Random time transformation analysis of Covid19 2020

Isaac Meilijson
Tel Aviv University

Seminario en la Universidad de Chile

18 de Junio de 2020

Joint work with Nitay Alon

The SIR epidemiological model - Kermack & McKendrick (1927)

Susceptible cases $S(t) = K - X(t)$
Affected cases $X(t)$
Asymptote K, maximal X (perhaps population size N)
Removed cases $R(t)$, dead or recovered
Infected cases $I(t) = S(t) - R(t)$

$$\begin{aligned} dX(t) &= \beta(t)g(I(t))(1 - X(t)/K)dt \\ dR(t) &= \gamma I(t)dt \\ I(t) &= X(t) - R(t) \end{aligned}$$

where $K = N$ and $g(x) = x$.

Basic reproduction number

$R_0 = \frac{\beta(t)}{\gamma}$

Passage below 1 indicates transition from increasing to decreasing $I(\cdot)$

Covid19 2020 SIR may accept almost exact solution with β (besides γ) constant, provided $g(x) = x^{\alpha}$ for some $\alpha < 1$ (Bjørnstad, Finkenstädt and Grenfell 2002) and K free parameter.

Exponential or asymptotically linear growth?

$\alpha = 1$ and $\alpha < 1$ differentiate between exponential and sub-exponential growth for infinite population.

$\alpha = 1$ equations: exponential growth. $dI(t) = (\beta - \gamma)I(t)dt$ derived from

$$\begin{aligned} dX(t) &= \beta(X(t) - R(t))dt \\ dR(t) &= \gamma(X(t) - R(t))dt \\ (dI(t) &= (\beta - \gamma)I(t)dt) \end{aligned} \tag{1}$$

$$I(t) = I(0)\exp\{(\beta - \gamma)t\}$$
$$X(t) = X(0)\exp\{(\beta - \gamma)t\}\ ,\ \ R(t) = R(0)\exp\{(\beta - \gamma)t\}$$

$\alpha < 1$ equations: linear growth. $dI(t) = (\beta I(t)^\alpha - \gamma I(t))dt$ derived from

$$\begin{aligned} dX(t) &= \beta(X(t) - R(t))^\alpha dt \\ dR(t) &= \gamma(X(t) - R(t))dt \end{aligned}$$

admits constant solution $I(t) \equiv I_0 = (\frac{\beta}{\gamma})^{\frac{1}{1-\alpha}}$ with

$$X(t) \approx X(0) + \frac{\beta^{\frac{1}{1-\alpha}}}{\gamma^{\frac{\alpha}{1-\alpha}}} t \ ; \ R(t) \approx R(0) + \frac{\beta^{\frac{1}{1-\alpha}}}{\gamma^{\frac{\alpha}{1-\alpha}}} t$$

But population size is finite, and target sub-population may be small.

If $g(x) = x^\alpha$ with $\alpha < 1$ and K fitted with α, β, γ, solutions invariably show typical Covid19 behavior of empirical data, number of infected cases $I(t)$ that increase to a maximum and then decrease, towards zero hopefully.

Profile likelihood function of alpha (Murphy and Van Der Vaart 2000) generally quite flat, α somewhat noisily estimated. But definitely in $(0, 1)$.

End of May 2020 most countries and the world as a whole displayed $I(\cdot)$ increasing with time, making estimation more speculative than for countries where the basic reproduction number R_0 crossed already below 1.

Quite popular to welcome R_0 transition, wrongly referred to as leaving behind exponential growth in favor of a period of recovery.

Transition occurs when number of infected cases is maximal, where care should be perhaps strengthened rather than relaxed.

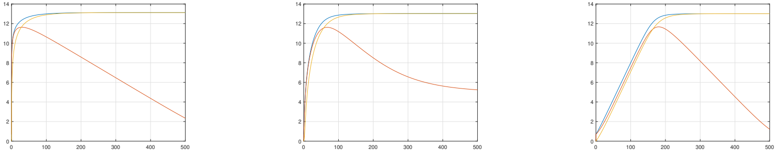

Figure: Log X, I, R for $\gamma = 0.05, K = .5M$. $\alpha = 0, \beta = 10000$. $\alpha = 0.75, \beta = 2$. $\alpha = 1, \beta = 0.125$

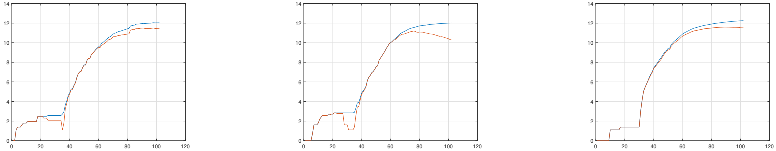

Figure: Logarithm of X, I, March to June 2020. L: France. M: Germany. R: Italy

Preliminary data handling - empirical data X, I, R and β

New removed cases proportional to number of infected cases.

R should be proportional to cumulative sum of I.

Linear regression with slope γ and zero intercept should manifest this relationship.

Empirically measured X kept intact but its split into R and I modified minimally so that the regression relation will hold.

B is $n \times n$ matrix with 0 above diagonal and 1 on and below diagonal.
A is $2n \times n$ matrix with γB in first n rows and γB plus identity matrix in last n rows.
V is column vector with R in top half and X in bottom half.

Regression equation $A\hat{I} \approx V$ gives "regression coefficients" $\hat{I}$ that are a compromise to manifest the requirement.

$$\begin{pmatrix} \gamma & 0 & 0 & \cdots & \cdots & 0 \\ \gamma & \gamma & 0 & \cdots & \cdots & \vdots \\ \vdots & \vdots & \ddots & \cdots & \cdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \cdots & \vdots \\ \vdots & \vdots & \vdots & \cdots & \ddots & \cdots \\ \gamma & \cdots & \cdots & \cdots & \cdots & \gamma \end{pmatrix} \qquad \left.\begin{pmatrix} \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \end{pmatrix}\right\} \hat{I} \approx \begin{pmatrix} \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \vdots \end{pmatrix} \begin{matrix} \left.\begin{matrix} \\ \\ \\ \end{matrix}\right\} R \\ \left.\begin{matrix} \\ \\ \\ \end{matrix}\right\} X \end{matrix}$$

$$\begin{pmatrix} \gamma+1 & 0 & 0 & \cdots & 0 \\ \gamma & \gamma+1 & 0 & \cdots & \vdots \\ \vdots & \vdots & \ddots & \cdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \cdots & \cdots \\ \gamma & \cdots & \cdots & \cdots & \gamma+1 \end{pmatrix}$$

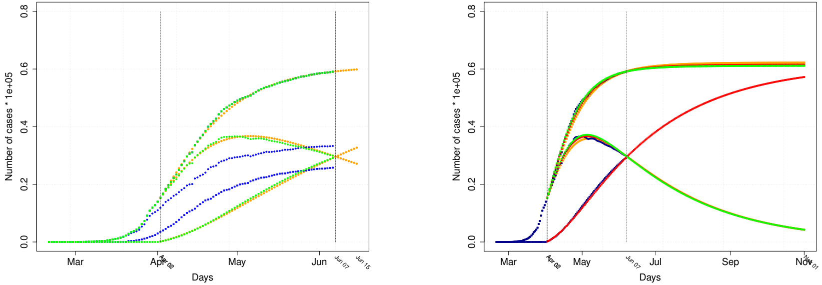

Figure: SIR model, data pre-proc and RTT solution, Belgium, March-June 2020

Regression pre-processing as tool for reverse engineering. Belgium recorded as affected cases also those under doubt. It holds as if record high deaths/M , and infected cases still growing on June 7th. Pre-processing is in sharp disagreement with the raw data, placing Belgium in the same standard stage as its neighbors, well past the transition to R_0 below 1, and suggesting lethality not above standard European levels.

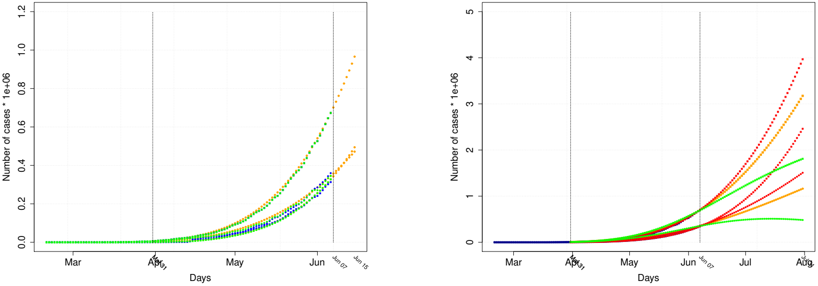

Figure: SIR model, data pre-processing and RTT solution, Brazil, March to June 2020

Data of Brazil show agreement, but still at early stage in the epidemic.

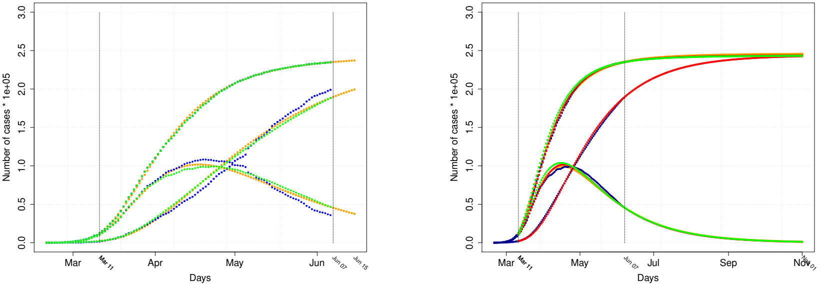

Figure: SIR model, data pre-processing and RTT solution, Italy, March to June 2020

The data of Italy and USA show close agreement between raw and pre-processed versions, with the pre-processed version smoothing relatively sharp jumps in the raw data.

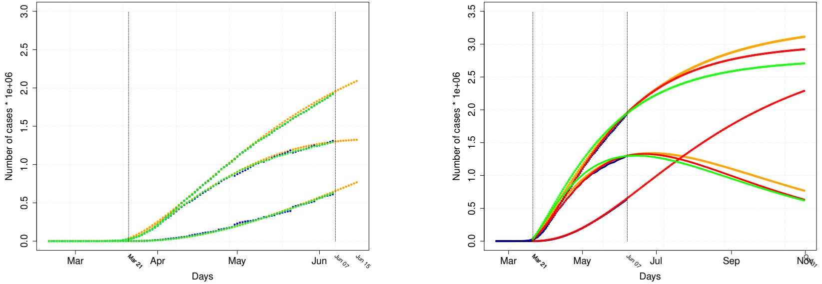

Figure: SIR model, data pre-processing and RTT solution, USA, March to June 2020

RTT method to solve differential equations with data subject to noise - Bassan, Marcus, Meilijson and Talpaz (1997)

Relatively novel theoretical contribution, RTT method to mimic ODE by stochastic counterparts, simpler but related to the SDE based on Diffusion processes via Fokker-Planck. Based on Skew-product transformations in Ergodic theory.
Diffusion methods place noise vertically, RTT method adopts the solution to the ODE, but considers it as evaluated at a random time process that advances on the average like chronological time. Errors are horizontal.

Empirical data $(X_1, R_1), (X_2, R_2), \ldots, (X_n, R_n)$. $Y_j = X_j - R_j$ inferred and regression-modified. X and R are assumed or forced to be non-decreasing.

No attempt to solve SIR analytically. Instead, time increment $\delta = \frac{1}{M}$ set (say, $M = 100$), and ODE is solved numerically as a difference equation.

Fix β, γ, K and the function g, initiate functions x and r as X_1 and R_1, initiate i as $X_1 - R_1$ and proceed with the definition for $j \geq 2$

$$\begin{aligned} x(j) &= x(j-1) + \beta g(i(j-1))(1 - \frac{x(j-1)}{K})\delta \\ r(j) &= r(j-1) + \gamma i(j-1)\delta \\ i(j) &= x(j) - r(j) \end{aligned}$$

Define the *random time trajectory* as starting at $T_1(1) = 1$, $T_2(1) = 1$.

For $m \geq 2$, let $T_1(m)$ be the smallest $\frac{j}{M}$ for which $x(j) \geq X_j$ and let $T_2(m)$ be the smallest $\frac{j}{M}$ for which $r(j) \geq R_j$.

Better yet, let $T_1(m)$ and $T_2(m)$ be the linear interpolants.

Solve for β and γ so that $T_1(n) = T_2(n) = n$. Incremental time has average 1.

$\Delta_1(m) = T_1(m+1) - T_1(m)$ and $\Delta_2(m) = T_2(m+1) - T_2(m)$ are the (mean-1) increments of the T_1 and T_2 processes.

Likelihood function induced by RTT

View $(\Delta_1(m), \Delta_2(m))$ as observations from a bivariate mean-1 Gaussian distribution, with empirical covariance matrix Σ. Up to a multiplicative constant, the normal density evaluated at these data is $(\det(\Sigma))^{-\frac{n-1}{2}}$.

Alternatively, view $T_1(m) - m$ and $T_2(m) - m$ as bivariate Gaussian random walk bridges. These two models yield equivalent Gaussian density functions. As a result, the simpler as-if i.i.d. formulation is adopted.

Perhaps a better stochastic model would be to let $(\Delta_1(m), \Delta_2(m))$ be first passage times of constant heights by BM. Density is well known. In this context, the RTT method is similar to SDE with diffusion term proportional to the drift term. Details are skipped but included below.

Likelihood function induced by RTT - contd

The random time likelihood function for the RTT model is obtained by multiplying the random time density above by the Jacobian of the transformation, the ratio 1 over the product over the sample of the differential terms $\beta g(X_m - R_m)(1 - X_m/K)$ and $\gamma(X_m - R_m)$.

K and α are MLE-estimated by maximizing the logarithm of the profile likelihood function, and their standard errors (and correlation coefficient, if needed) are estimated as usual, via the empirical Fisher information.

RTT model findings

Country	α	β	γ	σ_X	σ_R	ρ	K	X_{max}
Belgium	0.497	16.06	0.014	0.296	0.071	0.240	61510	59072
Brazil	0.849	0.57	0.046	0.255	0.148	0.578	N/A	672846
Germany	0.282	131	0.070	0.243	0.044	0.127	198917	185450
Italy	0.605	9.97	0.030	0.193	0.079	0.410	246234	234801
Switzerland	0.586	10.8	0.067	0.228	0.535	0.252	31575	30956
USA	0.246	1380	0.011	0.118	0.023	0.192	3370399	1920061

Table: Parameter estimation based on data until June 7, 2020

Country	α	β	γ	σ_X	σ_R	ρ	K	X_{max}
Italy	0.605	9.99	0.028	0.168	0.063	0.277	245431	230158
USA	0.272	478	0.011	0.143	0.042	0.294	2699321	1662302

Table: Parameter estimation based on data until May 25, 2020

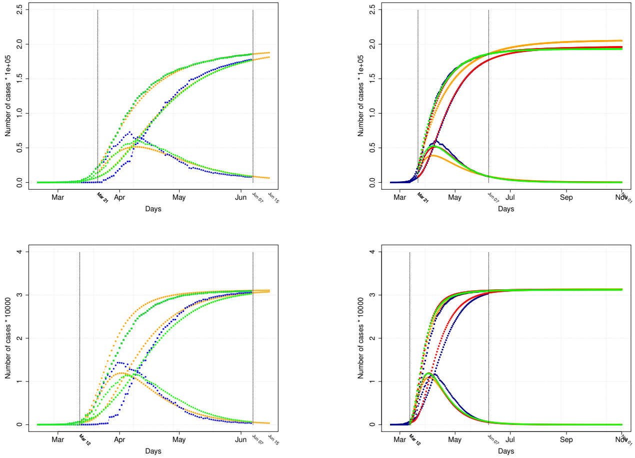

Figure: Data pre-processing and RTT solution. Germany (top), Switzerland (bottom) March-June 2020

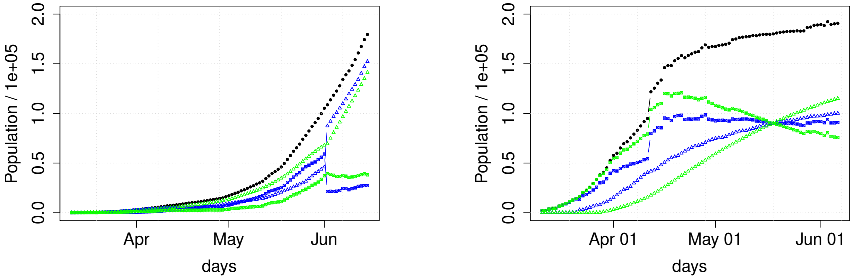

Figure: Data pre-processing: Raw-blue, modified-green. Chile (left), France (right), March-June 2020

The data of Chile and France show drastic corrections in the number of recovered cases, rendering analysis difficult.

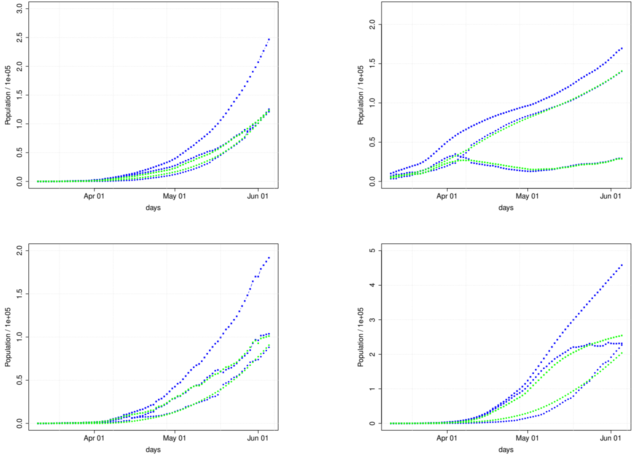

Figure: Data pre-processing: Raw-blue, modified-green. India (top left), Iran (top right), Peru (bottom left), Russia (bottom right), March-June 2020

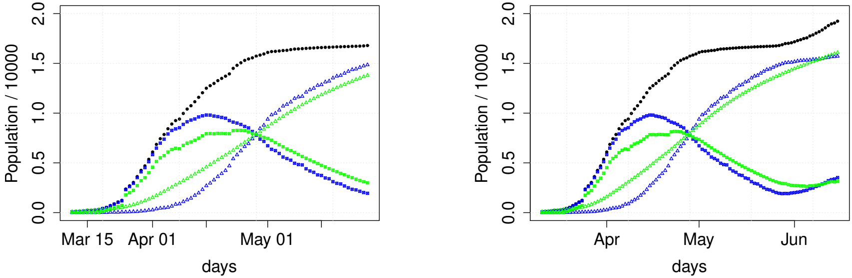

Figure: Israel data pre-processing: Raw-blue, modified-green. 25/05/2020 (left), 16/06/2020 (right)

The data of Iran and Israel show clear evidence of exit from steady conditions.

Appendix: A connection between RTT and SDE

If X and R are diffusions and time increments are small enough, local behavior is BM with drift given by the integrands in the SIR ODE and some diffusion coefficients.

As such, the incremental random times τ are first passage times of BM, first hitting times of heights D, with well known density functions, that should have been used to build the likelihood model.

In the benefit of simplicity and the ease with which the normal approximation handles dependent random times, we opted for the Gaussian inaccuracy.

Appendix: A connection between RTT and SDE contd

Let the BM B_t dealt with have local drift $\mu > 0$ and standard deviation σ.

Since $B_t - \mu t$ and $(B_t - \mu t)^2 - \sigma^2 t$ are mean-zero Martingales, their expected values at τ are zero too.
Hence, τ has mean $\frac{D}{\mu}$ and variance $\frac{\sigma^2 D}{\mu^3}$. Since τ has mean 1 and constant variance, σ must be proportional to μ. So the SDE corresponding to the SIR ODE has diffusion term proportional to the drift term.
If sampling is frequent enough, the Fokker-Planck equations lead to a likelihood function close to the Gaussian likelihood with exponential term

$$\exp\{-\frac{1}{2\sigma^2}\sum(\frac{X(i+1)-X(i)}{x(\frac{i+1}{\delta})-x(\frac{i}{\delta})}-1)^2\}$$

$x(\frac{i+1}{\delta}) - x(\frac{i}{\delta})$ can be replaced by the mid-interval differentials.

Thanks are due to Ilan Eshel from prompting this study and to Vahid Bokharaie, Amit Huppert, Eytan Ruppin, Laura Sacerdote and David Steinberg for helpful suggestions.
The data analyzed in this work is taken from the COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University.

Alon N. (2019), The effect of drift change on Skorohod embedded distributions with applications in Finance. Master's Thesis, Tel Aviv University

Bassan, B., Marcus, R., Meilijson, I. and Talpaz, H. (1997). Parameter estimation in differential equations, using random time transformations. *Journal of the Italian Statistical Society*, **6**, 177–199.

Fortiana, J. and Cuadras, C. M. (1997). A family of matrices, the discretized Brownian bridge, and distance-based regression. *Linear Algebra and its Applications*, **264**, 173—188.

Grenfell, B.T., Bjørnstad, O.N. and Filkenstädt, B. A. (2002) Dynamics of Measles epidemics: scaling, noise, determinism and predictability with the TSIR model. *Ecological Monographs*, **72(2)**, 185-–202.

Karatzas, I. and Shreve, S. E. (1998). *Brownian Motion and Stochastic Calculus*. Graduate texts in Mathematics, Springer.

Kermack, W. O., McKendrick, A. G. (1927). A Contribution to the Mathematical Theory of Epidemics. *Proceedings of the Royal Society A*, **115 (772)**, 700-–721.

Krengel, U. (1985). *Ergodic theorems*. de Gruyter.

Murphy, S. A. and Van Der Vaart, A. W. (2000). On Profile Likelihood. *Journal of the American Statistical Association*.