Simulating Hawkes processes with an efficient thinning algorithm

Michele Veronesi
Politecnico di Torino
michele.veronesi@studenti.polito.it

Abstract—In this report, we discuss the implementation of a thinning algorithm for simulating simple, non-homogeneous Poisson point processes with independent increments of the time. Firstly, it will be used to simulate a Hawkes process in the first 100 days, without any intervention by governments. After that, we define a simple metric for the intervention factor in function of the number of deaths, thus we simulate the behavior of the epidemic process in a whole year.

I. ALGORITHM

A. Description

We use the thinning algorithm proposed in [1], and reported in algorithm 1, with few implementation variants, in order to make it run in a feasible time.

Algorithm 1 Thinning algorithm for simulating simple non-homogeneous Hawkes process

---

Require:

m: parameter of the Poisson r.v. for the number of infected for each sample.

λ: the parameter for the function h.

σ: the intensity for generating the ancestors.

η: the death rate of the disease.

δ: the constant factor for obtaining ρ, i.e., the intervention parameter.

k: number of days of simulation.

Ensure:

INF$[1 \ldots k]$ the number of infections for each day in $[1 \ldots k]$.

DEATH$[1 \ldots k]$ the number of deaths for each day.

INF$[1 \ldots k] \leftarrow \{0 \ldots 0\}$
$t \leftarrow 0$ //the time of the simulator
$\bar{\gamma} = \sigma(t)$
$T \leftarrow \{0\}$
while $t < k$ **and** $\bar{\gamma} > 0$ **do**
  $w \leftarrow Exp(\bar{\gamma})$ //exponential r.v.
  $t \leftarrow t + w$
  $\gamma = \sigma(t) + m/\rho_\delta(t) \cdot \sum_{\tau \in T} h_\lambda(t - \tau)$
  $u \leftarrow U(0, 1)$ //uniform r.v.
  if $\bar{\gamma} \cdot u < \gamma$ **and** $t < k$ **then**
    INF$[$int$(t)]$ += 1
    $T \leftarrow T \cup \{t\}$
  end if
  $\bar{\gamma} \leftarrow \gamma$
end while
DEATH $\leftarrow \eta \cdot$ INF
return INF, DEATH

B. Implementation

The first noteworthy approximation is the one performed in the computation of deaths values. It is done at the end of the simulation, since it is reasonable that an infected people spread the disease, and then die after a while (we can assume a month). We also know that, for each possible choice of the function h, after one month the contribution of a sample to the stochastic intensity of the process approaches zero.

Another important approximation is done during the simulation process: to significantly speed-up the execution, we do not store any set of active ancestors T, instead we rely only on the array INF to compute the value of h. This implies that we do not consider, in the computation of the stochastic intensity γ at time Δt, the infected sample on the same time unit Δt (with time unit we mean a day). The algorithm behaves in two different ways, depending on the two choices of h reported respectively in eq. (1) and eq. (2).

$$h_\lambda(t) = \lambda \cdot \mathbb{1}_{[t < 1/\lambda]} \tag{1}$$

$$h_\lambda(t) = \lambda \cdot \exp\{-\lambda \cdot t\} \tag{2}$$

In the first case, we take the sum of INF$[\max\{0, t - 1/\lambda\} \ldots$ int$(t)]$ and multiply it by the factor λ. As you can notice, this operation is $\Theta(1)$ since λ is a constant. Instead, in the second case, we apply the function h to INF$[0 \ldots$ int$(t)]$ using the API of the library NumPy, thus resulting in a computational time equivalent to $\Theta(1)$.

A further discussion should be made on the choice of the factor of intervention $\rho(t)$, which is defined in eq. (3). Note that, in the code, the interventions can be deactivated, therefore $\rho_\delta(t)$ will return always 1, independently of t.

$$\rho_\delta(t) = \begin{cases} \max\{1, \eta \cdot \text{INF}[\text{int}(t) - 1]/\delta\} & \text{if } t \geq 20 \\ 1 & \text{otherwise} \end{cases} \tag{3}$$

II. EXPERIMENTS

All the experiments are run on an AMD Ryzen 7 CPU with 16 hyper-threads, each one is repeated 15 times (in parallel) to provide appropriate confidence interval.

The first experiment is run with $k = 100$, $\lambda = 1/20$ when h is uniform and $\lambda = 1/10$ when h is exponential. The parameter for generating the ancestors is reported in eq. (4). The death rate η is equal to 0.02, while $m = 2$. The factor δ is ignored because we deactivate interventions. Results, with h either uniform or exponential, are reported respectively in fig. 1 and fig. 2. The experiments took around 1 minute in the first case, while in the second around 5 (keep in mind that these results strongly depend on the machine you run the experiment and how many times you want to repeat it).

$$\sigma(t) = 20 \cdot \mathbb{1}_{[t < 10]} \tag{4}$$

Secondly, assuming h_λ uniform with $\lambda = 1/20$, we tried to limit the spread of the infection dividing the m constant by a

factor ρ, computed as reported in eq. (3), i.e., in function of the number of deaths in the previous day. In order to be compliant with the constrain of maximum 20k deaths, we choose $\delta = 32$. To do so, we seek this hyper-parameter on a discrete interval $\delta \in [20, 50] \subset \mathbb{N}$, but running the experiment only on some values of interest (i.e., the ones reported in fig. 4). The results when choosing $\delta = 32$ are reported in fig. 3, while the cost of making this choice, defined as in eq. (5), and the number of deaths, are reported both in fig. 4. The total number of deaths is 19550, while the total cost Φ_{32} is 757.7. The completion of this experiment, in our settings, takes around 4 minutes. The function $\rho_{32}(t)$ is reported in fig. 5.

$$\Phi_\delta(t) = \int_0^t \rho_\delta^2(s)ds \qquad (5)$$

REFERENCES

[1] Y. Y. Chen, "Thinning algorithms for simulating point processes," 2016. [Online]. Available: https://www.math.fsu.edu/~ychen/research/Thinning%20algorithm.pdf
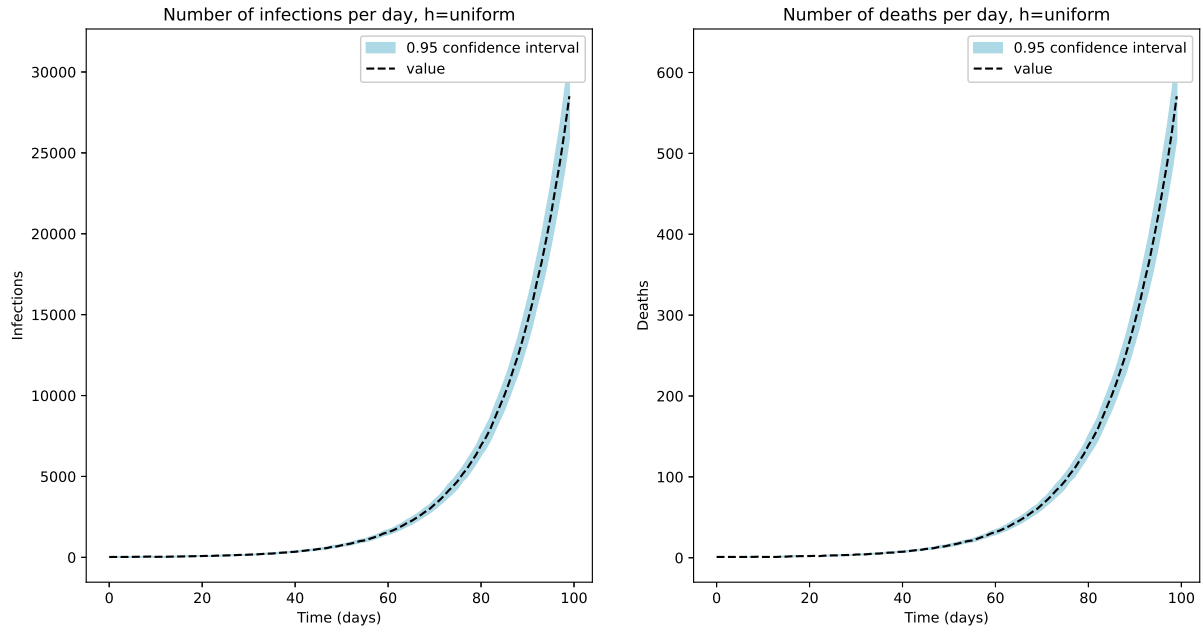
Fig. 1: Number of infections and deaths per day, with h uniform
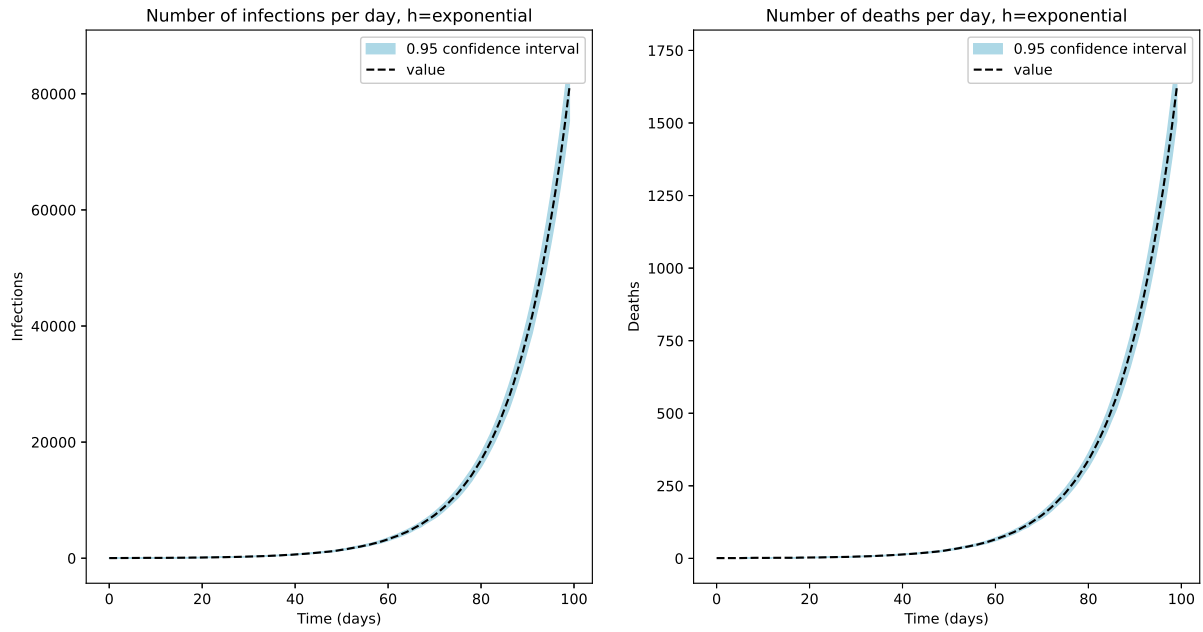

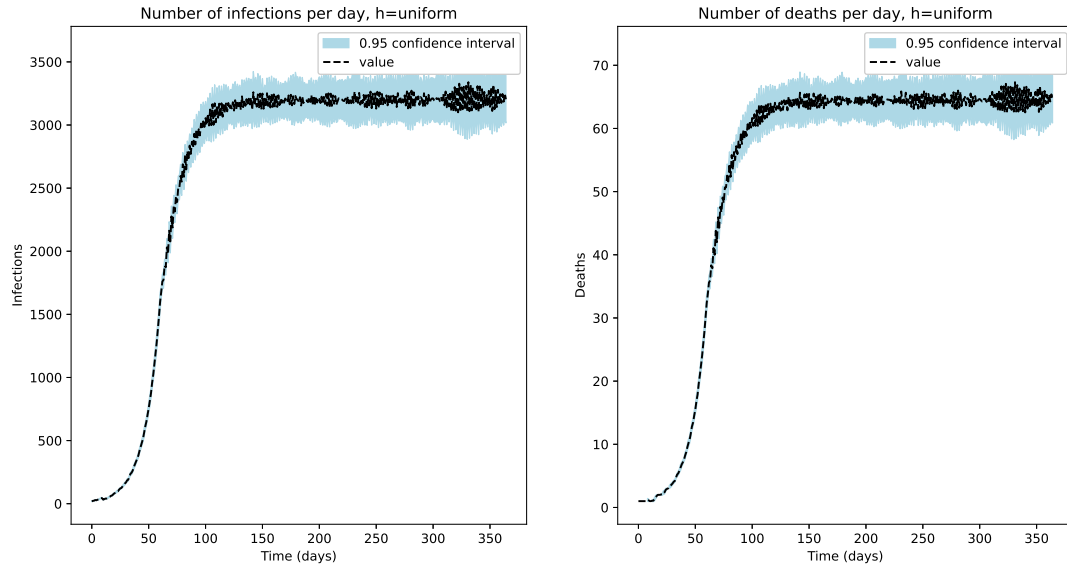Fig. 2: Number of infections and deaths per day, with h exponential

Fig. 3: Number of infections and deaths per day with intervention factor $\delta = 32$
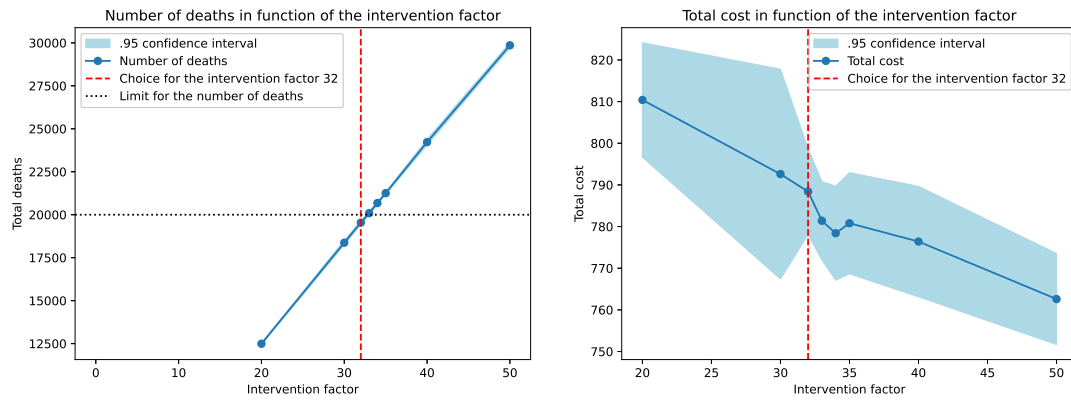

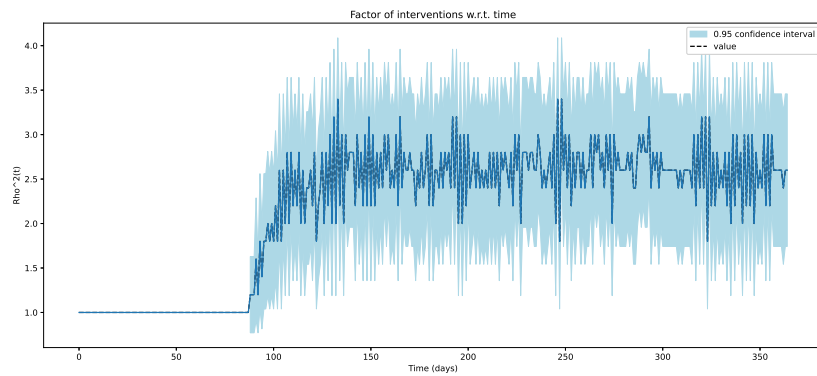Fig. 4: Results of the research of the factor of intervention δ

Fig. 5: Factor of intervention in function of the time, when choosing the intervention factor $\delta = 32$