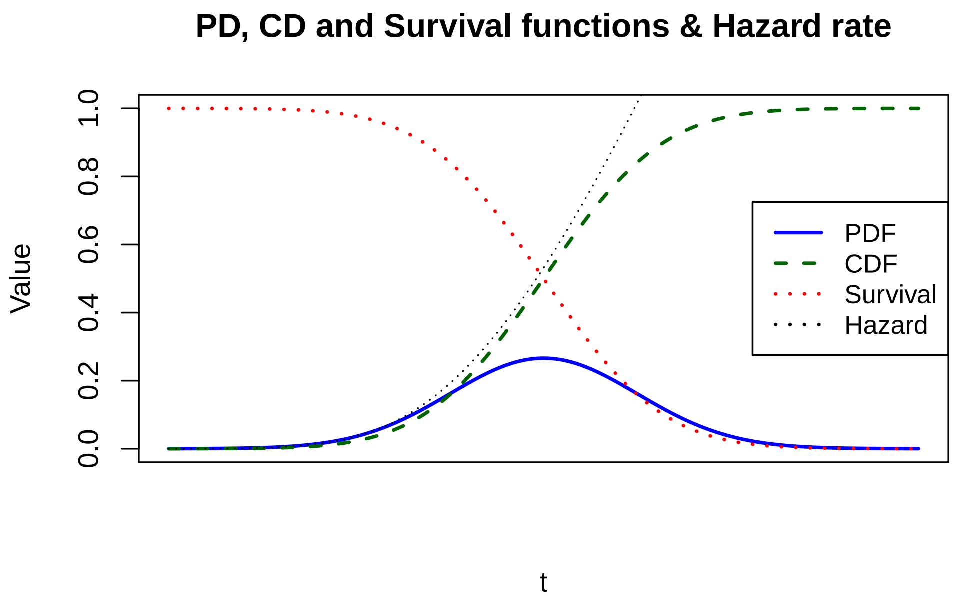
The **hazard rate** (or **failure rate**) is

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

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
Value
t
0.0
0.2
0.4
0.6
0.8
1.0
PDF
CDF
Survival
Hazard

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

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

(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

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

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

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

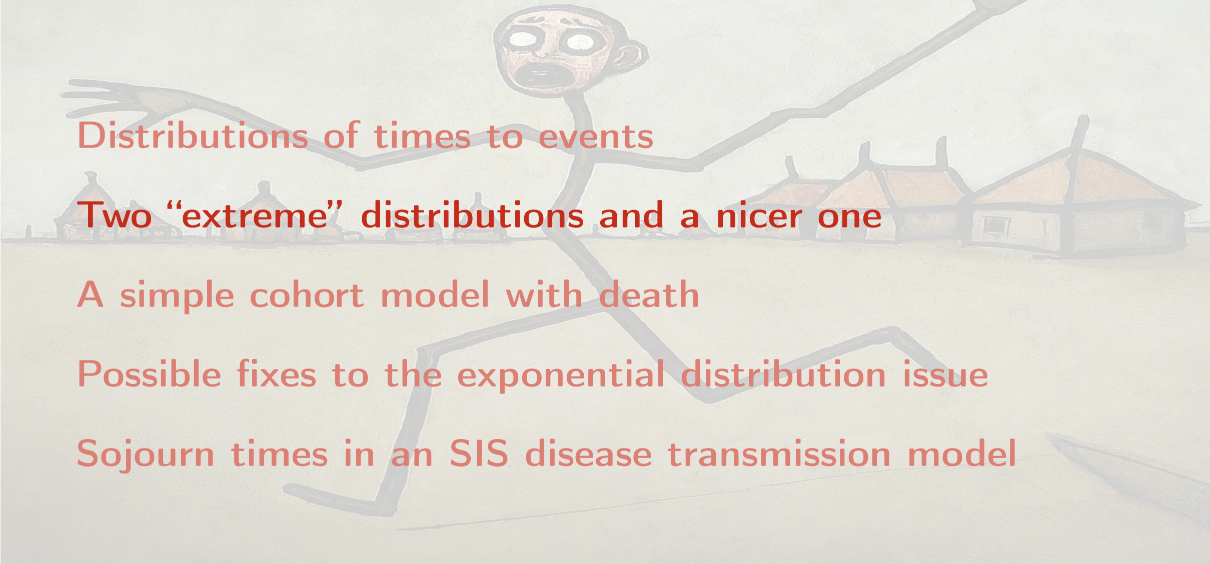
or, identically,

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

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

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



Distributions of times to events

Two "extreme" distributions and a nicer one

A simple cohort model with death

Possible fixes to the exponential distribution issue

Sojourn times in an SIS disease transmission model

The exponential distribution

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

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

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

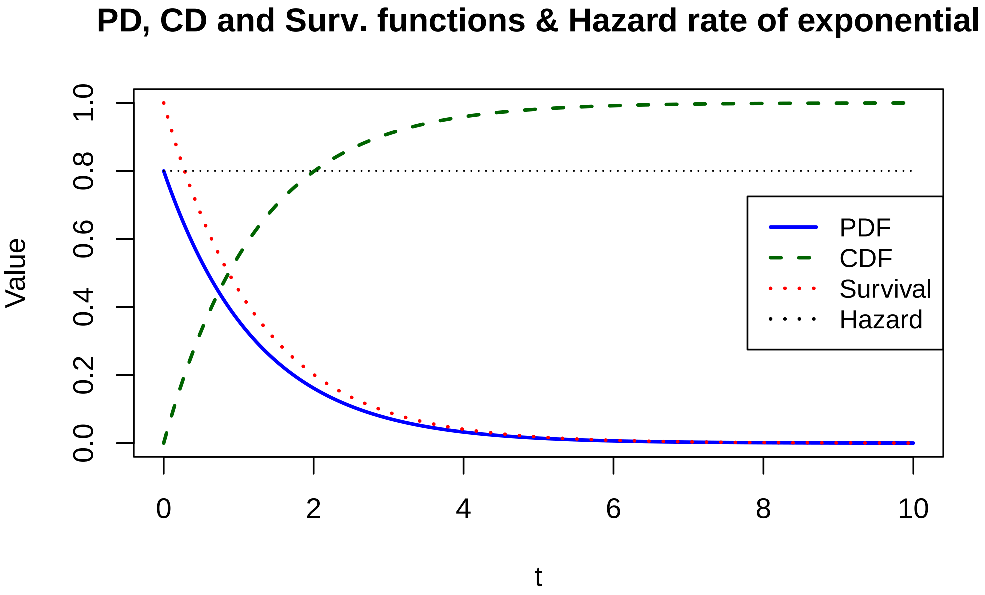
The standard deviation of an exponential distribution is also $1/\theta$. When estimating θ, it is impossible to distinguish the mean and the standard deviation

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
Value
t
0.0
0.2
0.4
0.6
0.8
1.0
0
2
4
6
8
10
PDF
CDF
Survival
Hazard

The Dirac delta distribution

If for some constant $\omega > 0$,

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

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

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

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

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

Properties of the Gamma distribution

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

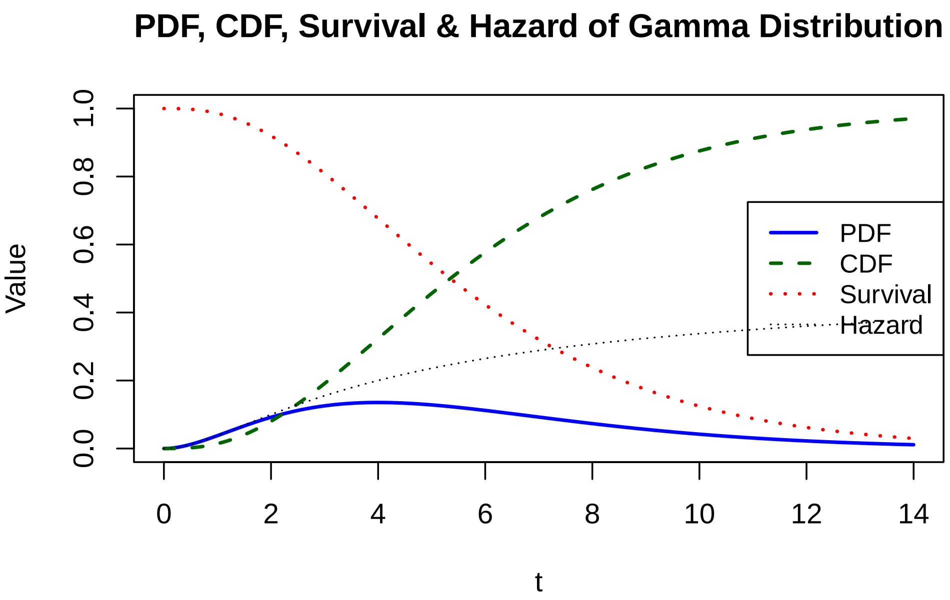
Survival function

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

where

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

is an incomplete Gamma function



PDF, CDF, Survival & Hazard of Gamma Distribution
Value
t
0.0
0.2
0.4
0.6
0.8
1.0
0
2
4
6
8
10
12
14
PDF
CDF
Survival
Hazard

Distributions of times to events

Two “extreme” distributions and a nicer one

A simple cohort model with death

Possible fixes to the exponential distribution issue

Sojourn times in an SIS disease transmission model

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 $\mathcal{S}(t)$

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

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

$\mathcal{S}(t)$ proportion of initial population still alive at time t, so $N_0 \mathcal{S}(t)$ 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 $\mathcal{S}(t) = e^{-dt}$, and (4) takes the form

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

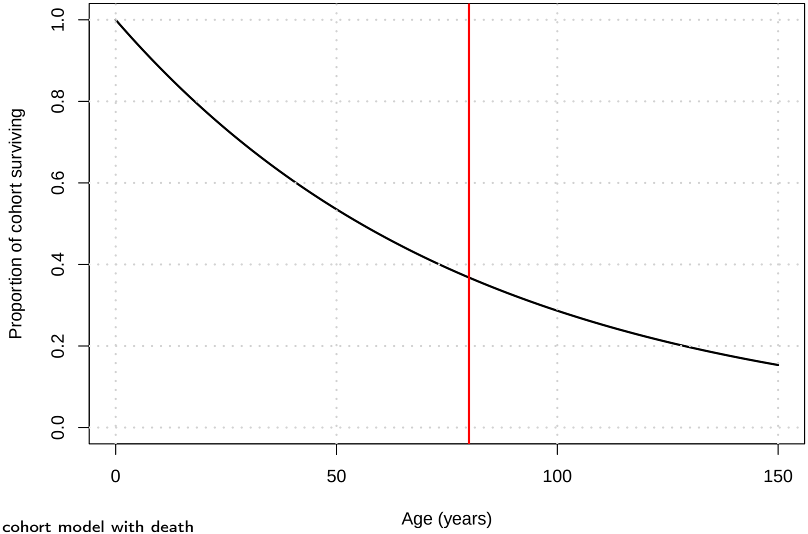
Now note that

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

with $N(0) = N_0$.

$\Rightarrow$ 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

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

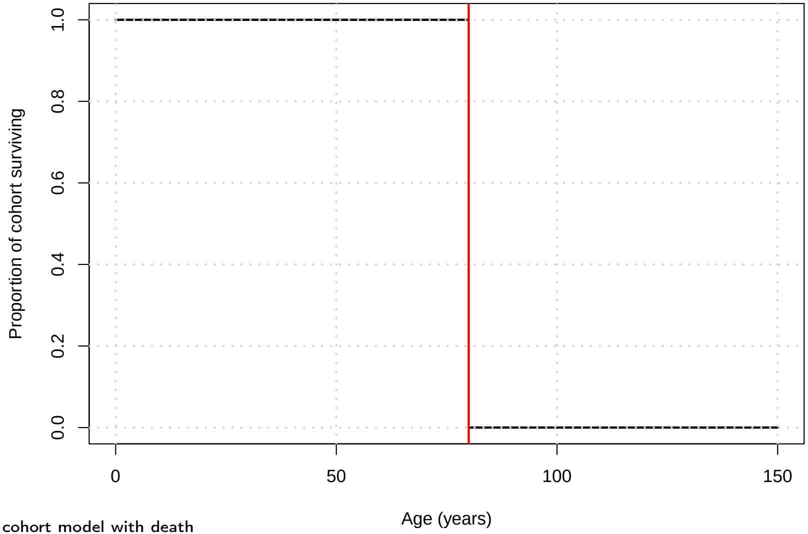
Then (4) takes the form

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

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



Distributions of times to events

Two “extreme” distributions and a nicer one

A simple cohort model with death

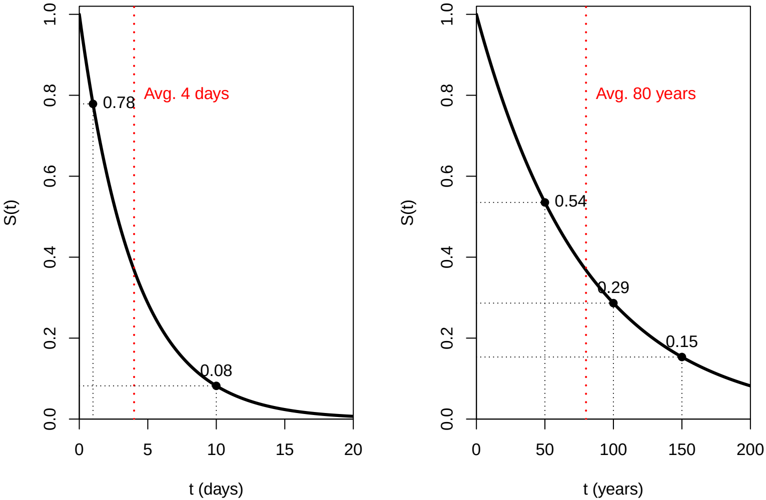
Possible fixes to the exponential distribution issue

Sojourn times in an SIS disease transmission model

Possible fixes to the exponential distribution issue

- **The "issue"**
- Fix 1 – Use info on the distribution as well
- Fix 2 – Use an Erlang distribution
- Finding the right Erlang
- Example – SLIAR-type model with Erlang

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

Possible fixes to the exponential distribution issue

Possible fixes to the exponential distribution issue

- The “issue”
- Fix 1 – Use info on the distribution as well
- **Fix 2 – Use an Erlang distribution**
- Finding the right Erlang
- Example – SLIAR-type model with Erlang

Exponential distributions are "bad" but also cool

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

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

The tool we use

Theorem 1

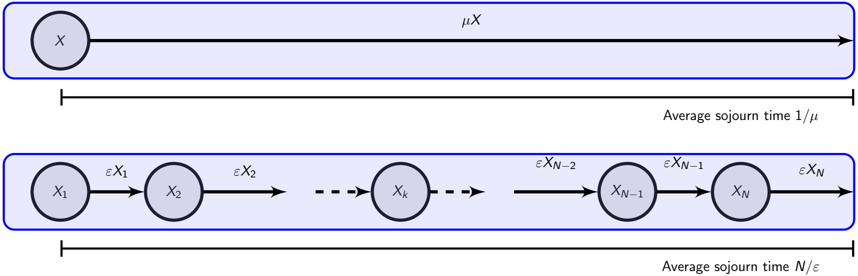
Let X_i be independent exponentially distributed random variables with parameter ξ and $Y = \sum_{i=1}^{n} X_i$

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

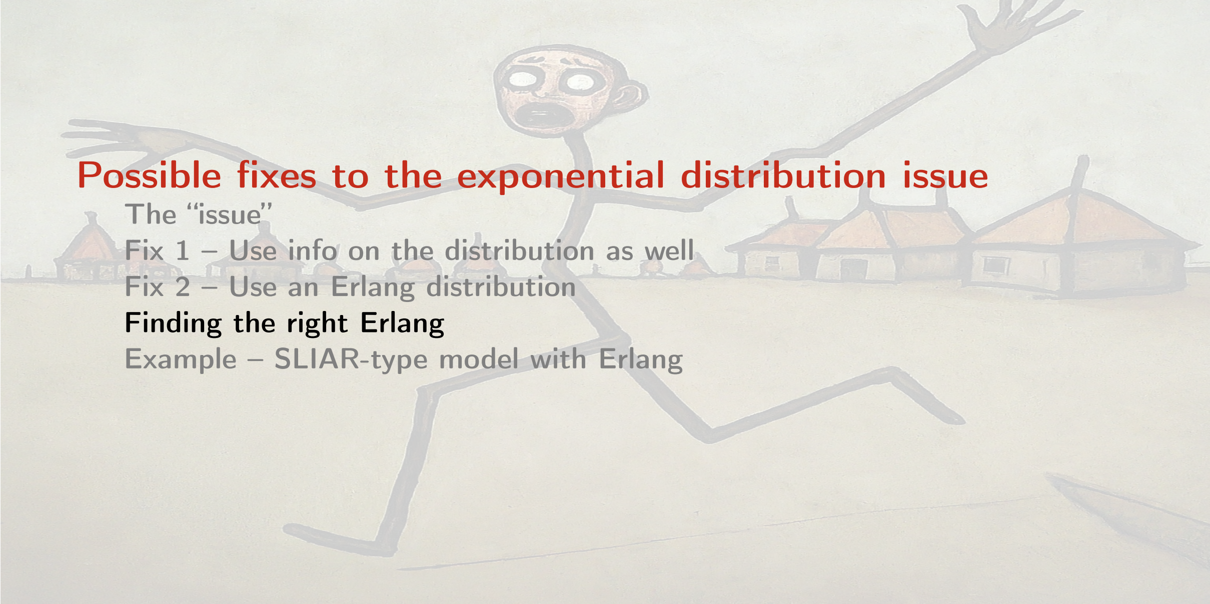
(Erlang distribution: Gamma distribution with integer shape parameter)

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 $\mathrm{Var}(Y) = n\xi^2$



I have a Shiny app for this :)



Possible fixes to the exponential distribution issue

- The “issue”
- Fix 1 – Use info on the distribution as well
- Fix 2 – Use an Erlang distribution
- **Finding the right Erlang**
- Example – SLIAR-type model with Erlang

Example: EVD incubation periods

During the 2014 Ebola Virus Disease (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

$\Longrightarrow$ compute RSS errors between data points generated from the given Gamma distribution and an Erlang

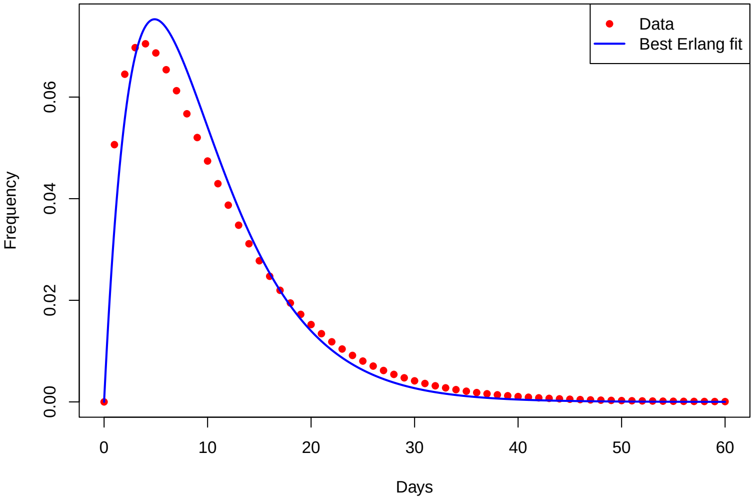
```
error_Gamma <- function(theta,shape,t,data) {
  test_points <- dgamma(t, shape = shape, scale = theta)
  ls_error <- sum((data-test_points)^2)
  return(ls_error)
}
```

```r
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,
                              data = 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)
}
```

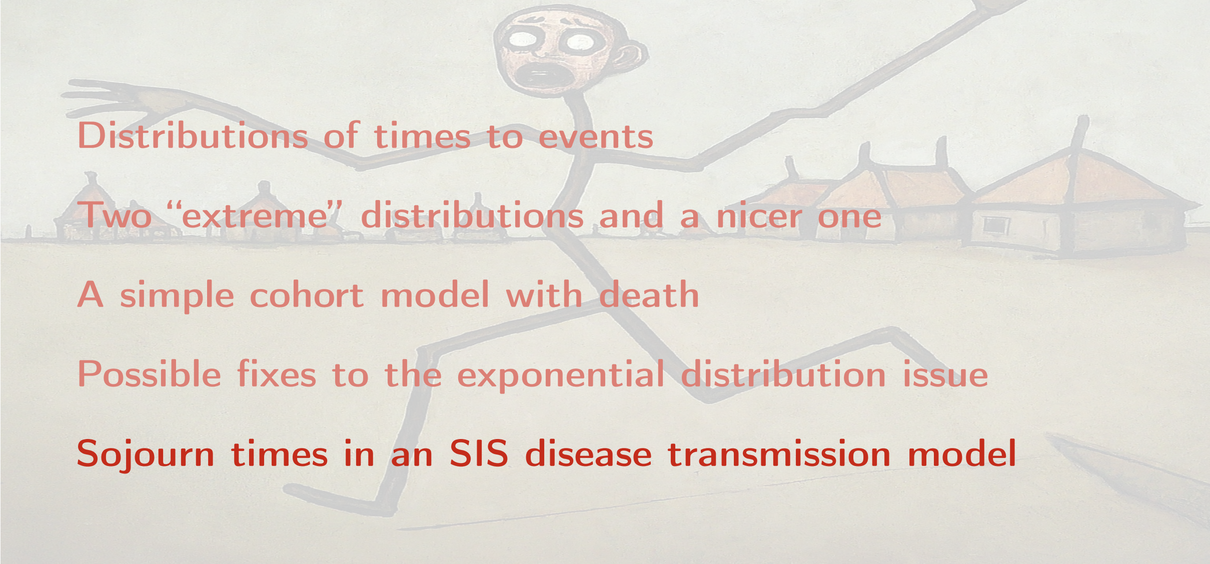
```r
time_points <- seq(0,60)
data_points <- dgamma(time_points, shape = 1.57,
                      scale = 6.53)
# Run the minimization
optim_fits <- optimize_gamma(time_points,data_points)
# Which is the best Erlang to fit the data
idx_best <- which.min(optim_fits$error)
```

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



Possible fixes to the exponential distribution issue

- The “issue”
- Fix 1 – Use info on the distribution as well
- Fix 2 – Use an Erlang distribution
- Finding the right Erlang
- **Example – SLIAR-type model with Erlang**



Distributions of times to events

Two "extreme" distributions and a nicer one

A simple cohort model with death

Possible fixes to the exponential distribution issue

Sojourn times in an SIS disease transmission model

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)
  $\Rightarrow$ 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 $\mathcal{S}(t - t_0)$ remain infective at time $t \geq t_0$
- $\Rightarrow$ For $t \geq 0$, $\mathcal{S}(t)$ is a survival function. As such, it verifies $\mathcal{S}(0) = 1$ and $\mathcal{S}$ 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)

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

- $I_0(t)$ 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 $\lim_{t\to\infty} I_0(t) = 0$
- $\mathcal{S}(t-u)$ 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:

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

The term

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

- $\beta(N - I(u))I(u)/N$ is the rate at which new infectives are created, at time u
- multiplying by $\mathcal{S}(t-u)$ 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 (4)

Equation (8) becomes

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

Taking the time derivative of (9) 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

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

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

In this case (8) becomes

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

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, (10) 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)

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

Bibliography I