A Bayesian approach to COVID-19 fatality rates

Luca Alberto Rizzo

Paris Machine Learning Meetup - 01/05/2020

Numberly

My LinkedIn profile

Introduction

The main scope of this project is:

- to do outreach
- to inspire domain experts using alternative statistical approaches

Github

Estimate of quarantine duration

Bayesian analysis of fatality rates

Estimate of the number of infected population

Bayesian analysis of the fatality rate

Why do we see such an impressive variety among fatality rate estimates across countries?

$$t = \frac{\text{num of casualties among people who contracted COVID-19}}{\text{num of people who contracted COVID-19}}$$

Frequentist approach → one estimate per country independent from other countries

Bayesian approach
- → one estimate per country dependent on other countries via a prior
- → offers an elegant way to estimate the "global fatality rate"

Bayesian analysis of the fatality rate

N_c^i Cumulative cases per country

N_d^i Cumulative casualties per country

$$P\left(t^i = \hat{t}^i \mid N_c^i, N_d^i\right) = \frac{P\left(N_c^i, N_d^i \mid t^i = \hat{t}^i\right) P\left(t^i\right)}{\int P(y)\, P\left(N_c^i, N_d^i \mid y\right) dy}$$

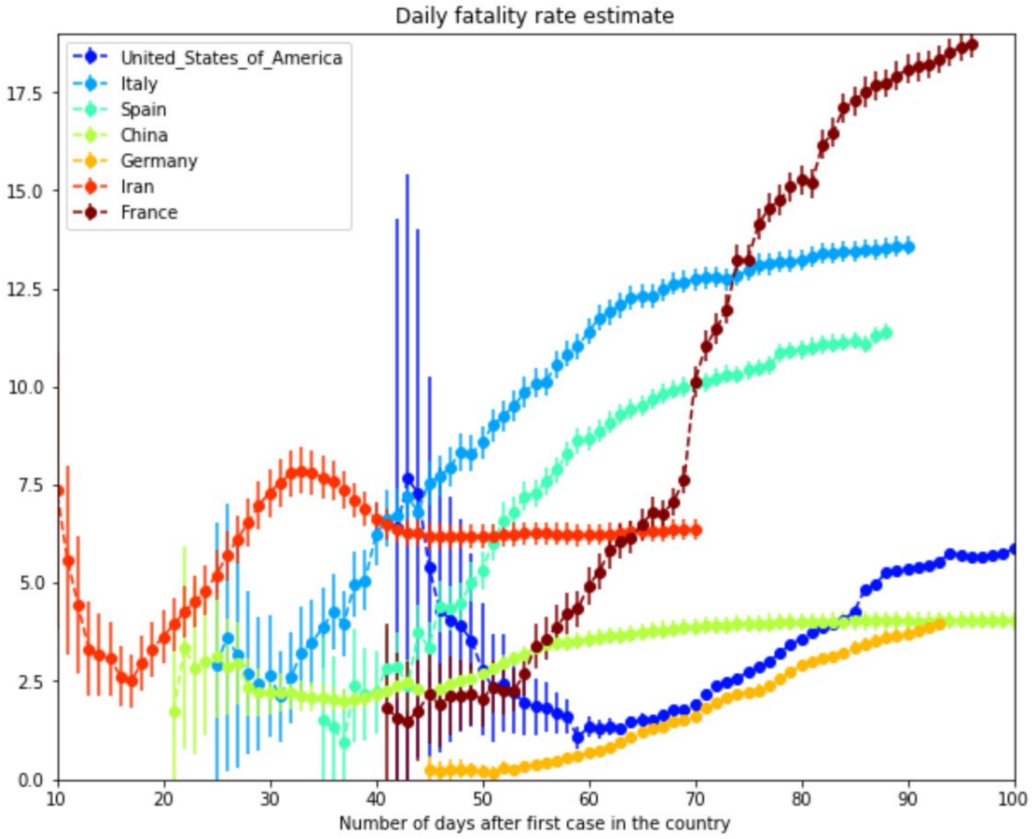
Bernoulli trial over cases and casualties

$$P\left(N_c^i, N_d^i \mid t^i = \hat{t}^i\right) = \binom{N_d^i}{N_c^i} t^{N_d^i} (1-t)^{N_c^i - N_d^i}.$$

With uniform prior :

$$P\left(t_C^i = \hat{t}_C \mid N_d^i, N_c^i\right) = \frac{t_C^{N_d^i}(1-t_C)^{N_c^i - N_d^i}}{B\left(N_d^i + 1, N_c^i - N_d^i + 1\right)}$$

$$= Beta\left(t_C, N_d^i + 1, N_c^i - N_d^i + 1\right),$$

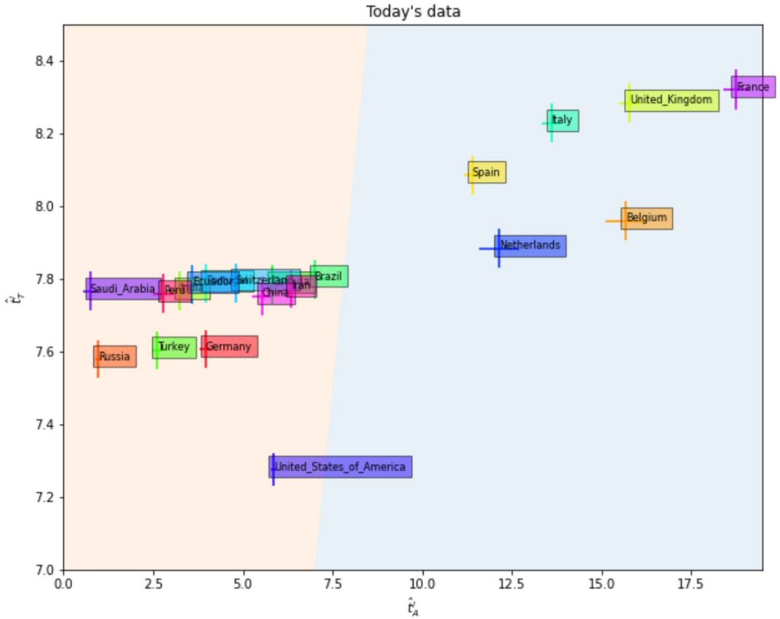


Bayesian analysis of the fatality rate

N_c^T Global cumulative cases N_d^T Global cumulative casualties

If we choose a Beta prior: $P(t^i) = Beta(t^i; N_d^T + 1, N_c^T + N_d^T + 1)$

$$\mathrm{P}\left(t_T^i = \hat{t}_T^i | N_d^T + N_d^i, N_c^T + N_c^i\right) = Beta\left(t_T, N_d^i + N_d^T + 1, \left(N_c^T + N_c^i\right) - \left(N_d^T + N_d^i\right) + 1\right)$$



Bayesian analysis of the fatality rate

Can we estimate the “global fatality rate” and its distribution?

Hierarchical Bayesian model for fatality rate

$$N_d^i \sim \mathrm{Bin}(N_c^i, t^i) \longleftarrow t^i \sim Beta(\alpha_d, \beta_d) \longleftarrow \alpha_d, \beta_d$$

Can we estimate the “global fatality rate” and its distribution?

$$p(\alpha_d, \beta_d, t_d \mid y_d) \propto p(y_d \mid t_d) p(t_d \mid \alpha_d, \beta_d) p(\alpha_d, \beta_d)$$

Weakly informative prior

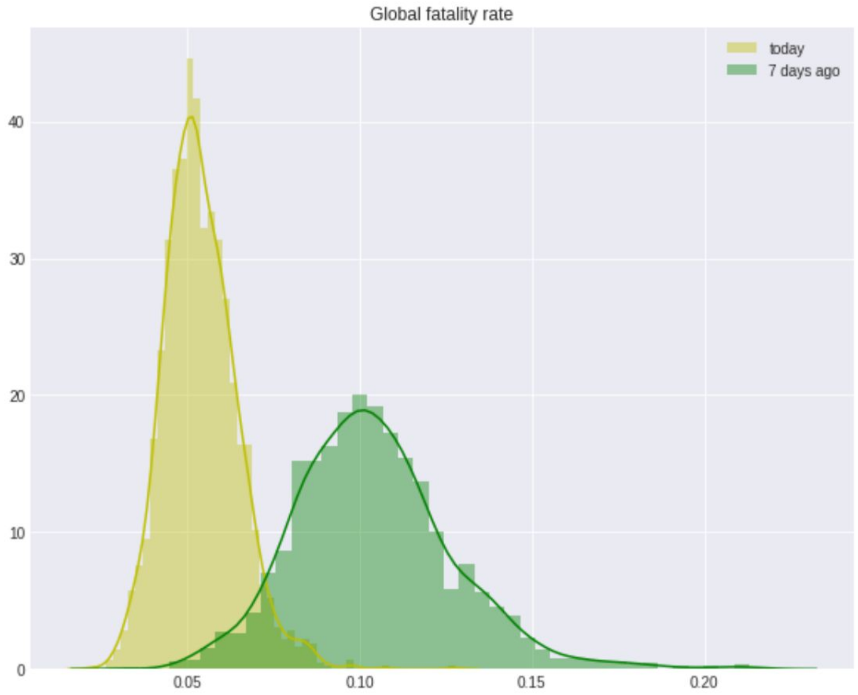
$$p(\alpha_d, \beta_d) \propto (\alpha_d + \beta_d)^{-5/2}$$

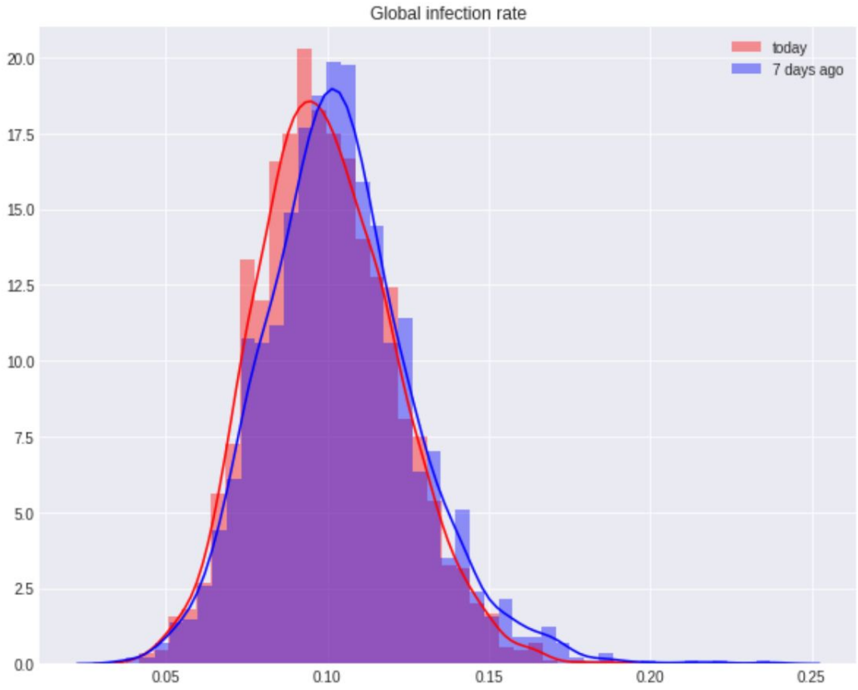
Bayesian analysis of the fatality rate

Can we estimate the “global fatality rate” and its distribution?

Hierarchical Bayesian model for fatality rate

Numerical estimate of hyperparameters with [PymC3]() : 1000 samples

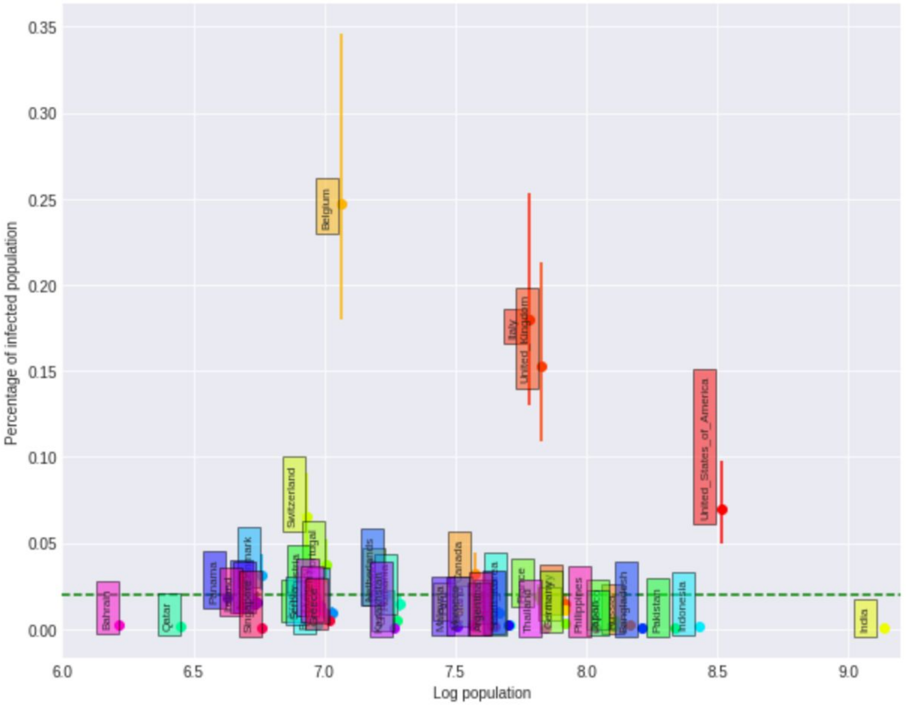


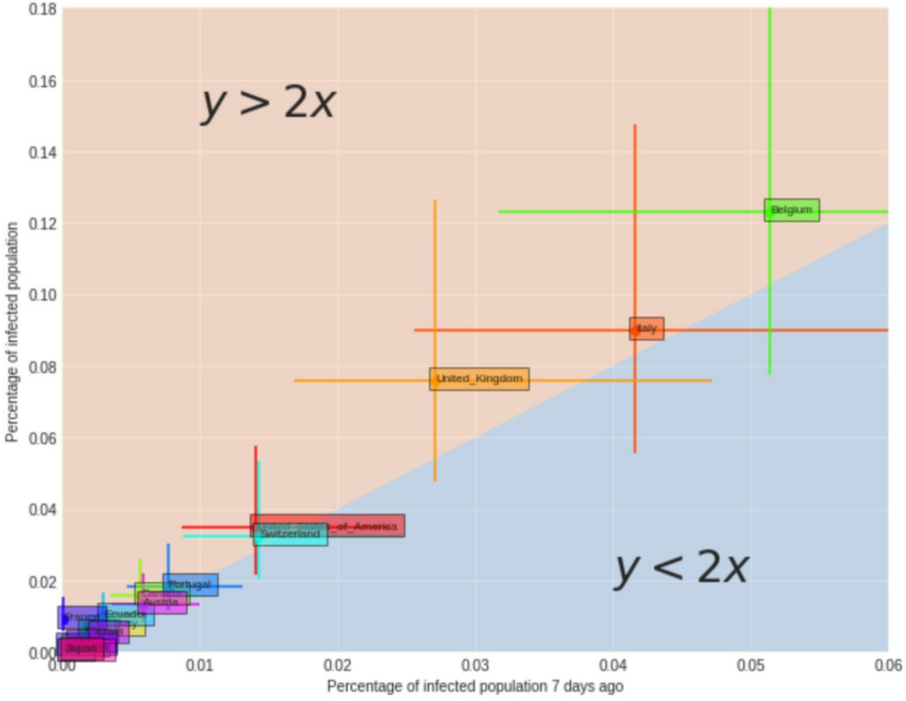


Bayesian analysis of the fatality rate

Estimate the “total number of infected people” = Number of people needed to generate all casualties

$$N_d^i \sim \mathrm{Bin}(N_e^i, t_e^i) \longleftarrow t_e^i \sim Beta(\alpha_d, \beta_d) \times Beta(\alpha_v, \beta_v)$$





Conclusions

An approach to studying fatality (and infection) rates using Bayesian Hierarchical Models

My educated guess is that most of the variability is explained by different phases in epidemic

Big variety in the total percentage of infected people among countries (1 to 20 %)

In the last 7 days the percentage of infected people has nearly doubled for the most impacted countries

Every predictive model must take into account
how fast parameters vary

Further studies

How to explain the “residual variation” among fatality rates?

Dynamical model to forecast the total number of infected population

Fatality rates for different groups of patients

...

Thanks for your attention

Please contact me if you have any comment and/or suggestion

Contact information

[My LinkedIn profile]

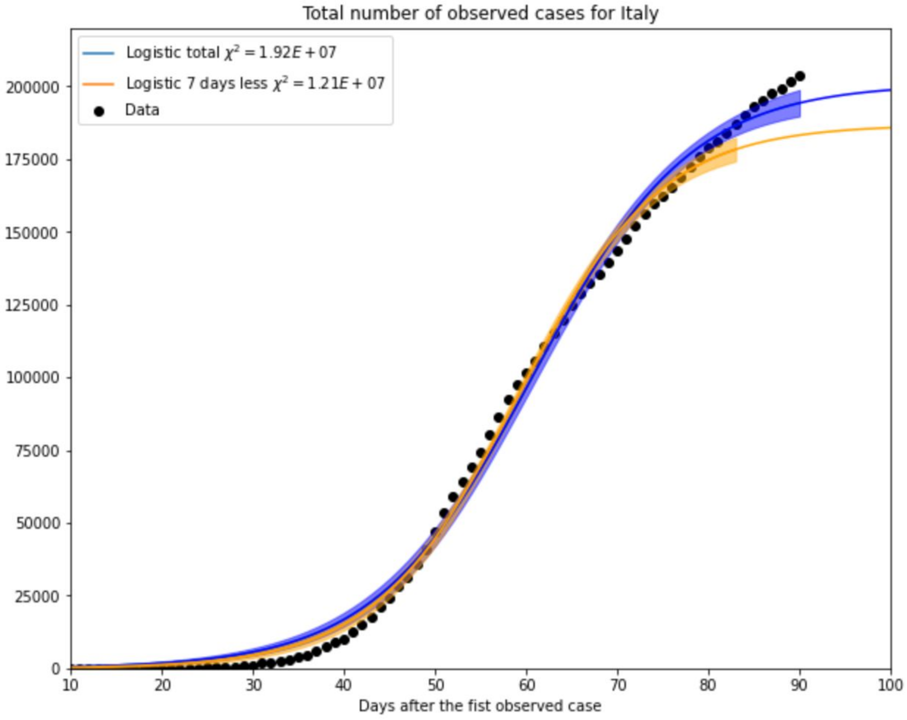
[Github]

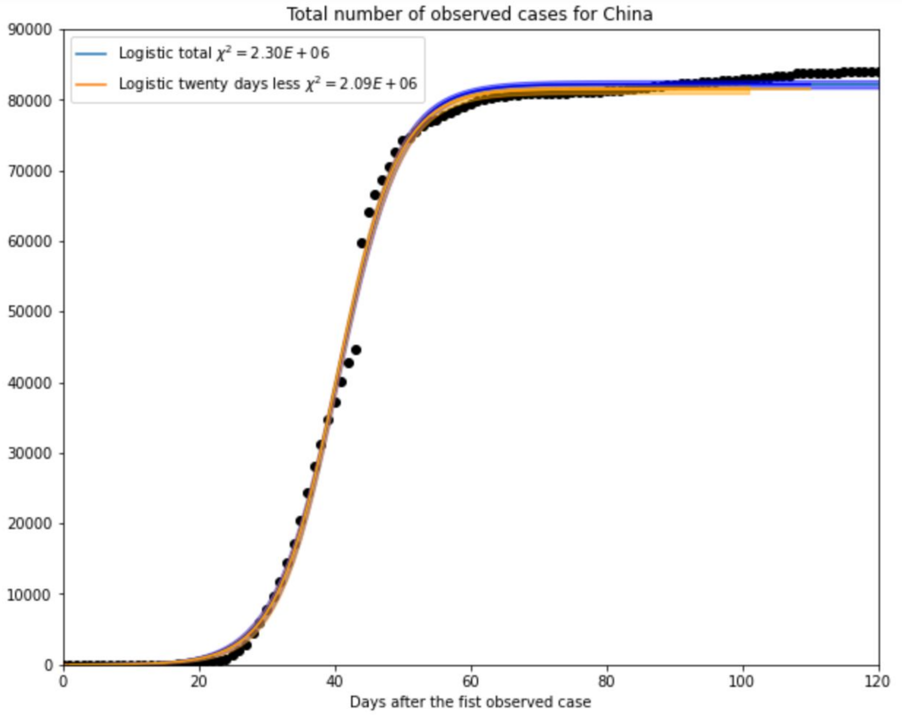
Data sources

[ECDPC]

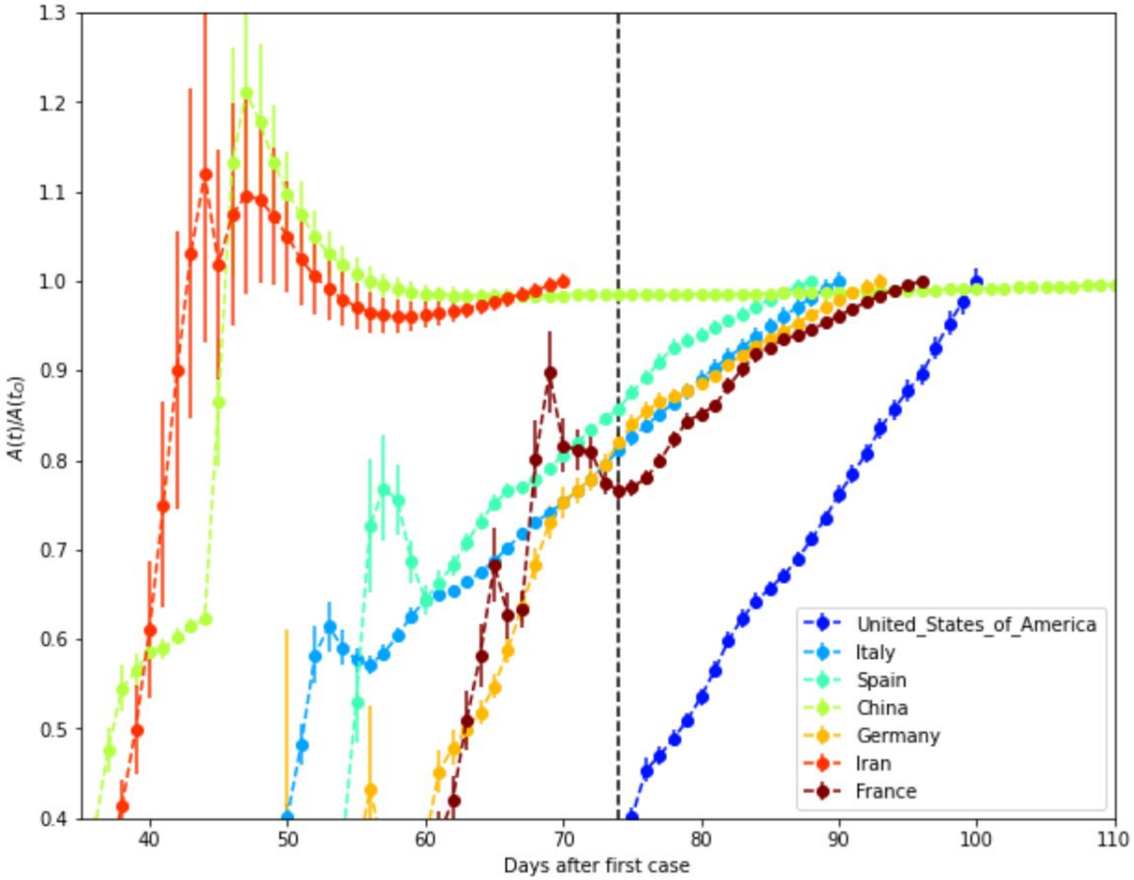
[Our World in Data]

Backup 1 : time stability of fits





Backup 2 : quarantine duration



Backup 3 : overestimate ratio