VIRALY

An experiment in discrete viral models.

github.com/ghomem/viraly

Basic ideia

A viral phenomenon is the spreading of a message (a disease, an idea, a name, a brand) across a population by a mechanism where the infected agents at time t determine the infected agents at time t + 1.

The variation of infections is influenced, among other things, by the number of contacts per unit of time and the probability of transmission during those contacts.

On these simplified models we will assume that all agents perform the same number of contacts per unit of time and behave roughly in the same way, so that the probability of transmission can be considered the same.

Parameters in use for the different models

h – average number of contacts per unit time (ex: day) of each agent (h = agent horizon)
p – probability of transmission in each contact
N_0 – initial number of infected agents
M – population size
T – infection duration

Parameter h depends on population behaviour whereas the probability of transmission p can be affected by both behavioral and environmental conditions.

1. Permanent infection, infinite population

$$N_{-1} = 0$$

$$N_0 = N_0$$

$$N_1 = N_0 + N_0 hp = N_0 (1 + hp)$$

$$N_2 = N_0 + N_0 hp + (N_0 + N_0 hp) hp$$

$$N_2 = N_1 (1 + hp)$$

...

$$N_t = N_{t-1} (1 + hp) = \underbrace{N_0 (1 + hp)^t}_{\text{explicit solution}}$$

In this case we have a geometric progression of ratio (1+ hp). This is the behaviour of all epidemics of large populations and high recovery times.

This is a simple exponential model.

2. Permanent infection, finite population correction

$$N_{-1} = 0$$

$$N_0 = N_0$$

$$N_1 = N_0 + N_0 hp \overbrace{\left(1 - \frac{N_0}{M}\right)}^{\text{correction}}$$

$$N_2 = N_0 + N_0 hp \left(1 - \frac{N_0}{M}\right) + \left(N_0 + N_0 hp \left(1 - \frac{N_0}{M}\right)\right) hp \left(1 - \frac{N_1}{M}\right)$$

$$N_2 = N_1 \left(1 + hp \left(1 - \frac{N_1}{M}\right)\right)$$

...

$$N_t = N_{t-1} \left(1 + hp \left(1 - \frac{N_{t-1}}{M}\right)\right)$$

The correction factor accounts for the fact that the effective horizon of infected agents narrows down as more and more infections happen.

This is a logistic growth model.

3. Temporary infection with fixed duration T, finite population correction

$$N_{-X} = 0$$

$$\dots$$

$$N_{-1} = 0$$

Definition necessary for the general formula to hold

$$N_0 = N_0$$

New cases (n) Outgoing cases (O)

$$N_1 = N_0 + N_0 hp\left(1 - \frac{M - m_0}{M}\right) - n_{1-T}$$

$$N_2 = N_1 + N_1 hp\left(1 - \frac{M - m_1}{M}\right) - n_{2-T}$$

$$N_t = N_{t-1} + N_{t-1} hp\left(1 - \frac{M - m_{t-1}}{M}\right) - n_{t-T}$$

Important: outgoing cases have in the past been new cases so let's not discount them twice.

$$m_0 = M - N_0$$

$$m_1 = m_0 - n_1$$

$$m_2 = m_1 - n_2$$

$$m_t = m_{t-1} - n_t$$

N_t = Infected
m_t = Susceptible
O_t = Removed (Recovered or Dead)

This is a SIR-like model.

In this model the outgoing cases at time t are simply the new cases at time t-T.

$$n_{t-T} = hp\, N_{t-T-1}$$

The model is equal to model 2 until the first removals. First removals are for t -T > 0, i.e.:

- at t=1 if T=1
- at t=2 if T=2
- ...
- at t=k if T=k

If T=1

New cases can be calculated like this because there are no removals yet

$$N_1 = N_0 + N_0 hp\left(1 - \frac{N_0}{M}\right) - N_0$$

If T=2

$$N_2 = N_1 + N_1 hp\left(1 - \frac{M - m_1}{M}\right) - N_0 \qquad N_2 = N_1 + N_1 hp\left(1 - \frac{M - (m_0 - n_1)}{M}\right) - N_0$$

$$N_2 = N_1 + N_1 hp\left(1 - \frac{M - (m_0 - (N_1 - N_0))}{M}\right) - N_0$$

$$N_2 = N_1 + N_1 hp\left(1 - \frac{M - ((M - N_0) - (N_1 - N_0))}{M}\right) - N_0$$

$$N_2 = N_1 + N_1 hp\left(1 - \frac{N_1}{M}\right) - N_0$$

If T=k

$$N_k = N_{k-1} + N_{k-1} hp\left(1 - \frac{N_{k-1}}{M}\right) - N_0$$

The conclusion is that the first removals are simply the initially infected agents being subtracted from the cases accumulated during k-1 iterations of the simple logistic growth model.

For a population which is very large compared to the initial number of infections the equation above is approximately:

$$N_k \simeq N_{k-1} + N_{k-1} hp - N_0$$

which for a generic iteration j >= T can be written:

$$N_{T+j} = N_{T+j-1} + N_{T+j-1} hp - N_{j-1} hp$$

This approximation holds because under the assumption of a very large population, recoveries start **way earlier than the difference before the logistic and the exponential growth is noticeable.**

Reference:

https://en.wikipedia.org/wiki/Linear_difference_equation

Setting β=hp we can write:

$$N_{T+j} = N_{T+j-1} + N_{T+j-1}\beta - N_{j-1}\beta$$

$$N_i = N_{i-1} + N_{t-1}\beta - N_{i-T-1}\beta$$

$$\boxed{N_{t+1} = N_t(1+\beta) - N_{t-T}\beta}$$

Solution = ?

$$N_t = C_1 1^t + C_2 \beta^t = N_0 \beta^t \qquad \text{T=1}$$

$$N_t = C_1 1^t + C_2 \lambda_2^t + C_3 \lambda_3^t \qquad \text{T=2}$$

$$\lambda_2 = \frac{\beta}{2}\left(1+\sqrt{1+\frac{4}{\beta}}\right) \quad >0 \qquad \lambda_2 < 1 \Rightarrow \beta < \frac{1}{2}$$

$$\lambda_3 = \frac{\beta}{2}\left(1-\sqrt{1+\frac{4}{\beta}}\right) \quad <0 \qquad |\lambda_3| < 1 \;\forall\; \beta > 0$$

Characteristic equation

$$\lambda^{T+1} - (1+\beta)\lambda^T + \beta = 0$$

$$(\lambda-1)(\lambda-\beta) = 0 \qquad \text{T=1}$$

$$(\lambda-1)(\lambda^2-\beta\lambda-\beta) = 0 \qquad \text{T=2}$$

$$(\lambda-1)(\lambda^3-\beta\lambda^2-\beta\lambda-\beta) = 0 \qquad \text{T=3}$$

...

Since $N_{-1} = N_{-2} = 0$:

$$N_0 = C_1 + C_2 + C_3$$

$$N_0(1+\beta) = C_1 + C_2\lambda_2 + C_3\lambda_3$$

$$N_1(1+\beta) - N_1 = C_1 + C_2\lambda_2^2 + C_3\lambda_3^2$$

For the general case the characteristic equation is:

$$P(\lambda) = \lambda^{T+1} - (1+\beta)\lambda^{T} + \beta = 0$$

It is easy to see that:

$$P(0) = \beta$$
$$P(1) = 0$$
$$P(+\infty) = +\infty$$

$$P'(0) = 0$$
$$P'(C) = 0$$

$$P'(\lambda) = \lambda^{T-1}[(T+1)\lambda - T(1+\beta)]$$

$$C = T\frac{1+\beta}{1+T}$$

Given that P' has two zeros, P has at most three real roots, one of which we already know is 1. Given that P decreases from β to 0 in [0,1] and grows to infinity for large λ, C has to be a minimum which means that if C > 1 there will be another real root to the right of C, therefore being such root itself >1.

From this we conclude that for stability it is **necessary** that:

$$C = T\frac{1+\beta}{1+T} < 1 \Rightarrow \beta < \frac{1}{T}$$

If the necessary condition is met there will be a real root in the interval]0,1[.

If T is even (T+1 is odd), there will also be a negative real root, because in that case:

$$P(-\infty)=-\infty$$

By evaluating directly we note that P is already negative at -1, which means that the extra real root has module smaller than 1. There are 3 real roots (one positive, one negative, 1) plus (T-2)/2 pairs of complex roots.

If T is odd

$$P(+\infty)=+\infty$$

so there are 2 real roots (one positive, 1) plus (T-1)/2 pairs of complex roots.

We note that the characteristic equation can be re-written in the form:

$$P(\lambda)=(\lambda-1)(\lambda^T-\beta\sum_{k=0}^{T-1}\lambda^k)=0$$

whose second term fulfills the Schur dominance condition [*], which is sufficient for stability:

$$|a_n|<\sum_{k=0}^{n-1}|a_k| \longrightarrow 1<\beta\sum_{k=0}^{T-1}1$$

if $\beta<\frac{1}{T}$

We conclude that:

$$\beta < \frac{1}{T}$$

is **necessary and sufficient** for the finite difference equation to be stable.

The solutions found using the large population approximation represent an upper bound for the general solutions where the finite character of the population is modeled. This means that the condition above can be used as a **sufficient** propagation decay condition for the more general model where the population is finite. In this case the condition is **not necessary** because the finite character of the population becomes a self-limiting propagation mechanism that ends up stopping the propagation at some point.

In that more general case:

$$N_{t+1} = N_t + N_t \beta \left(1 - \frac{M - m_t}{M}\right) - N_{t-T} \beta \left(1 - \frac{M - m_{t-T}}{M}\right)$$

and given what has been analyzed before, a less demanding but still sufficient condition follows:

$$\beta < \frac{1}{T\left(1 - \frac{M - m_t}{M}\right)}$$

where m_t is the number susceptible individuals evaluated at time t.

The last equation can also be written:

$$R_t < 1$$

with:

$$R_t = \beta T \left(1 - \frac{M - m_t}{M}\right)$$

or

$$R_t = R_0 \left(1 - \frac{M - m_t}{M}\right)$$

where we defined:

$$R_0 = \beta T$$

We can also write

$$R_t = R_0 \left(1 - \epsilon\right)$$

where ε represents the immune fraction of the population. It is clear that for the epidemic to be controlled it is necessary that:

$$\epsilon > 1 - \frac{1}{R_0}$$

We have seen that the characteristic equation of:

$$N_{t+1} = N_t(1+\beta) - N_{t-T}\beta$$

has the following roots:

for even T:

- 1
- one positive root
- one negative root with | | <1
- (T-2)/2 pairs of complex roots

for odd T:

- 1
- one positive root
- (T-1)/2 pairs of complex roots

A solution for generic constant β would therefore be written:

$$N_t = C_1 + C_+ \lambda_+^{\ t} + C_- \lambda_-^{\ t} + \sum_{j=0}^{T-2} C_j \lambda_j^t$$

zero

(it is the only possible value)

| | <1

(fades over time)

What about these roots?

Theorem 3.9 (simplified from doi:10.1007/s10474-018-0896-6).

Let

$$P(z) = z^m + Az^{m-1} + B \qquad G(z) = z^m - Az^{m-1} - B$$

- **If B > 1/(m-1)** P(z) has m-1 zeros in $\delta_{m-1} < |z| < 1$ and one zero in (1, +∞) where δ_{m-1} is the unique positive zero of G
- **If B <= 1/(m-1)** all the zeros of P(z) are located in $\delta_{m-1} <= |z| <= 1$

This applies directly to our polynomial. Making m=T+1 and B=β we find:

- **If β > 1/T** P(z) has T zeros in $\delta_{m-1} < |z| < 1$ and one zero in (1, +∞) where δ_{m-1} is the unique positive zero of G
- **If β <= 1/T** all the zeros of P(z) are located in $\delta_{m-1} <= |z| <= 1$

THEOREM 3.9. *Let* $f_p(z) = z^m + a_p z^p - a_0$ *and* $g_p(z) = z^m - a_p z^p - a_0$. *The zeros of* $P_p(z) = z^m + a_p z^p + a_0$, $p = 1$ *or* $p = m - 1$, *under conditions* (1.10), *satisfy*:

(1) *If* $a_p \le -\frac{m}{p}$ (*or* $a_0 \ge \frac{m-p}{p}$), *all the zeros of* $P_1(z)$ *are located in* $1 \le |z| \le \gamma_1$, *where* γ_1 *is the unique positive zero of* $f_1(z)$. *If* $a_p > -\frac{m}{p}$ (*or* $a_0 < \frac{m-p}{p}$), $P_1(z)$ *has* $m-1$ *zeros in* $1 \le |z| \le \gamma_1$ *and one zero in* $(0, 1)$.

(2) *If* $a_p < -\frac{m}{p}$ (*or* $a_0 > \frac{m-p}{p}$), $P_{m-1}(z)$ *has* $m-1$ *zeros in* $\delta_{m-1} \le |z| \le 1$ *and one zero in the interval* $(1, +\infty)$, *where* δ_{m-1} *is the unique positive zero of* $g_{m-1}(z)$. *If* $a_p \ge -\frac{m}{p}$ (*or* $a_0 \le \frac{m-p}{p}$), *all the zeros of* $P_{m-1}(z)$ *are located in* $\gamma_{m-1} \le |z| \le 1$.

Since all the complex roots have | | <1, the **long term solution** is of the form:

$$N_t \simeq C_+ \lambda_+^{\ t}$$

which, due to the positive nature of the root, can be re-written:

$$N_t \simeq C_+ \left(e^A\right)^t = C_+ e^{At}$$

and conveniently expressed as:

$$N_t \simeq C_+ e^{\alpha(\beta T - 1)t}$$

This can only be applied to the more general expression:

$$N_{t+1} = N_t + N_t \underbrace{\beta_t \left(1 - \frac{M - m_t}{M}\right)}_{\beta'_t} - N_{t-T} \underbrace{\beta_{t-T} \left(1 - \frac{M - m_{t-T}}{M}\right)}_{\beta'_{t-T}}$$

Otherwise we will see growth or decay that is not exponential and the solution does not apply.

if we consider that the fraction of susceptible population **is large** and **varies slowly within an time inverval of size T** (early phase of a controlled epidemic)**.** In that case we can simply write:

$$N_{t+1} = N_t \left(1 + \beta'\right) - N_{t-T} \beta'$$

Therefore, we have:

$$N_t \simeq C_+ e^{\alpha(\beta' T - 1)t}$$

for intervals in which the time varying factors, including the base value of β, are changing slowly compared to the interval size. An equivalent expression, valid in such an small interval, would be:

$$N_t \simeq N_0 e^{\alpha(R-1)t}$$

For the new cases, we find, by substitution:

$$n_t = N_{t-1}\beta' \simeq \frac{R}{T} N_0 e^{\alpha(R-1)(t-1)} = \frac{R}{T} N_0 e^{-\alpha(R-1)} e^{\alpha(R-1)t}$$

$$n_t \simeq n_0 e^{\alpha(R-1)t}$$

* References:

http://www.m-hikari.com/ces/ces2017/ces5-8-2017/p/chooCES5-8-2017.pdf

https://scihubtw.tw/10.1080/0020717830893300

https://scihubtw.tw/10.1007/s10474-018-0896-6

Equations recap

General equation:

$$N_{t+1} = N_t + N_t \underbrace{\beta_t \left(1 - \frac{M - m_t}{M}\right)}_{\beta'_t} - N_{t-T} \underbrace{\beta_{t-T} \left(1 - \frac{M - m_{t-T}}{M}\right)}_{\beta'_{t-T}}$$

Simplification for small T, in terms of the time scale for changes in behavioral and environmental parameters:

$$N_{t+1} = N_t + N_t \underbrace{\beta_t \left(1 - \frac{M - m_t}{M}\right)}_{\beta'_t} - N_{t-T} \beta_t \left(1 - \frac{M - m_{t-T}}{M}\right)$$

Equations recap

Simplification for small T in terms of the time scale for changes in the fraction of susceptible population:

$$N_{t+1} = N_t + N_t \underbrace{\beta_t \left(1 - \frac{M - m_t}{M}\right)}_{\beta'_t} - N_{t-T} \underbrace{\beta_t \left(1 - \frac{M - m_t}{M}\right)}_{\beta'_t}$$

→ Epidemic behaviour with oscilations forced by behavioral changes in the population if β>1/T and exponential decay otherwise.

Simplification for constant behavioral and environmental parameters:

$$N_{t+1} = N_t + N_t \underbrace{\beta \left(1 - \frac{M - m_t}{M}\right)}_{\beta'_t} - N_{t-T} \underbrace{\beta \left(1 - \frac{M - m_{t-T}}{M}\right)}_{\beta'_t}$$

→ Epidemic curve with a peak if β>1/T and exponential decay otherwise. **Fundamental equation.**

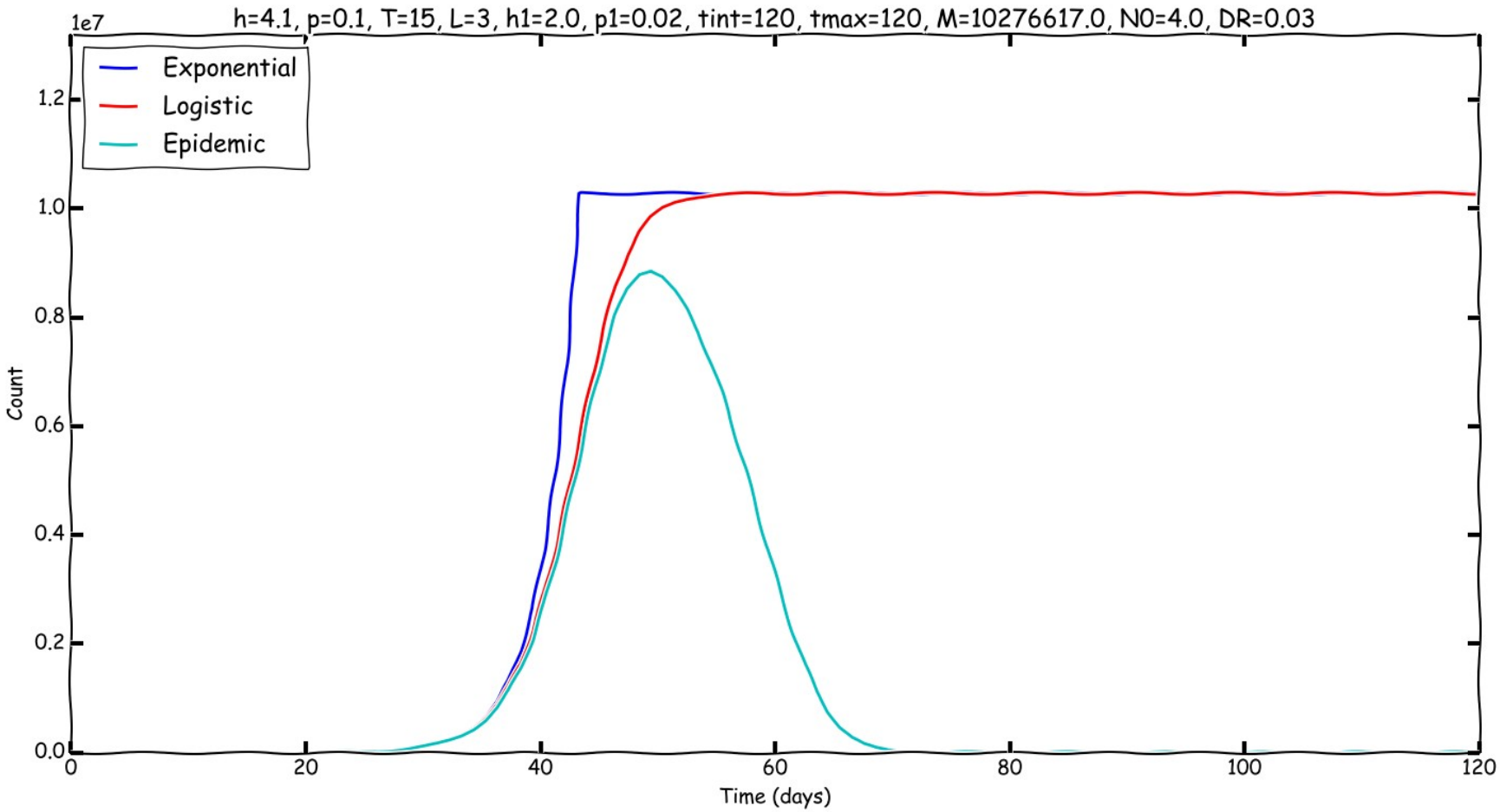
Simplification for a very early epidemic stage:

$$N_{t+1} = N_t + N_t \beta - N_{t-T} \beta$$

→ Pure exponential growth or decay, once the complex roots related solutions fade out.

Comparison of models

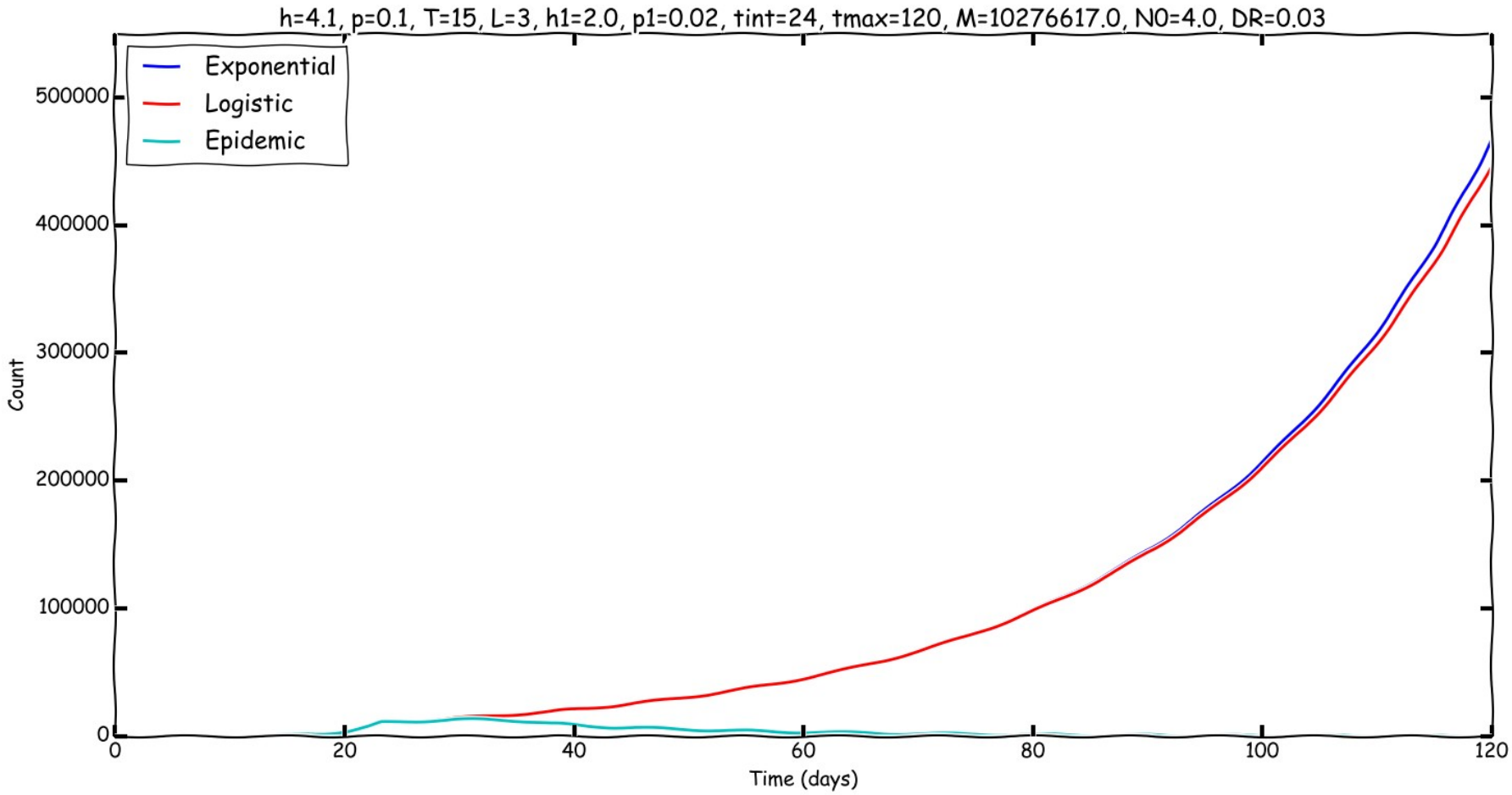
Active cases



https://github.com/ghomem/viraly

Change of parameters

Active cases with an imposed parameters change at t=24



4. Temporary infection with gaussian duration of average T and std dev L, finite population correction

$$N_{-X}=0$$

$$\dots$$

$$N_{-1}=0$$

Definition necessary for the general formula to hold

$$N_0=N_0$$

New cases (n) Outgoing cases (OC)

People infected at time 0 that will recover between 0 and 1

$$N_1=N_0+N_0\beta\left(1-\frac{M-m_0}{M}\right)-n_0F_0(0,1)$$

$$N_2=N_1+N_1\beta\left(1-\frac{M-m_1}{M}\right)-n_0F_0(1,2)-n_1F_1(1,2)$$

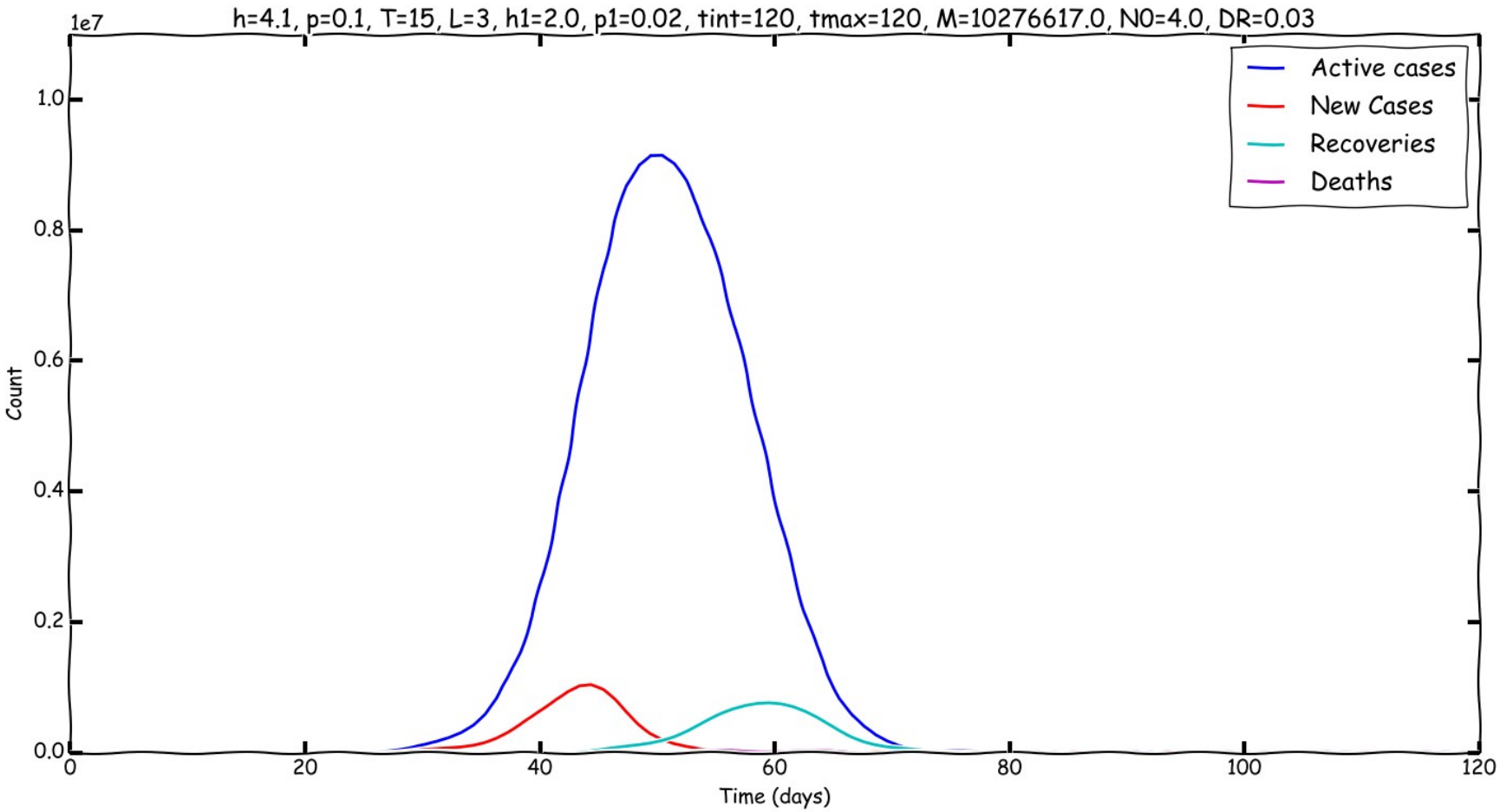
$$N_t=N_{t-1}\left(1+\beta\left(1-\frac{M-m_{t-1}}{M}\right)\right)-\sum_{j=0}^{t-1}n_jF_j(t-1,t)$$

with

$$F_j(t-1,t)=\int_{k-1}^{k}\frac{1}{L\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{j-T}{L\sqrt{2}}\right)^2}$$

People infected at time j that will recover between t-1 and t

Free epidemic situation



49 9158122.132366545

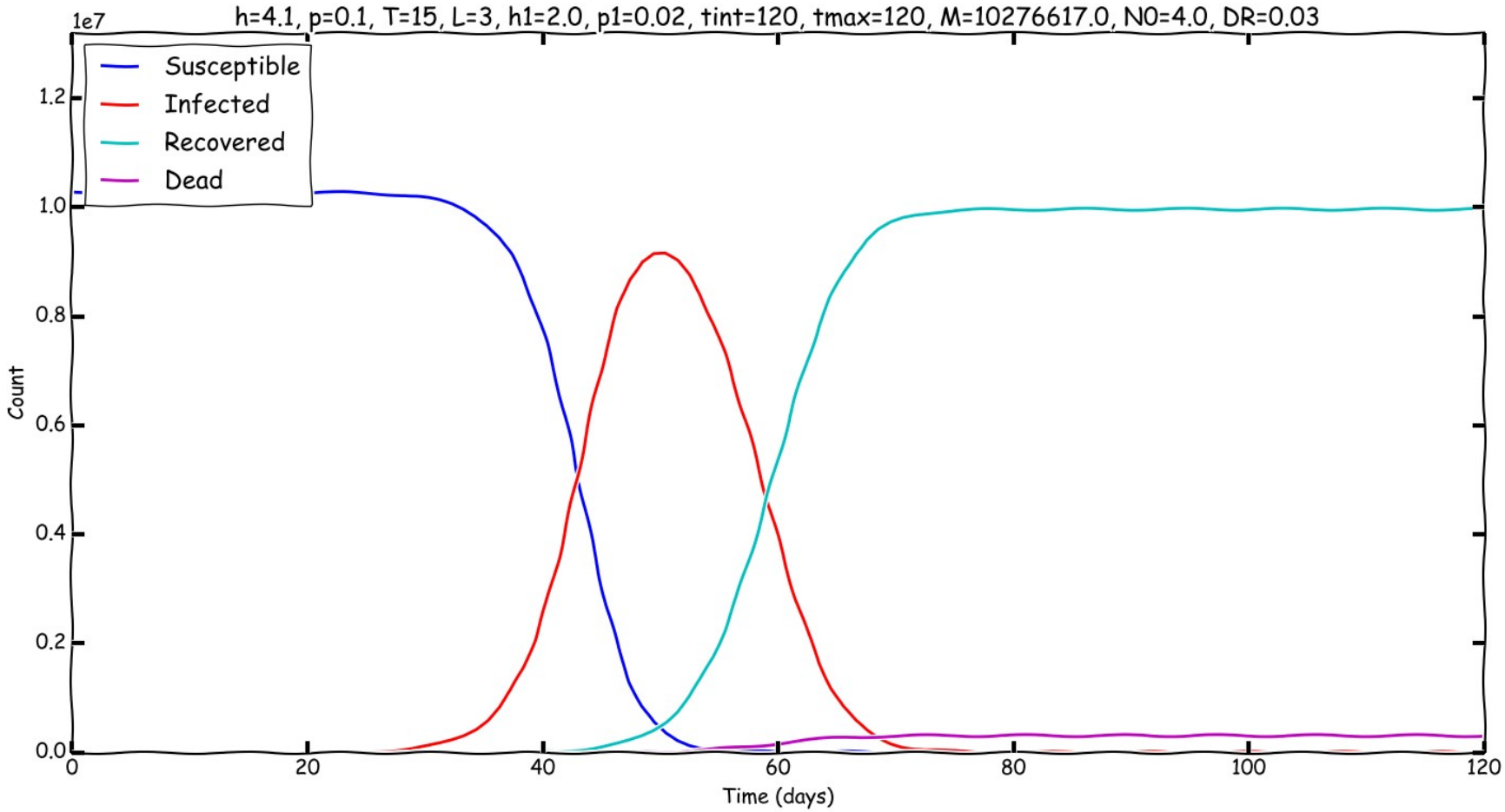
python3 viraly.py "4.10,0.1,15,3,2,0.02,120,120,10276617,4,0.03"

Totals:

transmissions, infections, recoveries, deaths

10270507.20087203 10270511.20087203 9962389.11813645 308115.1273650449

Free epidemic situation



49 9158122.132366545

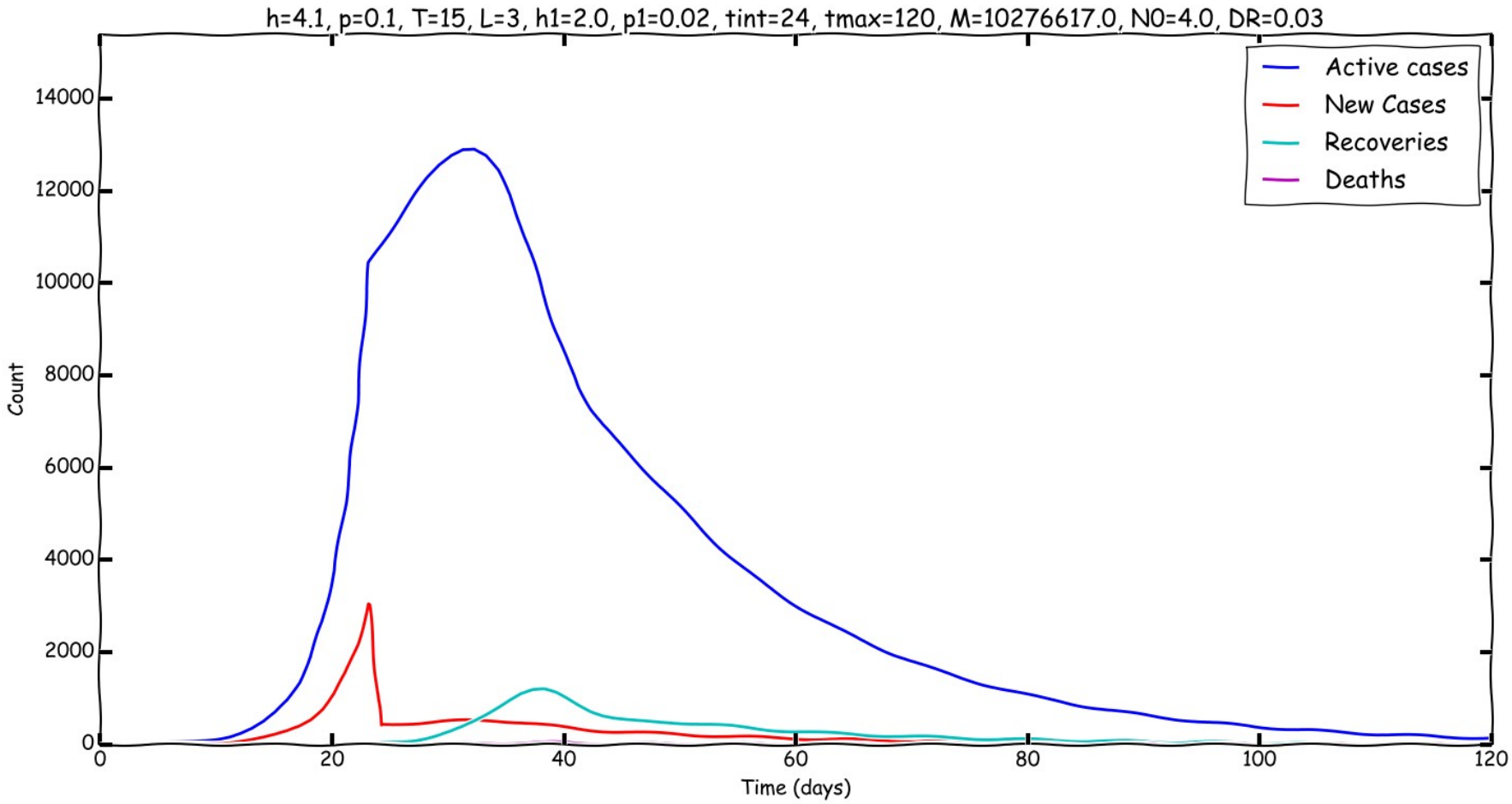
python3 viraly.py "4.10,0.1,15,3,2,0.02,120,120,10276617,4,0.03"

Totals:

transmissions, infections, recoveries, deaths

10270507.20087203 10270511.20087203 9962389.11813645 308115.1273650449

Containment starting at t=24



Maximum value

33 18161.206978689457 python3 viraly.py "4.10,0.1,15,3,2,0.02,24,120,10276617,4,0.03"

Totals:

transmissions, infections, recoveries, inactivations

34881.88671629874 34885.88671629874 33642.307338365674 1040.4837321144023

Model 3 – Fundamental equation

$$N_{t+1} = N_t + N_t \beta \left(1 - \frac{M - m_t}{M}\right) - N_{t-T} \beta \left(1 - \frac{M - m_{t-T}}{M}\right)$$

$$N_{t+1} = N_t + N_t \frac{\beta}{M} m_t - n_{t-T}$$

$$m_t = m_{t-1} - n_t$$

$$n_{t-T} = \frac{\beta}{M} N_{t-T-1} m_{t-T-1}$$

N_t = Active
m_t = Susceptible
n_{t-T} = New cases at T-t

$$N_{t+1} = N_t + N_t \frac{\beta}{M} m_t - O_t$$

$$m_t = m_{t-1} - O_t$$

$$O_t = \frac{\beta}{M} N_{t-T-1} m_{t-T-1}$$

N_t = Active
m_t = Susceptible
O_t = Outgoing

$$N_{t+1} - N_t = N_t \frac{\beta}{M} m_t - O_t$$

$$m_{t+1} - m_t = O_{t+1}$$

$$O_{t+1} = \frac{\beta}{M} N_{t-T} m_{t-T}$$

Model 3 - Rewriting in SIR notation

$$I_{-X} = 0$$

$$\dots$$

$$I_{-1} = 0$$

Definition necessary for the general formula to hold

$$I_0 = I_0$$

New cases (n) Outgoing cases (O)

$$I_1 = I_0 + I_0 \beta \left(1 - \frac{M - S_0}{M}\right) - n_{1-T}$$

$$I_2 = I_1 + I_1 \beta \left(1 - \frac{M - S_1}{M}\right) - n_{2-T}$$

$$I_t = I_{t-1}\left(1 + \beta\left(1 - \frac{M - S_{t-1}}{M}\right)\right) - n_{t-T}$$

$$S_t = S_{t-1} - n_t = S_{t-1} - I_{t-1}\beta\left(1 - \frac{M - S_{t-1}}{M}\right)$$

S_{t-1}

$$O_t = n_{t-T} = I_{t-T-1}\beta\left(1 - \frac{M - S_{t-T-1}}{M}\right)$$

S_{t-T-1}

I_t = Infected
S_t = Susceptible
O_t = Outgoing (Recovered or Dead)

(Avoiding the R letter to not cause notation confusions with R_t)

Model 3 - Rewriting in SIR notation II

$$I_{-X}=0$$

$$\dots$$

$$I_{-1}=0$$

Definition necessary for the general formula to hold

$$I_0=I_0$$

$$I_1=I_0+\frac{\beta}{M}S_0 I_0-V_1$$

$$S_0=M-I_0$$

$$O_1=I_{-T}\beta S_{-T}$$

$$I_2=I_1+\frac{\beta}{M}S_1 I_1-V_2$$

$$S_1=S_0-\frac{\beta}{M}S_0 I_0$$

$$O_2=I_{1-T}\beta S_{1-T}$$

$$\beta=hp$$

$$I_t=I_{t-1}+\frac{\beta}{M}S_{t-1}I_{t-1}-V_t$$

$$S_{t-1}=S_{t-2}-\frac{\beta}{M}S_{t-2}I_{t-2}$$

$$O_t=I_{t-T-1}\beta S_{t-T-1}$$

Model 3 - Rewriting in SIR notation III

$$I_{-X}=0$$

$$\dots$$

$$I_{-1}=0$$

$$I_0=I_0$$

Definition necessary for the general formula to hold

$$\begin{cases} S_{t-1}=S_{t-2}\left(1-\frac{\beta}{M}I_{t-2}\right) \\ I_k=I_{t-1}\left(1+\frac{\beta}{M}S_{t-1}\right)-O_t \\ O_t=\frac{\beta}{M}I_{t-T-1}S_{t-T-1} \end{cases} \rightarrow \begin{cases} S_t-S_{t-1}=-\frac{\beta}{M}S_{t-1}I_{t-1} \\ I_t-I_{t-1}=\frac{\beta}{M}I_{t-1}S_{t-1}-O_t \\ O_t=\frac{\beta}{M}I_{t-T-1}S_{t-T-1} \end{cases} \rightarrow \begin{cases} S_{t+1}=S_t-\frac{\beta}{M}I_tS_t \\ I_{t+1}=I_t+\frac{\beta}{M}I_tS_t-O_{t+1} \\ O_{t+1}=\frac{\beta}{M}I_{t-T}S_{t-T} \end{cases}$$

$$\beta=hp$$

I_t = Infected
S_t = Susceptible
O_t = Outgoing (Recovered or Dead)

Somewhat similar to the ODE based SIR model but here the removals are not differential but rather defined directly as a function of past I and S.

Model 3 - Rewriting in SIR notation III

$$\begin{cases} S_{t+1} = S_t - \frac{\beta}{M} I_t S_t \\ I_{t+1} = I_t + \frac{\beta}{M} I_t S_t - O_{t+1} \\ O_{t+1} = \frac{\beta}{M} I_{t-T} S_{t-T} \end{cases}$$

$$\begin{cases} S_{t+1} = S_t - \frac{\beta}{M} I_t S_t \\ I_{t+1} = I_t + \frac{\beta}{M} I_t S_t - \frac{\beta}{M} I_{t-T} S_{t-T} \end{cases}$$

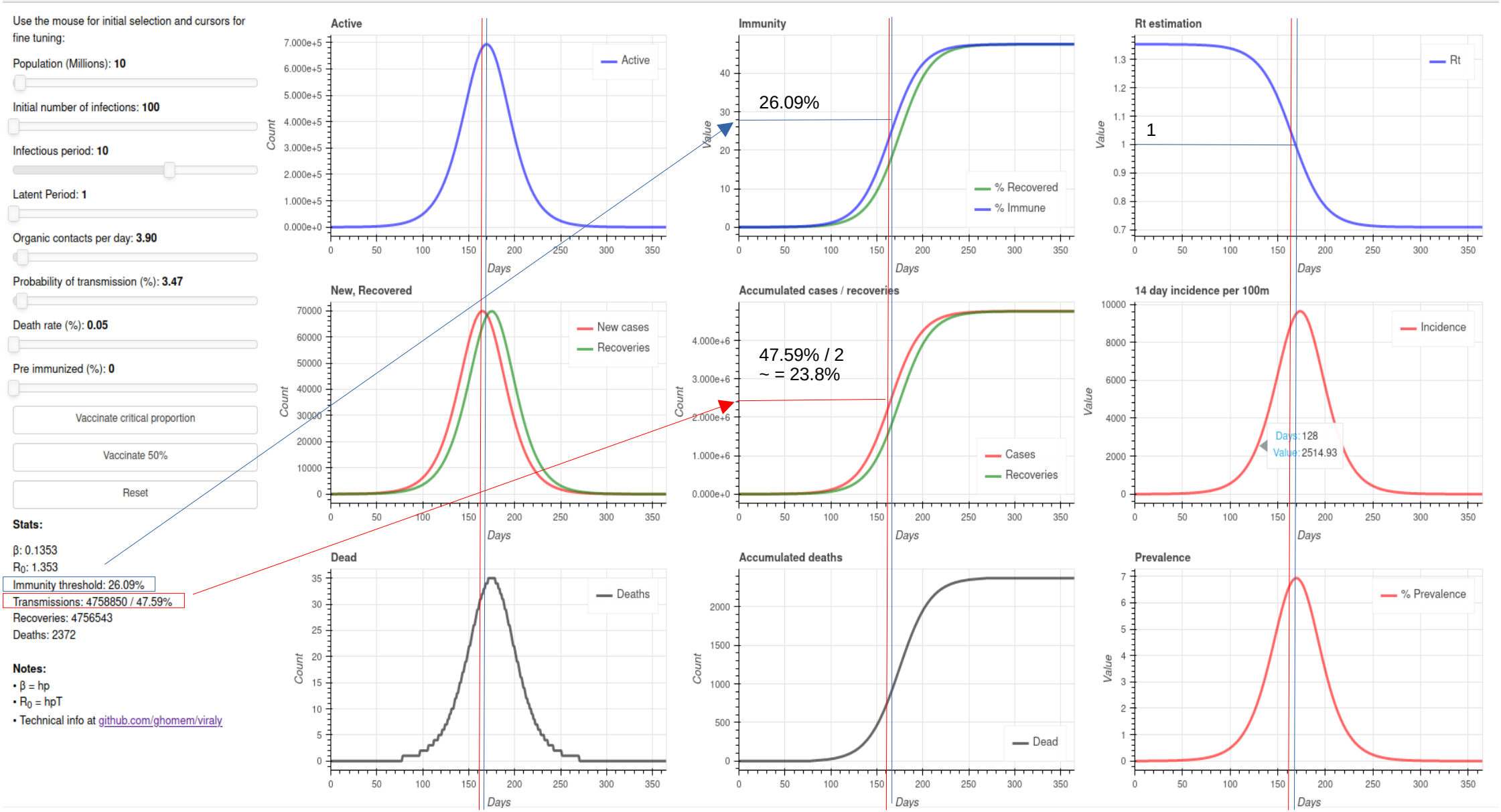
Compacted the last 2 equations

$$\begin{cases} S_{t+1} = S_t - \frac{\beta}{M} I_t S_t \\ I_{t+1} = I_t + \frac{\beta}{M} I_t S_t - \frac{\beta}{M} I_{t-T} S_{t-T} \\ V_{t+1} = V_t + \frac{\beta}{M} I_{t-T} S_{t-T} \end{cases}$$

Added accumulated removals but the other two equations do not depend on it.

$$\beta = hp$$

I_t = Infected
S_t = Susceptible
V_t = Removed (accumulated Recovered or Dead)



New cases peak at half of the total infection reach.
The curve of active cases peaks later, at the herd immunity threshold.
R_t crosses 1 at the peak of active cases.

Estimating R_t from epidemic data

In the presence of real epidemic data we can theoretically invert the propagation equation to estimate R_t:

$$N_{t+1} = N_t + N_t \underbrace{\beta_t \left(1 - \frac{M - m_t}{M}\right)}_{\beta'_t} - N_{t-T} \underbrace{\beta_{t-T} \left(1 - \frac{M - m_{t-T}}{M}\right)}_{\beta'_{t-T}}$$

$$N_{t+1} = N_t + N_t \beta'_t - N_{t-T} \beta'_{t-T}$$

Real data makes available, with varying accuracy, the number of new cases, whereas the number of active cases is usually not available (or it is too uncertain due to the logistics it depends on). We note that, under our simple model, the active cases at a certain moment are simply the sum of new cases of the last T days (today included):

$$N_t = \sum_{j=t-T+1}^{t} n_j$$

We also note that even though we have two different values of β' in the expressions above, one of them can be expressed as a function of the number of new cases:

$$N_{t+1} = N_t + N_t \beta_t' - N_{t-T} \beta' \longrightarrow N_t = N_{t-1} + N_{t-1} \beta'_t - \underbrace{N_{t-T-1} \beta'_{t-T}}_{\text{New cases at t-T}}$$

$$N_t = N_{t-1} + N_{t-1} \beta'_t - n_{t-T}$$

$$\frac{N_t}{N_{t-1}} = 1 + \beta'_t - \frac{n_{t-T}}{N_{t-1}} \longrightarrow \frac{N_t + n_{t-T}}{N_{t-1}} - 1 = \beta'_t \longrightarrow R_t = \left(\frac{N_t + n_{t-T}}{N_{t-1}} - 1\right) T$$

From

$$N_t = \sum_{j=t-T+1}^{t} n_j$$

R_t is, by definition, β_t divided by T.

It follows that:

$$R_t = \left(\frac{n_{t-T} + \sum_{j=t-T+1}^{t} n_j}{\sum_{j=t-T}^{t-1} n_j} - 1\right) T \longrightarrow R_t = \left(\frac{\sum_{j=t-T}^{t} n_j}{\sum_{j=t-T}^{t-1} n_j} - 1\right) T \longrightarrow R_t = \frac{n_t T}{\sum_{j=t-T}^{t-1} n_j}$$

R_t appears as the ratio between a hypothetical accumulated result from having T days with the current number of new cases and the actual accumulated result of the previous T days. It can also be expressed as:

$$R_t = \frac{n_t}{\bar{n}_{t-1}}$$

where the average number $\bar{n}_{t-1}$ could be called the **activity** over the previous T days.

Extrapolating trends from epidemic data

If we estimated R_t as a certain value R, in a particular moment, we can forecast the evolution of new cases in a scenario where it remains constant, by simply inverting the previous formula and calculating sequentially:

$$n_{t+1} = R\,\bar{n}_t$$

$$n_{t+2} = R\,\bar{n}_{t+1} \qquad \bar{n}_{t+1} = \bar{n}_t + n_{t+1} - n_{t-T} = \bar{n}_t\left(1+R\right) - n_{t-T}$$

$$n_{t+3} = R\,\bar{n}_{t+2} \qquad \bar{n}_{t+2} = \ldots = \bar{n}_{t+1}\left(1+R\right) - n_{t-T+1}$$

However, (positive) constant R means, by definition, that we are on a temporary situation where the growth of active cases and new cases is exponential and in such situation it is easy to extrapolate. For new cases we would have:

$$n_t \simeq n_0\, e^{\lambda t}$$

The parameter λ represents the relative daily variation and is easy to estimate by linear regression of the logarithm of new cases. The duplication time is given by:

$$T_{dup} = \frac{\ln\left(2\right)}{\lambda}$$

Extrapolating trends from epidemic data

For the particular situation of exponential growth, the value of R can be easily determined from the exponential parameter:

$$R = \frac{\lambda T}{1 - e^{-\lambda T}}$$

This formula holds because if

$$n_t = n_0 e^{\lambda t}$$

then

$$n_{t+1} = n_0 e^{\lambda (t+1)} = n_0 e^{\lambda t} e^{\lambda} = e^{\lambda} n_t$$

but also *

$$\bar{n_t} = \frac{1 - e^{-\lambda T}}{\lambda T} e^{\lambda} n_t$$

from which results:

$$n_{t+1} = R \bar{n_t} = R \frac{1 - e^{-\lambda T}}{\lambda T} e^{\lambda} n_t$$

Obtained by comparing the two expressions

$$R = \frac{\lambda T}{1 - e^{-\lambda T}}$$

* next slide

Extrapolating trends from epidemic data

The average number of cases during exponential growth can be estimated as:

$$n_t = n_0 e^{\lambda t} \longrightarrow \bar{n}_t = \frac{n_0}{T} \sum_{j=t-T+1}^{t} e^{\lambda j} \longrightarrow \bar{n}_t \simeq \frac{n_0}{T} \int_{t-T+1}^{t+1} e^{\lambda h} dh$$

$$\bar{n}_t \simeq \frac{n_0}{\lambda T} \left[e^{\lambda (t+1)} - e^{\lambda (t-T+1)} \right]$$

$$\bar{n}_t \simeq \frac{n_0}{\lambda T} e^{\lambda (t+1)} \left[1 - e^{-\lambda T} \right] = \frac{1 - e^{-\lambda T}}{\lambda T} e^{\lambda} n_t$$

Adjustment of the interval due to the sum to integral approximation Without this we would have 0 for T=1 and high error in general for small T. We are using the upper bound from

https://en.wikipedia.org/wiki/Summation#Approximation_by_definite_integrals

Extrapolating trends from epidemic data

Note: this expression contains an implicit definition of the exponential parameter as a function of R and T:

$$R = \frac{\lambda T}{1 - e^{-\lambda T}} \longrightarrow \lambda = \lambda(R, T)$$

Let's make k = λT

$$(k - R)\, e^{k-R} = -\frac{1}{e^R} R \longleftarrow \text{> -1/e}$$

$$(k - R) = W_0\left(-\left(\frac{1}{e^R}\right) R\right) \longleftarrow \text{Positive branch of the Lambert function}$$

https://en.wikipedia.org/wiki/Lambert_W_function

$$k = R + W_0\left(\frac{-R}{e^R}\right)$$

$$\lambda = \frac{1}{T}\left[R + W_0\left(\frac{-R}{e^R}\right)\right]$$

Exact solution

$$\lambda \simeq \frac{1}{T}\left[R - 1 + \sqrt{2}\sqrt{1 - \frac{R}{e^{R-1}}}\right]$$ For R ~ 1 (from *, equations 3 and 4)

$$\lambda \simeq \frac{1}{T}\left[2(R - 1)\right]$$ For R ~ 1 (with further simplifications)

$$\lambda \simeq \frac{1}{T} R\left(1 - \frac{1}{e^R}\right)$$ For large R (Taylor series for W_0, around x=0)

* https://link.springer.com/article/10.1007/s10444-017-9530-3

Extrapolating trends from epidemic data

Note on approximation for R ~ 1:

Asymptotic expansions are known at both the end-points of the branch. At the limit point $x = -1/e$, W goes as

$$W \approx -1 + \sqrt{2}y - \frac{2}{3}y^2 + \frac{11}{18\sqrt{2}}y^3 - \frac{43}{135}y^4 + \dots, \qquad (3)$$

where

$$y \equiv \sqrt{1 + ex}. \qquad (4)$$

$$W_0 \simeq -1+\sqrt{2}\sqrt{1-e\frac{R}{e^R}} = -1+\sqrt{2}\sqrt{1-\frac{R}{e^{R-1}}} \simeq -1+\sqrt{2}\sqrt{1-\frac{R}{1+R-1+\frac{(R-1)^2}{2}+\dots}}$$

$$W_0 \simeq -1+\sqrt{2}\sqrt{\frac{(R-1)^2}{2R+(R-1)^2}} = -1+(R-1)\sqrt{2}\sqrt{\underbrace{\frac{1}{2R+(R-1)^2}}_{\sim 2}} \simeq -1+(R-1)$$

$$\lambda \simeq \frac{1}{T}[R-1+(R-1)] = \frac{1}{T}[2(R-1)] \longrightarrow n_t \simeq n_0 e^{\frac{2}{T}(R-1)t} \quad \text{For R ~ 1}$$

* https://link.springer.com/article/10.1007/s10444-017-9530-3

Why is R_t interesting as a metric?

Some of the reasons include:

- unlike the variation of new cases R_t is independent of the population size
- unlike β, R_t can describe any viral growth independently from the underlying reality (because parameters h,p and T are all part of R_t)
- R_t allows for an intuitive interpretation:
  - the average number of secondary infections each infected agent generates

$$R_t \simeq \beta T \left(1 - \frac{M - m_t}{M}\right) = \beta' T$$

Furthermore:

The active cases per capita N_t/M along with R_t form a population and phenomenon independent (state, growth) pair of variables.

Caveat: this independence requires a reliable estimation of T.