Brief Analysis of Norovirus Data

Andrew Martin

June 6, 2019

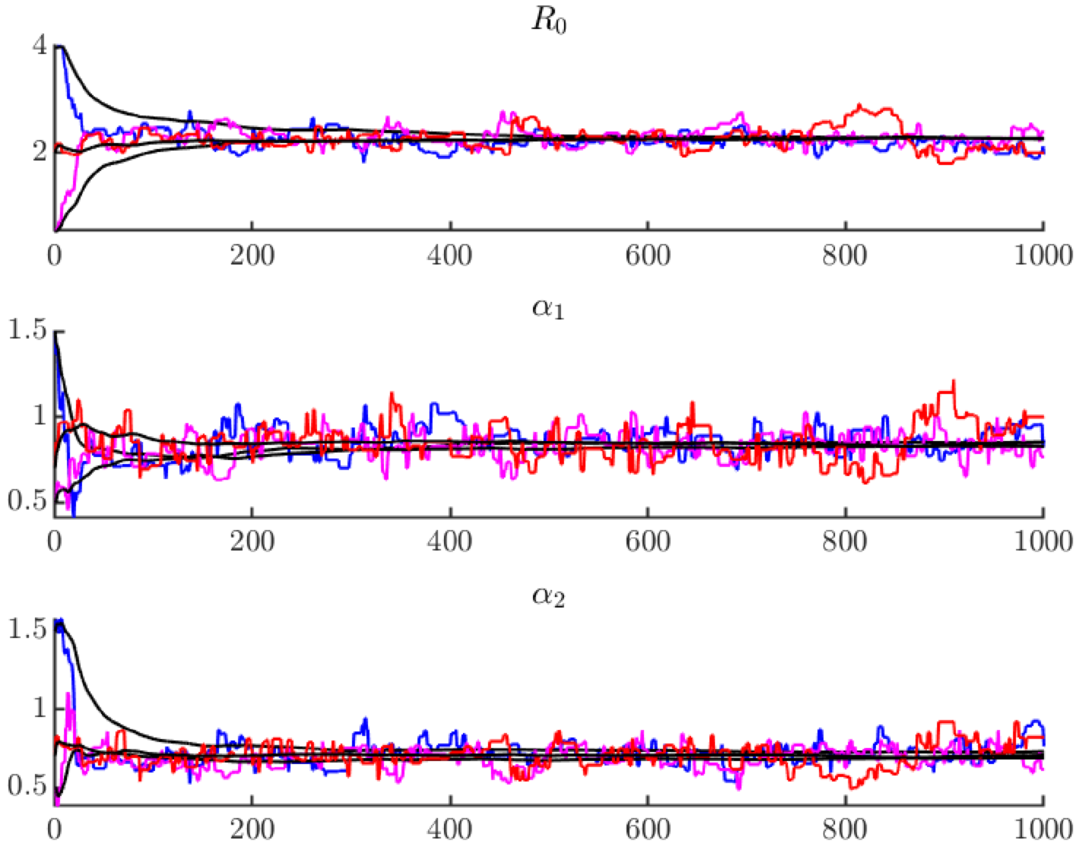

Figure 1: Caption here

Advise the government on effectiveness of interventions, and which should be implemented (if any).

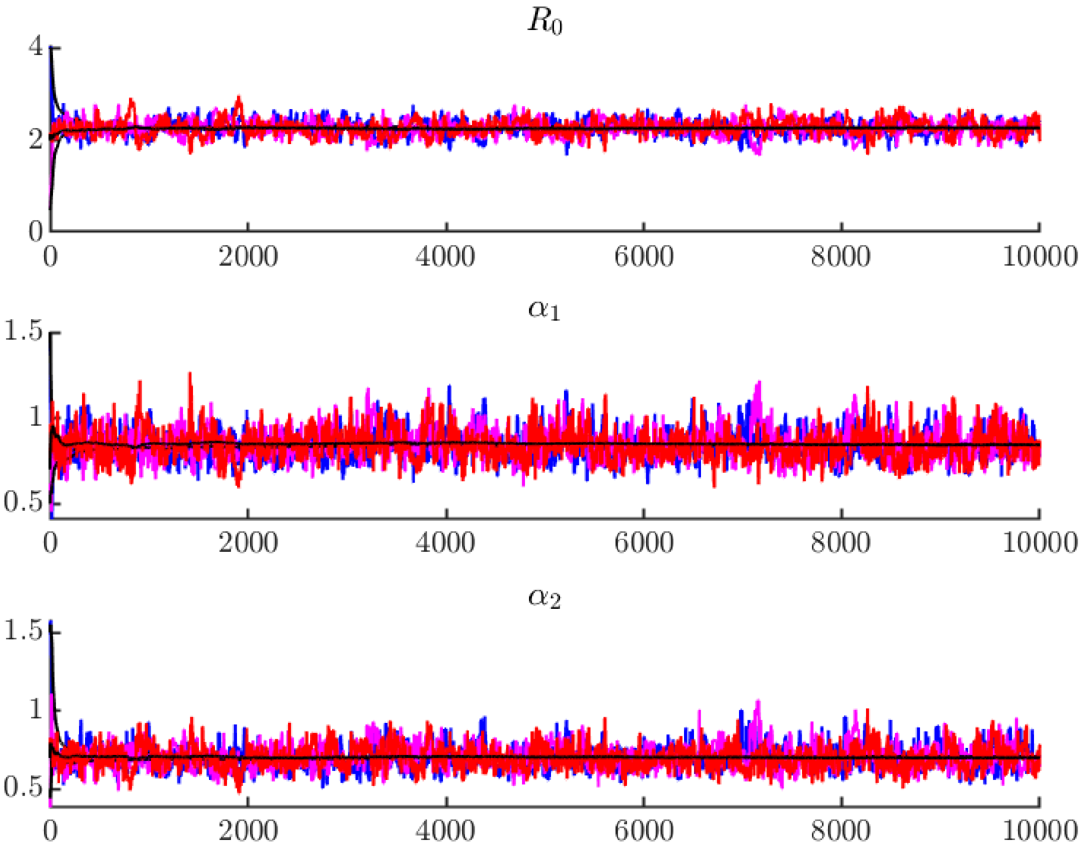

Figure 2: Caption here

.1 Code

.1.1 Main Script (R0Predict.m)

```
%Pretty plots
set(groot, 'DefaultLineLineWidth', 1, ...
    'DefaultAxesLineWidth', 1, ...
    'DefaultAxesFontSize', 12, ...
    'DefaultTextFontSize', 12, ...
    'DefaultTextInterpreter', 'latex', ...
    'DefaultLegendInterpreter', 'latex', ...
    'DefaultColorbarTickLabelInterpreter', 'latex', ...
    'DefaultAxesTickLabelInterpreter','latex');



%Speeds up the code for multiple runs
if ~exist('dataMat','var')
    dataMat = readmatrix('NorovirusDataA3.txt');
    %print the first few rows
    dataMat(1:5,:)
end

%Observation of dataset
%how many of them correspond to the disease effectively dying out?
amountOfData = length(dataMat(:,1))
propDiedOut = sum(dataMat(:,2)<=0.1*dataMat(:,1))/amountOfData

%dataMat = [#people, #infected, #action]
%consistency
rng(1)



%since we are expecting approximately 66%
%getting the right prior
%we want the dist to be N'(2.5,sigma)
%where sigma is the variance to have approx 66% in (2-3).
%N' since we will drop values below 0 and above 5
%N' will still be symmetric
sigma = -(3-2.5)/norminv((1-0.66)/2);
variance = sigma* speye(3);

varR0 = 0.01;
distStruct.ProposalDist = @(x) mvnrnd(x,diag([varR0,0.01,0.01]));
distStruct.PriorPDF = @(x) normpdf(x,2.5,sigma);
distStruct.LogLikelihoodFunc =@(params,data) LogLikelihood(params,data);
%constrain a_1,a_2 in [0,2]
distStruct.ProposalConstraints = @(vars) ProposalConstraints(vars,[0,0,0],[5
%overshoot, undershoot and approximately close
```

```
startVars = [4,1.5,1.5;
             0.5,0.5,0.5;
             2,0.7,0.7];

vars = zeros(10000,3,3);
for index = 1:3
    startPoint = startVars(index,:)
    exitflag = 1;
    while exitflag~=0
        [varsTemp,accrate,exitflag]= MetropolisHastingsPassLikelihood(distStr
        if exitflag==-1
            %acceptance was too low
            %variance is too high
            varR0 = varR0 *2/3;
            distStruct.ProposalDist = @(x) mvnrnd(x,diag([varR0,0.01,0.01]));
            warning('retrying with decreased variance')
        else if exitflag==1
            %acceptance was too high
            %variance is too low
            varR0 = varR0*2;
            distStruct.ProposalDist = @(x) mvnrnd(x,diag([varR0,0.01,0.01]));
            warning('retrying with decreased variance')
            end
        end
        vars(:,:,index) = varsTemp;
    end


end
%%
cumVars = cumsum(vars,1)./(1:length(vars))';
tstr = ["$$R_0$$","$$\alpha_1$$","$$\alpha_2$$"];
for burnin = 0:1
    figure
    for i=1:3
        hold on
        subplot(3,1,i)
        hold on
        plot(vars(:,i,1),'b')
        plot(cumVars(:,i,1),'k')
        plot(vars(:,i,2),'m')
        plot(cumVars(:,i,2),'k')
        plot(vars(:,i,3),'r')
        plot(cumVars(:,i,3),'k')
        title(tstr(i))

        if burnin
            axis([0,1000,-inf,inf])
        end
    end
```

```
        saveas(gcf,"MHplot"+num2str(burnin)+".eps","epsc")
end
    axis([0,1000,-inf,inf])
    xlabel('Iteration')

%%

%%Clean new data
%just take one set since they all converged
%and omit the burnin
varsClean = vars(500:end,:,3);

close all
%density plot
%and obtain estimates for R0, a1, a2

labs = ["$$R_0$$","$$\alpha_1$$","$$\alpha_2$$"];

est = [0,0,0];
for i=1:3
    figure
    [prob,val] = ksdensity(varsClean(:,i));
    prob = prob./sum(prob);
    plot(val,prob)
    xlabel(labs(i))
    ylabel("Probability")
    title("Probability density for "+labs(i))
    [~,ind] =max(prob);
    %assuming there is no dependence
    est(i) = val(ind);
    saveas(gcf,"Probdensity"+num2str(i),"epsc")
end
est
est(2:3) * est(1)

for i=1:2
    for j=i+1:3
        figure
        binscatter(varsClean(:,i),varsClean(:,j),60,'HandleVisibility','off')
        hold on
        scatter(est(i),est(j),'xr')
        xlabel(labs(i))
        ylabel(labs(j))
        legend("Independent approximation")
        colormap(gca,'parula')
        saveas(gcf,"BinScatter"+num2str(i)+num2str(j),"epsc")
    end
end
```

.1.2 MetropolisHastingsPassLikelihood.m

```
function [vars,accRate,exitflag]= MetropolisHastingsPassLikelihood(distStruct
%Most generic MetropolisHastings
%distStruct has fields
%-priorDist - the prior to pull from
%-proposalDist - the proposed distribution

numAccepted = 0;
PriorPDF = distStruct.PriorPDF;
ProposalDist = distStruct.ProposalDist;
BreaksProposalConstraints = distStruct.ProposalConstraints;
LogLikelihoodFunc = distStruct.LogLikelihoodFunc;
vars = zeros(numIterations,length(startVars));
vars(1,:) = startVars;

for i=2:numIterations
    proposal = ProposalDist(vars(i-1,:));
    %if breaks constraints
    if BreaksProposalConstraints(proposal)
        vars(i) = vars(i-1);
    else
        candidateProbTop = LogLikelihoodFunc(proposal,data)  + log(PriorPDF(
        candidateProbBottom = LogLikelihoodFunc(vars(i-1,:),data) + log(Prior
        candidateProb = candidateProbTop - candidateProbBottom;
        acceptProb= log(rand);

        if acceptProb< candidateProb
            vars(i,:) = proposal;
            numAccepted = numAccepted +1;
        else
            vars(i,:) = vars(i-1,:);
        end
    end
end
accRate = numAccepted/numIterations;

if  accRate < 0.2
    warning(”Bad acceptance rate - too low, accRate = ”+num2str(accRate));
    exitflag = -1;
else if accRate >0.27
    exitflag = 1;
    warning(”Bad acceptance rate - too high, accRate = ”+num2str(accRate));
else
    exitflag = 0;
    end
end


end
```

.1.3 LogLikelihood.m

```matlab
function logLikelihood = LogLikelihood(params,data)
%Amended SIR finalsize code from Josh
%Amended by Andrew Martin
%calculates log likelihood for a given R0,alpha1,alpha2
logLikelihood = 0;
% R0      = params(1);
% alpha1  = params(2);
% alpha2  = params(3);
NVec = data(:,1);
%[R0, a1R0, a2R0]
paramsModified = [params(1),params(1)*params(2),params(1)*params(3)];
%yay matlab uses 1 based indexing
R0index = data(:,3)+1;


for iterator=1:length(data)
    N = NVec(iterator);
    relevantParam = paramsModified(R0index(iterator));
    q = zeros(N+1,1);
    q(2) = 1;
    %Proportions for each number of infection events
    %could vectorise, but this is more meaningful
    for Z2 = 0:N
        for Z1 = Z2+1:N-1
            %infection probability (jump prob)
            infProb = 1 / ( 1 + ((N-1)/(relevantParam*(N-Z1))));
            q(Z1+2) = q(Z1+2) + q(Z1+1)*infProb;
            q(Z1+1) = q(Z1+1)*(1-infProb);
        end
    end
    %sum of the log likelihoods (product of likelihoods)
    logLikelihood= logLikelihood + log(q(data(iterator,2)+1));
end


end
```

.1.4 ProposalConstraints.m

```matlab
function boolean = ProposalConstraints(vals,min,max)
%
%OUTPUT:
%boolean - true if the boundary constraints are broken
boolean = any(vals < min | vals > max);


% %hardcoded way
% boolean  =vals(1) < 0 || vals(1) > 5 ...
%          || vals(2) < min || vals(2) > max ...
%          || vals(3) < min || vals(3) > max;

```

```
end
```

Assignment III

Worth 15% of course assessment; due by 3pm on Friday 7th June, 2019.

Most-relevant lectures: Lectures 21 – 26.

The total marks for this assignment is 45.

Please provide code where appropriate.

Report to the Government on the effectiveness of interventions. [45 marks]

The CSV file `NorovirusDataA3` contains information regarding independent outbreaks in a set of 125 hospital wards of varying sizes. The first column contains the number of occupied beds in the ward, the second column contains the number of those patients which succumbed to norovirus during the outbreak, and the third column indicates the control action implemented; in the last column, 0 corresponds to standard practices, 1 corresponds to a trial intervention strategy, and 2 corresponds to different trial intervention strategy.

You are to analyse this data to advise the Government on the effectiveness of their interventions, and to advise them on which intervention should be adopted (if any).

You may assume that the interventions work to reduce the effective transmission rate parameter. The Chief Medical Officer has said that R_0 for norovirus in typical hospital settings is between 2 and 3 with (approximately) 66% probability; you may use this expert opinion as prior knowledge.

You are to prepare two reports. One is for the Government. The second is to provide detail on how you have performed the statistical analysis, including the model(s) used, any assumptions you have made and providing evidence that your approach/algorithms are working correctly, for example through the use of simulated data of similar form to the 'real' data (e.g., using trace plots from multiple independent chains, and kernel density estimators and box plots). This assignment is deliberately vague – you need to make decisions, but feel free to ask for feedback as you make progress.