Estimation of the final size of the coronavirus epidemic by the SIR model

Method · February 2020

1 author:

Milan Batista
University of Ljubljana
144 PUBLICATIONS **668** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Port of Koper Industrial Risk Assessment View project

Navigability Study; Tanker Berth "JET" View project

Estimation of the final size of the coronavirus epidemic by the SIR model

Milan Batista

University of Ljubljana, Slovenia

milan.batista@fpp.uni-lj.si

(Feb 2020)

Abstract.

In the note, the SIR model is used for the estimation of the final size of the coronavirus epidemic. The current prediction is that the size of the epidemic will be about 85 000 cases. The note complements the author's note [1]

1. Introduction

In this note, we will try to estimate a final epidemic size by the SIR model [2, 3]. The program implements the model is available at

https://www.mathworks.com/matlabcentral/fileexchange/74658-fitviruscovid19

2. The SIR model

The model equations are

$$\frac{dS}{dt} = -\frac{\beta}{N} IS \,, \tag{1}$$

$$\frac{dI}{dt} = \frac{\beta}{N} IS - \gamma I \,, \tag{2}$$

$$\frac{dR}{dt} = \gamma I \,, \tag{3}$$

where t is time, $S(t)$ is the number of susceptible persons at time t, $I = I(t)$ is the number of infected persons at time t, $R(t)$ is the number of recovered persons in time

1

t. β is the contact rate, .and $1/\gamma$ is the average infectious period. From (1),(2),(3) we obtain total population size N

$$N = S + I + R = \text{const.} \tag{4}$$

The initial conditions are $S(0) = S_0$, $I(0) = I_0$, $R(0) = R_0$.

Eliminating I from (1) and (3) yield

$$S = S_0 \exp\left[-\frac{\beta}{N\gamma}(R - R_0)\right]. \tag{5}$$

In the limit $t \to \infty$ the number of susceptible people left S_∞ is

$$S_\infty = S_0 \exp\left[-\frac{\beta}{N\gamma}(R_\infty - R_0)\right], \tag{6}$$

where R_∞ is the final number of recovered persons. Because the final number of infected people is zero, we have, using (4),

$$N = S_\infty + R_\infty . \tag{7}$$

From this, and (6) the equation for R_∞ is

$$R_\infty = N - S_0 \exp\left[-\frac{\beta}{N\gamma}(R_\infty - R_0)\right]. \tag{8}$$

In order to use the model, we must estimate model parameters β, γ and initial values S_0 and I_0 from available data (we set $R_0 = 0$ and $I_0 = C_1$).

Now the available data is a time series of the total number of cases C, i.e.,

$$C = I + R . \tag{9}$$

We can estimate the parameters and initial values by minimizing the difference between the actual and predicted number of cases, i.e., by minimizing

$$\left\| C_t - \hat{C}_t\left(\beta, \gamma, S_0\right) \right\|^2 = \min , \tag{10}$$

where $C_t = \left(C_1, C_2, ..., C_n\right)$ are given number of cases in times $t_1, t_2, ..., t_n$ and $\hat{C}_t = \left(\hat{C}_1, \hat{C}_2, ..., \hat{C}_n\right)$ are corresponding estimates calculated by the model. For the

practical calculation, we use MATLAB's function *fminsearch* and for the integration of the model equation MATLAB's function *ode45.*

3. Results

The results of the calculation are shown in Table 1 and on Figure 1. From data in Table 1, we see that all data sets have high R^2 (>0.98). Also, we can see that the final number of recovered persons converge and the predicted values do not differ substantially; however, the predicted total population involved differs substantially. Here we note that from day 28, the collection of data changed. Until day 28, the estimated epidemic size was about 52 000 infections; after that prediction change to about 85 700 infections.

Table 1. Convergence study. After day 28, the method of data collection change.

|	Day	N	S_∞	R_∞	$R_{\infty,n}\big/R_{\infty,n-1}$	β	γ	$\beta\big/\gamma$	R^2
12.feb.20	28	551513	473888	77625	1.429	2.897	2.689	1.077	0.988
13.feb.20	29	1300538	1189616	110922	1.048	4.026	3.854	1.045	0.988
14.feb.20	30	1434310	1318035	116275	0.925	4.157	3.988	1.042	0.990
15.feb.20	31	1203132	1095609	107523	0.932	3.905	3.730	1.047	0.991
16.feb.20	32	1002252	901990	100262	0.953	3.624	3.441	1.053	0.992
17.feb.20	33	864976	769448	95528	0.969	3.387	3.198	1.059	0.993
18.feb.20	34	774076	681530	92546	0.969	3.205	3.010	1.065	0.994
19.feb.20	35	683791	594070	89722	0.978	2.998	2.798	1.071	0.994
20.feb.20	36	619431	531645	87787	0.985	2.835	2.630	1.078	0.994
21.feb.20	37	574963	488460	86503	0.990	2.712	2.504	1.083	0.994
22.feb.20	38	544793	459126	85667	0.993	2.624	2.413	1.087	0.995
23.feb.20	39	522591	437513	85078	0.993	2.557	2.343	1.091	0.995
24.feb.20	40	506492	421823	84669	0.995	2.506	2.291	1.094	0.995
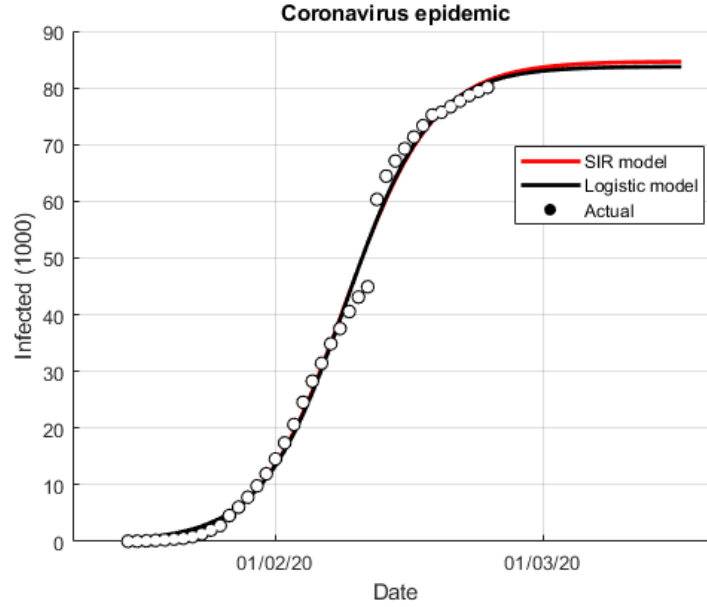
Figure 1. Actual and predicted number of cases by the SIR model and logistic model (data up to 25 Feb 2020)

Having a series of final predictions, we can estimate the series limit by Shanks transformation [4, 5]

$$R_\infty = \frac{R_{\infty,n+1}R_{\infty,n-1} - R_{\infty,n}^2}{R_{\infty,n+1} + R_{\infty,n-1} - 2R_{\infty,n}}. \tag{11}$$

For data from Table 1, the current prediction is 84085 cases (Table 2).

Table 2. Iterated Shanks transformation.

day	R_∞	$R(R_\infty)$	$R(R(R_\infty))$	$R(R(R(R_\infty)))$	$R(R(R(R(R_\infty))))$
32	100262				
33	95528	87470			
34	92546	39247	62344		
35	89722	83575	83974	84181	
36	87787	83971	84179	84084	84085
37	86503	84107	84003	84499	
38	85667	83673	83731		
39	85078	83740			
40	84669				

4. Conclusion

If a method of data collection will not change again, and if the situation will remain
stable, then by the SIR model, the predicted size of the epidemic is about 84 100 cases.
This prediction is comparable with the current prediction 84 000 infections, by the
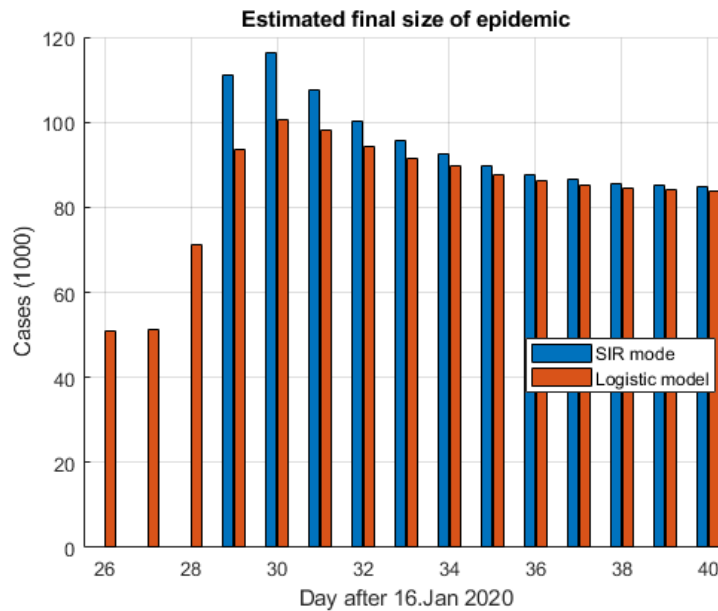empirical logistic model [6] (see Fig 2).

Figure 2. Comparison of models predictions (data up to 25.Feb 2020)

Appendix. MATLAB script

```
%SIR model - Dynamic of epidemic
%
% Reference:
%   https://en.wikipedia.org/wiki/Compartmental_models_in_epidemiology
%
% Variables
%   S - # of susceptible
%   I - # of infected
%   R - # recovered
%
% Parameters:
%   beta          -- 1/beta  = Tc - typical time between contacts
%   gamma         -- 1/gamma = Tr - typical time until recovery
%   N             --  = S + I + R = const
%
% Derived parameters:
%   RN = beta/gamma  -- basic reproduction number (ratio)
%
% History:
%   23.Feb.2020    MB Created

global beta gamma N    % parameters
global S0 I0 R0        % initial values (normalized S0=1)
```

```matlab
    global C               % number of infected
    global init            % initial guess

    close all

    % get data
    [CC,date0] = getData();

    fprintf('%12s %3s %10s %10s %10s %7s %7s %7s %7s\n',...
    'Date','Day','N','Sinf','Rinf','beta','gamma','R0','R2')
    init = false;
    for ii = 28:length(CC)

        % get data
        C = CC(1:ii);

        % estimate model parameters
        b = parest();

        % final number
        Rinf = Rmax();
        Sinf = S0*exp(-beta/gamma/N*(Rinf - R0));

        % calculate R2
        tspan = 0:length(C)-1;  % final time
        ic    = [S0 I0 R0]';     % initial conditions
        opts  = [];              % no options set
        [~,z] = ode45( @SIR, tspan, ic, opts);
        z = z(:,2)+z(:,3);
        zbar = sum(C)/length(C);
        SStot = sum((C - zbar).^2);
        SSres = sum((C - z').^2);
        R2 = 1 - SSres/SStot;

        % print results
        fprintf('%12s %3d %10d %10d %10d %7.3f %7.3f %7.3f %7.3f\n',...
            datestr(date0+ceil(length(C)-
1)),ceil(length(C)),round(N,0),...
            round(Sinf,0),round(Rinf,0),beta,gamma,beta/gamma,R2)
        % fprintf('Estimated parameters\n')
        % fprintf(' End date
%s\n',datestr(date0+ceil(length(C)-1)));
        % fprintf(' Day number           %d\n',ceil(length(C)));
        % fprintf(' Population size       %d\n',round(N,0));
        % fprintf(' Initial infected      %d\n',round(z(1,2),0));
        % fprintf(' Remain susceptible    %d\n',round(Sinf,0));
        % fprintf(' Total recovered       %d\n',round(Rinf,0));
        % fprintf(' Contact rate          %g\n',beta);
        % fprintf(' recovery rate         %g\n',gamma);
        % fprintf(' Recovery time         %g\n',1/gamma);
        % fprintf(' Reproduction number   %g\n',beta/gamma);
        % fprintf(' R2                    %g\n',R2);
    end

    % set parameters
    tspan = 0:2*length(C);  % final time
    ic    = [S0 I0 R0]';     % initial conditions
    opts  = [];              % no options set

    % simulate
    [t,z] = ode45( @SIR, tspan, ic, opts);
```

```matlab
% plot results
figure
hold on
plot(t,z,'LineWidth',2)
legend('Susceptible','Infected','Recovered',...
    'Location','best','FontSize',12)
xlabel('Day (after 16 jan 2020)')
ylabel('Cases')
grid on
hold off
shg

% plot comparsion
figure
hold on
plot(t,(z(:,2)+z(:,3))','k','LineWidth',2)
scatter(1:length(C),C,50,'filled')
legend('Predicted','Actual',...
    'Location','best','FontSize',12)
xlabel('Day (after 16 jan 2020)')
ylabel('Cases')
grid on
hold off
shg

% save to global
ta = t;
Ca = z(:,2)+z(:,3);

function dzdt = SIR(~,z)
%SIR model
% x(1) = S,   x(2) = I, x(3) = R
    global beta gamma N
    S = z(1);
    I = z(2);
    R = z(3);
    N = S + I + R;
    dzdt = [ -beta*I*S/N; beta*I*S/N - gamma*I; gamma*I];
end

function b = iniguess()
%INIGUESS Obtain initial guess
    global beta gamma
    global S0 I0 R0         % initial values
    global C
    global init
    if ~init
        beta = 1/0.00267103;
        gamma = 1/0.00267232;
        S0 = 1e8;
        I0 = C(1);
        R0 = 0;
        init = true;
    end
    b(1) = beta;
    b(2) = gamma;
    b(3) = S0;
end

function b = parest
```

```matlab
%PAREST Parameter estimation
    global beta gamma N
    global S0 I0 R0        % initial values
    maxiter = 20000;
    maxfun  = 20000;
    b0 = iniguess();
    options = optimset('Display','off','MaxIter',maxiter,...
        'MaxFunEvals',maxfun);
    [b, fmin,flag] = fminsearch(@fun,b0,options);
    %      disp('Exit condition:')
    %      disp(flag)
    %      disp('Smallest value of the error:');
    %      disp(fmin)
    fun(b); % obtain x ped
    beta  = b(1);
    gamma = b(2);
    S0    = b(3);
    N     = S0 + I0 + R0;
end

function f = fun( b)
%FUN Optimization function
    global beta gamma S0 I0 R0
    global C Ca

    % set parameters
    beta  = b(1);
    gamma = b(2);
    S0    = b(3);

    tend  = length(C);

    tspan = 0:tend-1;   % time interval
    ic = [S0 I0 R0]';   % initial conditions

    % solve ODE
    try
        [tsol,zsol] = ode45(@SIR,tspan,ic);
    catch
        f = NaN;
        return
    end

    % check if calculation time equals sample time
    if length(tsol) ~= length(tspan)
        f = NaN;
        return
    end

    Ca = (zsol(:,2)+zsol(:,3))'; % calculated number of cases
    f = norm(C - Ca);

end

function r = Rmax()
%FSOLVE Calculate number of recoverd individuals after t=inf
    global N S0 beta gamma R0
    RN = beta/gamma;
    r = fzero(@f,[0,S0]);
    %-----------------------
    function z = f(x)
```

```
        z = x - (N - S0*exp(-RN*(x - R0)/N));
    end
end

function [C,date0] = getData()
%GETDATA Coronavirus data
%   data from 16 Jan to 21 Jan https://i.redd.it/f4nukz4ou9d41.png
%   data from from 22 Jan 2020 to 13 Feb 2020 are from
%   https://www.worldometers.info/coronavirus/
date0=datenum('2020/01/16'); % start date
C = [
45
62
121
198
291
440
580
845
1317
2015
2800
4581
6058
7813
9821
11948
14551
17389
20628
24553
28276
31439
34876
37552
40553
43099
44919
60326
64437
67100
69197
71329
73332
75198
75700
76676
77673
78651
79400
80088
%----------- add new data here
]';
end
```

References

[1] M. Batista, Estimation of the final size of the coronavirus epidemic by the logistic model, medRxiv (2020) 2020.02.16.20023606.

[2] H.W. Hethcote, The Mathematics of Infectious Diseases, SIAM Review 42(4) (2000) 599-653.

[3] I. Nesteruk, Statistics based predictions of coronavirus 2019-nCoV spreading in mainland China, medRxiv (2020) 2020.02.12.20021931.

[4] D. Shanks, Non-linear Transformations of Divergent and Slowly Convergent Sequences, Journal of Mathematics and Physics 34(1-4) (1955) 1-42.

[5] C.M. Bender, S.A. Orszag, Advanced mathematical methods for scientists and engineers I asymptotic methods and perturbation theory, Springer, New York, 1999.

[6] M. Batista, Estimation of the final size of the coronavirus epidemic by the logistic model, 2020 DOI: 10.13140/RG.2.2.36053.37603.