Modeling the Covid-19 pandemic

Dani Gamerman

- Modelling the Covid19 pandemic

Modelling the Covid19 pandemic

Dani Gamerman - Graduate Program in Statistics - UFMG

1st semester 2020

(inspired on notes by José Marcos Andrade Figueiredo - UFMG)

Basic logistic growth model

$$Y(t) \sim N(\mu(t), \sigma^2), \qquad t = 1, 2, \ldots$$

where $Y(t)$ is the **cumulated number of confirmed cases** by day t in a given region, with $\mu(t) = \frac{a \exp\{ct\}}{1 + b \exp\{ct\}}$.

Special case:

- $b = 0$ (exponential growth) $\rightarrow \mu(t) = a \exp\{ct\}$;
- adequate for early stages of the pandemic.

Problems of the basic model:

a. data are **counts**, and the normal distribution assumes continuous data;
b. variance should increase with data magnitude.

Characteristics of interest

The most important characteristics are:

1) Infection rate

- c measures the acceleration of growth and reflects the infection rate of the disease.

2) Assintote

$\lim_{t \to \infty} \mu(t) = \lim_{t \to \infty} \frac{a \exp\{ct\}}{1 + b \exp\{ct\}} = \frac{a}{b}$

- Reflects the total number of cases accumulated throughout the whole trajectory of the pandemic.
- Exponential growth ($b = 0$): assintote $= \infty$!

3) Peak of the pandemic

- Defined as the time t^* where number of new cases stops growing and starts to decrease.
- Exponential growth ($b = 0$): number of new cases never stops growing!

4) Prediction

- What can be said about $Y(t+k), \forall k$, for t fixed (today)?

  It depends on the distribution of $Y(t)$ but will always be given by the predictive distribution of $Y(t+k)$ given $Y(1:t) = \{Y(1), \ldots, Y(t)\}$ - what was observed.

  It works as the posterior distribution of $Y(t+k)$.

Useful result: If Z and W are any 2 r. v.'s then:

- $E[Z] = E[E(Z \mid W)]$
- $Var[Z] = Var[E(Z \mid W)] + E[Var(Z \mid W)]$

In particular, $E[Y(t+k) \mid Y(1:t)] = E\{E[Y(t+k) \mid \mu(1:t)] \mid Y(1:t)\} = E[\mu(t+k)] \mid Y(1:t)]$, the posterior mean of $\mu(t+k)$.

Inference about all that was described above should be reported through point estimators (eg: posterior means), along with respective credibility intervals.

5) Reproducibility rate R_0

R_0 is the expected number of secondary cases of a disease caused by an infected individual.

At time t, it is defined as $R_0 = \frac{\mu(t)-\mu(t-1)}{\mu(t-1)} = \frac{\mu(t)}{\mu(t-1)} - 1$.

- **Beginning of the pandemic:** $1 \gg b \exp\{ct\} \rightarrow \mu(t) \approx a \exp\{ct\} \rightarrow R_0 \approx e^c - 1$
- **End of the pandemic:** $1 \ll b \exp\{ct\} \rightarrow \mu(t) \approx a/b \rightarrow R_0 \approx 0$
- **Middle of the pandemic:** R_0 is a function of parameters (a, b, c) and time t, and is given by $R_0(t) = e^c \, \frac{1+be^c e^{ct}}{1+be^{ct}} \; - \; 1$.

For any fixed t, one can obtain its posterior distribution (via MCMC sample) and calculate mean, quantiles and credibility intervals.

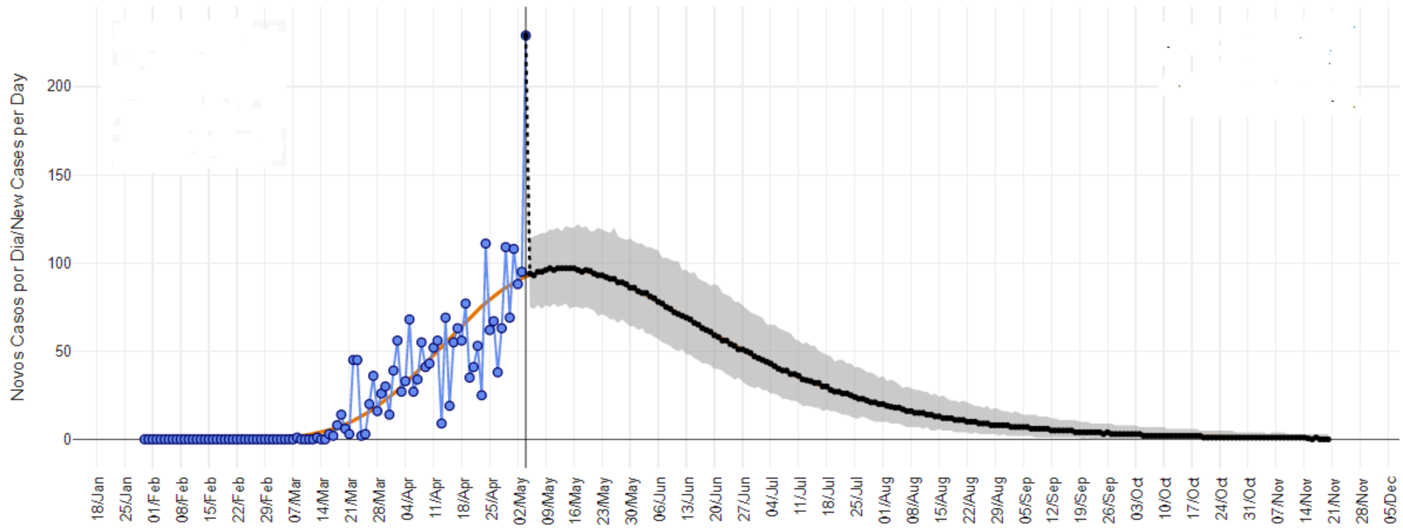
6) Mean number of new cases (MNNC)

Mean number of new cases at time $t+k$:

$$n_t(k) = E[Y(t+k) - Y(t+k-1)] = \mu(t+k) - \mu(t+k-1)$$

Thus, MNNC is also a function of parameters (a, b, c) and can be easily calculated.

For any fixed t and k, one can obtain its posterior distribution (via MCMC sample) and calculate mean, quantiles and credibility intervals.



Alternatives:

1.1) $Y(t) \sim Poisson(\mu(t))$ with $E[Y(t)] = \mu(t)$ and $Var(Y(t)) = \mu(t)$

1.2) $Y(t) \sim N(\mu(t), \sigma^2\ \mu(t))$ with $E[Y(t)] = \mu(t)$ and $Var(Y(t)) = \sigma^2\ \mu(t)$

Observações:

- Model (1.2) admits overdispersion if $\sigma^2 > 1$
- Alternative (1.2) only handles comment (b)
- Alternative (1.1) handles the two comments but does not allow overdispersion

Poisson with overdispersion

1.3) $Y(t) \mid \epsilon(t) \sim Poisson(\mu(t) + \epsilon(t))$ with $E[\epsilon(t)] = 0$ and $Var(\epsilon(t)) = \sigma^2$

1.4) $Y(t) \mid \epsilon(t) \sim Poisson(\mu(t) \times \epsilon(t))$ with $E[\epsilon(t)] = 1$ and $Var(\epsilon(t)) = \sigma^2$

Considering the usefull results presented above:

Mod(1.3):

- $E[Y(t)] = E[E(Y(t) \mid \epsilon(t))] = E[\mu(t) + \epsilon(t)] = \mu(t) + E[\epsilon(t)] = \mu(t)$
- $Var[Y(t)] = Var[E(Y(t) \mid \epsilon(t))] + E[Var(Y(t) \mid \epsilon(t))] = Var[\mu(t) + \epsilon(t)] + E[\mu(t) + \epsilon(t)] = \sigma$

Mod(1.4):

- $E[Y(t)] = E[E(Y(t) \mid \epsilon(t))] = E[\mu(t) \times \epsilon(t)] = \mu(t) \times E[\epsilon(t)] = \mu(t)$
- $Var[Y(t)] = Var[E(Y(t) \mid \epsilon(t))] + E[Var(Y(t) \mid \epsilon(t))] = Var[\mu(t) \times \epsilon(t)] + E[\mu(t) \times \epsilon(t)] = \mu$

Both preserve Poisson mean but increase Poisson dispersion.

Dynamic extensions

Previous models assume static behaviour:

- shape of the disease does not modify along time;

- infection rate will always be the same, assintote will always be the same, …

Dynamic models make it flexible.

1. **Dynamic models**

$\mu(t) = \frac{a(t)\ \exp\{c(t)\ t\}}{1+b(t)\ \exp\{c(t)\ t\}}$

with: $a(t) = a(t-1) + w_a(t)$, where $w_a(t) \sim N(0, W_a), \forall t$.

$b(t) = b(t-1) + w_b(t)$, where $w_b(t) \sim N(0, W_b), \forall t$.

$c(t) = c(t-1) + w_c(t)$, where $w_c(t) \sim N(0, W_c), \forall t$.

Advantages:

a. $E[a(t) \mid a(t-1)] = a(t-1)$, and the same goes for $b(t)$ and $c(t) \Rightarrow$ local constancy.

b. $Var[a(t) \mid a(t-1)] = W_a$, and the same goes for $b(t)$ and $c(t) \Rightarrow$ increase in uncertainty.

Problems:

a. variances W_a, W_b, W_c unknown $\Rightarrow$ difficult to specify;

b. variances W_a, W_b, W_c unknown $\Rightarrow$ difficult to estimate.

c. it os not possible to simplify $W_a = W_b = W_c = W$ (different magnitudes of (a, b, c)).

2. **Multiplicative effect:**

Another form to introduce dynamics, now **multiplicative**:

$a(t) = a(t-1) \times w_a(t)$, where $w_a(t) \sim Gamma(d_a, d_a), \forall t$.

$b(t) = b(t-1) \times w_b(t)$, where $w_b(t) \sim Gamma(d_b, d_b), \forall t$.

$c(t) = c(t-1) \times w_c(t)$, where $w_c(t) \sim Gamma(d_c, d_c), \forall t$.

Advantages:

a. $E[a(t) \mid a(t-1)] = a(t-1)$ and the same goes for $b(t)$ and $c(t) \Rightarrow$ local constancy.

b. $Var[a(t) \mid a(t-1)] = d_c^{-1}$ and the same goes for $b(t)$ and $c(t) \Rightarrow$ increase in uncertainty.

c. Hiperparameters d_a, d_b, d_c easier to specify.

Examples: $d = 1000 \ \rightarrow\ 0,90 = P(0,95 < w(t) < 1,05) = P\left(0,95 < \frac{a(t)}{a(t-1)} < 1,05\right)$

$d = 1500 \ \rightarrow\ 0,95 = P(0,95 < w(t) < 1,05) = P\left(0,95 < \frac{a(t)}{a(t-1)} < 1,05\right)$

Disadvantages:

a. Magnitudes of a, b, c still interfere in the increase in uncertainty.

b. Not sure if free software works fine with Gammas with such high parameter values.

3. **Multiplicative evolution with normal errors**

Consider the multiplicative evolution below for parameter a:

$$a(t) = a(t-1) \times \exp\{w_a(t)\}, \text{ where } w_a(t) \sim N(0, W_a)$$

Taking logarithm on both sides, one obtains:

$$\log\ a(t) = \log\ a(t-1) + w_a(t), \text{ where } w_a(t) \sim N(0, W_a)$$

Passing $\log\ a(t-1)$ to the left, one obtains:

$$\log\ a(t) - \log\ a(t-1) = \log\left[\frac{a(t)}{a(t-1)}\right] = w_a(t), \text{ where } w_a(t) \sim N(0, W_a)$$

Specification of W_a: one can think of percentual increase, as before.

$$0,95 = P\left(0,95 < \frac{a(t)}{a(t-1)} < 1,05\right) = P(-0,05 < w_a(t) < 0,05)$$

This implies $2\sqrt{W_a} = 0,05$, that implies $\sqrt{W_a} = 0,025 \ \Rightarrow W_a = (0,025)^2$.

The same specification is valid for W_b and W_c, since magnitudes of b and c do not matter.

- **Special case**

Based on Gamerman, Santos and Franco (J. Time Series Analysis, 2013):

$$\mu(t) = \frac{a(t)\ \exp\{c\,t\}}{1 + b\ \exp\{c\,t\}}$$

$a(t) = a(t-1) \times w_a(t)$, where $w_a(t) \sim Beta, \forall t$.

It may also be used for exponencial growth($b = 0$).

Advantage:

a. Allows exact calculation, thus avoiding (MCMC) approximations.

Disadvantage:

a. Does not allow dynamic b and c.

Generalizations of the logistic curve

So far, logistic curve was used to specify the mean $\mu(t)$ as $\mu(t) = \frac{a \exp\{ct\}}{1+b \exp\{ct\}} = \frac{a}{b+\exp\{-ct\}}$.

This expression is the simplest logistic form. It can be generalized in many ways. One possible form of the **generalized logistic** is

$$\mu(t) = d + \frac{a-d}{(b + \exp\{-ct\})^f}$$

The logistic curve is obtained by taking $d = 0$ and $f = 1$.