Forecasting Model for the Spread of SARS-CoV-2

David Champredon

Department of Pathology and Laboratory Medicine
Western University (London, Ontario)
david.champredon@gmail.com

April 28, 2020

Warning: This is a draft documentation and not as clear as I would like it to be: do not hesitate to contact me for any questions or comments. The model described here is a constantly evolving work and is used to provide daily forecasts to Public Health Ontario. A link to a Github repo will be posted soon.

1 Data

The data used to fit the model are the confirmed (positive) COVID-19 cases that are considered a proxy for the true (unobserved) incidence of infections in the whole population. Moreover, to inform the generation interval distribution, I use data from line lists of COVID-19. My data source for COVID-19 reports in Canada is compiled and curated by Michael Li (https://github.com/wzmli/COVID19-Canada).

2 Model

2.1 Reported cases and true incidence

The number of positive cases reported at time t is P_t and the true, unobserved, incidence of infections is I_t. I assume a stochastic relationship between the positive reports and the true incidence:

$$I_t \sim \text{Gamma}\left(\text{mean} = \frac{P_{t+\ell}}{\rho}, \text{cv} = v_{obs}\right) \tag{1}$$

The probabilistic distribution is a convenient way to represent reporting errors and changes in the testing strategy. The shape parameter of this Gamma distribution is set such that the coefficient of variation v_{obs} has a pre-determined value. The parameter ℓ is a constant that represents the reporting delay.

2.2 Severity cascade

I assume the numbers of hospitalized cases at time H_t is directly proportional to the number of positive cases P_t. Similarly, the number of patients needing critical care (in ICUs with or without ventilators) C_t is proportional to the hospitalized cases, and finally the number of deaths D_t is proportional to the number of critically ill patients. Assuming again a stochastic relationship:

$$H_t \sim \text{Poisson}(h\, P_t) \tag{2}$$
$$C_t \sim \text{Poisson}(k\, H_t) \tag{3}$$
$$D_t \sim \text{Poisson}(d\, C_t\, e^{\varepsilon t}) \tag{4}$$

The term $e^{\varepsilon t}$ is an adjustment for the different growth rate observed between death and reported cases. The parameter ε is fitted on data.

2.3 Transmission process

I use a probabilistic representation of the renewal equation to model the disease transmission process:

$$I_t \sim \text{Gamma}(m_t, \phi) \tag{5}$$

with ϕ a dispersion parameter and m the mean incidence defined as

$$m_t = \left(\frac{S_t}{N}\right)^{1+\alpha} \mathcal{R}_0\, B_t \sum_{j\geq 1} I_{t-j} g(j) \tag{6}$$

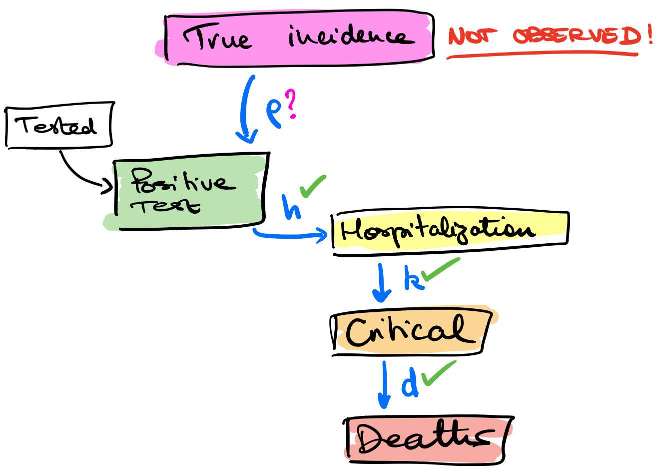

Figure 1: Model of the severity cascade of COVID-19 cases

where g is the intrinsic generation interval distribution; α is a parameter modelling the mixing heterogeneity ($\alpha = 0$ means homogeneous mixing; the larger α, the more heterogeneous the implicit contact structure); N is the effective population size; $\mathcal{R}_0$ the basic reproduction number; and B_t a function that models behavioural change at time t.

The function B_t can take various shapes in order to represent different intervention scenarios.

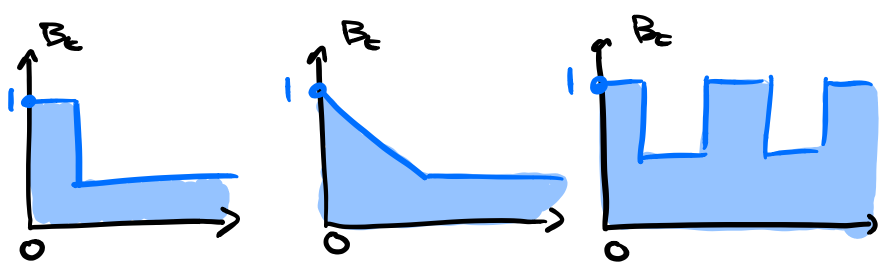

Figure 2: Examples of possible shape for B_t to represent interventions.

3 Fitting

3.1 Exponential growth phase

The generation interval is informed by available line lists of COVID-19 cases worl-wide.

The basic reproduction number $\mathcal{R}_0$ is fitted on the positive tests reported using the Dushoff and Park Gamma moment approximation:

$$\mathcal{R}_0 = (1 + \kappa r \bar{g})^{1/\kappa} \tag{7}$$

where r is the exponential growth rate of positive cases reported, $\bar{g}$ is the mean generation interval and κ a dispersion parameter informed by the observed variance of the generation interval.

The growth rate r is estimated with a linear regression on the log number of positive tests. The confidence interval are used to propagate uncertainty in $\mathcal{R}_0$ and consequently in the transmission process via the renewal equation Equation 6

The parameters for the severity cascade are directly observed from the daily epidemiological reports.

3.2 Deceleration phase

When the incidence count start to deviates from an exponential growth, the approximation defined by Equation 7 is not valid anymore.

In this "deceleration" phase, the model parameters are fitted using an approximate bayesian computation method (ABC). More specifically, the parameters $\mathcal{R}_0$, α and potentially B_t are fitted jointly on the incidence curve. For example, to assess the effect of social distancing in Ontario, I take B_t as a Heaviside step function with a discontinuity at the time of the declaration of the state of emergency (March 17, 2020) and a step size of $1 - \lambda$ (so λ is the effect of the intervention).

The distribution prior used are listed in Table 1.

The fit on Ontario data shows that $\mathcal{R}_0$ and λ are well identified by a unimodal posterior distribution but the posterior distribution of the heterogeneity parameter α is, unsurprisingly, not informed by the data.

4 Model Parameters

Table 1:

Parameter	Prior	Comments
$\mathcal{R}_0$	Normal(2, 0.7)	weak prior
α	Uniform(3, 5)	strong prior: contact structure highly heterogeneous
λ	Uniform(0, 1)	weak prior

Parameter	**Value**	**Comments**
Basic reprod. num. $\mathcal{R}_0$	fitted	Equation 7 or ABC
Contact heterogeneity α	fitted	arbitrary strong prior
Intervention effect λ	fitted	weak prior
Mean generation interval	4.5 days	from various studies
Variance of generation interval	7 days^2	from various studies
Susceptible proportion	0.8	assumption
Prop. of true incidence reported positive	Beta(20, 60)	CMMID
Prop. of positive to hospitalized h	0.14	Daily report from PHO; adj. daily
Prop. of hospitalized to critical k	0.39	Daily report from PHO; adj. daily
Prop. of critical to death d	0.60	Daily report from PHO; adj. daily
Coeff. of variation obs. process v_{obs}	0.2	assumption
Dispersion of transmission process ϕ	0.7	assumption