**Contents of the file ‘EpiEstimation.R’ and brief description of syntax**

**15 March 2011**

**Note that the library ‘Hmisc’ is required for any program that simulates data.**

**Contents**

1. GetR0.WP()
2. getR0SI.WP()
3. getRt.WP()
4. getRt.WT()
5. likN.MN()
6. likN.gam()
7. likN.logit.SIknown()
8. likN.logit.SIunknown()
9. getNsMN()
10. getNsGam()
11. getNSMNL()
12. getNsGamL()
13. getSummaryMN()
14. covProbsWP()

**Functions to perform estimation**

1. getR0.WP(N,pars,distn="gamma",k=20)

Obtains an estimate of R0 for a known serial interval.

*Inputs:*

N: vector of the number of new cases each time unit (day, week, etc)

pars: variables that describe the serial interval distribution. Given as a vector of pj that add to one if a multinomial distribution or c(shape,rate) if a gamma distribution. Note that shape/rate=mean of gamma and shape/rate2 = variance of a gamma distribution.

distn: indicates the shape of the serial interval. “gamma” or “MN” are the options.

k: maximal length of the serial interval

1. getR0SI.WP(N,distn=”gamma”, k=6)

Get the estimate of R0 and the serial interval using the White and Pagano method for either a multinomial or gamma distributed serial interval. If distn=”gamma” then the output will be in the form of (R0,shape,rate) where shape and rate are the parameters of a gamma distribution describing the serial interval. If distn=”MN”, then the output is in the form (R0,p­1,…,pk), where p­1,…,pk describe the serial interval distribution.

*Inputs:*

N: vector of the number of new cases each time unit (day, week, etc)

distn: Indicates the parametric form of the serial interval. Either “gamma” or “MN”.

k: maximum length of the serial interval. Defaults to 6.

1. getRt.WP(N,distn=”gamma”,dist.pars,k=6)

Get the estimate of Rt using the White and Pagano method, assuming that Rt follows a four parameter logistic, and that the serial interval is known. Makes use of the likelihood likN.logit.SIknown().

*Inputs:*

N: vector of the number of new cases each time unit (day, week, etc)

distn: Indicates the parametric form of the serial interval. Either “gamma” or “MN”.

dist.pars: parameters of the serial interval. If distn=”gamma” then c(shape,rate); if distn=”MN” then c(p1,…,pk).

k: maximum length of the serial interval

1. getRt.WT(N,distn=”gamma”,dist.pars,k=6)

Get the estimate of Rt using the Wallinga and Teunis method. Returns the value of Rt that is the same length as N.

*Inputs:*

N: vector of the number of new cases each time unit (day, week, etc)

distn: Indicates the parametric form of the serial interval. Either “gamma” or “MN”.

dist.pars: parameters of the serial interval. If distn=”gamma” then c(shape,rate); if distn=”MN” then c(p1,…,pk).

k: maximum length of the serial interval

**Likelihood Functions for White and Pagano Method**

1. likN.MN(params,N)

Provides the value of the –log of the likelihood for estimation with a multinomial serial interval and R0 for the initial exponential growth phase of an epidemic.

*Inputs:*

params: take the form (R0p1, …,R0pk)

N: vector of the number of new cases each time unit (day, week, etc)

1. likN.gam(params,N,k)

Provides the value of the –log of the likelihood for estimation with a multinomial serial interval and R0 for the initial exponential growth phase of an epidemic.

*Inputs:*

params: take the form (R0,shape,rate)

N: vector of the number of new cases each time unit (day, week, etc)

k: maximum length of the serial interval

1. likN.logit.SIknown(params,distn="gamma",dist.pars,N,k=6)

Provides the value of the –log of the likelihood for estimation with a known gamma or multinomial serial interval and Rt for an entire epidemic. Assumes that Rt follows a four parameter logistic distribution given by

**

*Inputs:*

params: take the form (a,b,c,d)

distn: specifies the parametric form of the serial interval. “gamma” and “MN” are the only options.

dist.pars: provides the parameters for the serial interval. If distn=”gamma” then c(shape,rate); if distn=”MN” then c(p1,…,pk).

N: vector of the number of new cases each time unit (day, week, etc)

k: maximum length of the serial interval

1. likN.logit.SIunknown(params,N,k=6)

Provides the value of the –log of the likelihood for estimation with an unknown gamma serial interval and Rt for an entire epidemic. Assumes that Rt follows a four parameter logistic distribution given by

**

*Inputs:*

params: take the form (a,b,c,d,shape,rate)

N: vector of the number of new cases each time unit (day, week, etc)

k: maximum length of the serial interval

**Function to simulate epidemic data (based on the likelihood of White and Pagano)**

1. getNsMN(R0,ps,num.samples=1000,epi.size,num.start,perc.off=0.02)

Simulate epidemic data assuming a multinomial distribution for the serial interval. The size of the simulated data is determined by the number of cases simulated.

The output is a matrix where each row is a simulated epidemic. After epi.size cases are generated, the rest of the row is filled in with 0’s, where necessary.

*Inputs:*

R0: the value of the reproductive number for the simulation.

ps: vector of pj  for the serial interval. Must add to 1.

num.samples: number of datasets to generation. Defaults to 1000.

epi.size: total number of cases to simulate

num.start: initial number of cases (i.e. No)

perc.off: Value between 0 and 1. The margin of error for the total number of cases simulated. If it is specified that perc.off=0.02 and epi.size=100 then the total number of cases will be within 2 of 100 (i.e. between 98 and 102). Defaults to 0.02.

1. getNsGam(R0,shape,rate,num.samples=1000,epi.size,num.start,perc-off=0.02)

Simulate epidemic data assuming a gamma distribution for the serial interval. The size of the simulated data is determined by the number of cases simulated.

The output is a matrix where each row is a simulated epidemic. After epi.size cases are generated, the rest of the row is filled in with 0’s, where necessary.

*Inputs:*

R0: the value of the reproductive number for the simulation.

shape, rate: shape and rate parameters for the gamma distributed serial interval.

num.samples: number of datasets to generation. Defaults to 1000.

epi.size: total number of cases to simulate

num.start: initial number of cases (i.e. No)

perc.off: Value between 0 and 1. The margin of error for the total number of cases simulated. If it is specified that perc.off=0.02 and epi.size=100 then the total number of cases will be within 2 of 100 (i.e. between 98 and 102). Defaults to 0.02.

1. getNsMNL(R0,ps,num.samples=1000,max.days,num.start)

Simulate epidemic data assuming a multinomial distribution for the serial interval. The size of the simulated data is determined by length (for instance number of days) of data simulated.

The output is a matrix with num.samples rows and max.days columns.

*Inputs:*

R0: the value of the reproductive number for the simulation.

ps: vector of pj  for the serial interval. Must add to 1.

num.samples: number of datasets to generation. Defaults to 1000.

max.days: total number of days to simulate

num.start: initial number of cases (i.e. No)

1. getNsGamL(R0,shape,rate,num.samples=1000,max.days,num.start)

Simulate epidemic data assuming a gamma distribution for the serial interval. The size of the simulated data is determined by the number of days of data to simulate.

The output is a matrix with num.samples rows and max.days columns.

*Inputs:*

R0: the value of the reproductive number for the simulation.

shape, rate: shape and rate parameters for the gamma distributed serial interval.

num.samples: number of datasets to generation. Defaults to 1000.

max.days: total number of days to simulate

num.start: initial number of cases (i.e. No)

**Miscellaneous Functions**

1. getSummaryMN(pars)

Returns the summary statistics (R0, mean of the serial interval, and sd of the serial interval) for MN Serial interval. pars takes on the form of (R0\*p\_1, R0\*p\_2,…,R0\*p\_k).

1. covProbsWP (parsEsts,k,epi.size,num.start,numRep=1000,distn="MN",est.si=T)

Function to get estimates of the coverage probability of the estimates obtained using the White and Pagano method. This uses a bootstrapping technique to simulate numRep epidemics from the estimated parameters. Estimates of R0 and the serial interval (if estimated, otherwise just the assumed parameters of the serial interval) are provided as a vector in parsEsts. The the parameters that are estimated are obtained from each of the numRep datasets and are used for calculating the coverage probability (the 2.5 and 97.5th percentiles). The returned matrix has a column for each estimated parameter. The rows give the 2.5th, 50rh (median), 97.5th percentiles, and the min and max from the bootstrap samples. If est.si=T this indicates that the serial interval is an estimate and coverage probabilities should be obtained for those parameters. If est.si=F, then the serial interval is assumed known and coverage probabilities are not obtained.

*Inputs:*

parsEsts: the parameters that we are interested in. If a multinomial distribution is used, this should be c(R0, **p**). If a gamma distribution is used, this will be c(R0,shape,rate). If the parameters of the serial interval are assumed known, they must still be included here, but est.si will be equal to F.

k: maximum length of the serial interval..

epi.size: the size of the epidemic that is analyzed (should come from the original data).

num.start: initial number of cases in observed data (should come from the original data).

numRep: number of simulated datasets to base the coverage probabilities on. Defaults to 1000.

distn: assumed distribution of the serial interval. Can be either “gamma” or “MN”.

est.si: indicator of whether or not the serial interval in parsEsts is an estimate or not. If it is set to T (the default) then it is assumed that the serial interval is estimated and coverage probabilities will be calculated for these parameters. If it is set to F, then the serial interval parameters are assumed to be fixed and known.