THE LANCET Infectious Diseases

Supplementary webappendix

This webappendix formed part of the original submission and has been peer reviewed. We post it as supplied by the authors.

This online publication has been corrected. The corrected version first appeared at thelancet.com/infection on May 19, 2015.

Supplement to: Althaus CL. Ebola superspreading. *Lancet Infect Dis* 2015; **15:** 507–8.

Appendix: Individual variation in infectiousness for Ebola virus disease

Christian L. Althaus (christian.althaus@alumni.ethz.ch)

Overview

This document describes the analysis of the number of secondary cases infected with Ebola virus disease (EVD), the serial interval distribution, and the resulting outbreak trajectories. All analyses were performed in the R software environment for statistical computing.[1]

```
library(fitdistrplus)
```

Distribution of secondary cases

The transmission trees in figure 2D from Faye et al.[2] can be used to obtain the distribution of the number of secondary cases generated by an infected index case.

```
# Number of individuals in the trees
n <- 152
# Number of secondary cases for all individuals
c1 <- c(1,2,2,5,14,1,4,4,1,3,3,8,2,1,1,4,9,9,1,1,17,
        2,1,1,1,4,3,3,4,2,5,1,2,2,1,9,1,3,1,2,1,1,2)
c0 <- c(c1,rep(0,n-length(c1)))
# Fitting a negative binomial distribution to the number of secondary cases
fit.cases <- fitdist(c0,"nbinom")
summary(fit.cases)
```

```
## Fitting of the distribution ' nbinom ' by maximum likelihood
## Parameters :
##      estimate Std. Error
## size   0.1814     0.0399
## mu     0.9538     0.1981
## Loglikelihood:  -177.2   AIC:  358.4   BIC:  364.4
## Correlation matrix:
##           size        mu
## size 1.0000000 0.0001384
## mu   0.0001384 1.0000000
```

Fitting a negative binomial distribution to the number of secondary cases provides maximum likelihood estimates of the mean (0.95, 95% confidence interval [CI]: 0.57-1.34) and the dispersion parameter $k = 0.18$ (95% CI: 0.1-0.26).

```
plot(fit.cases)
```

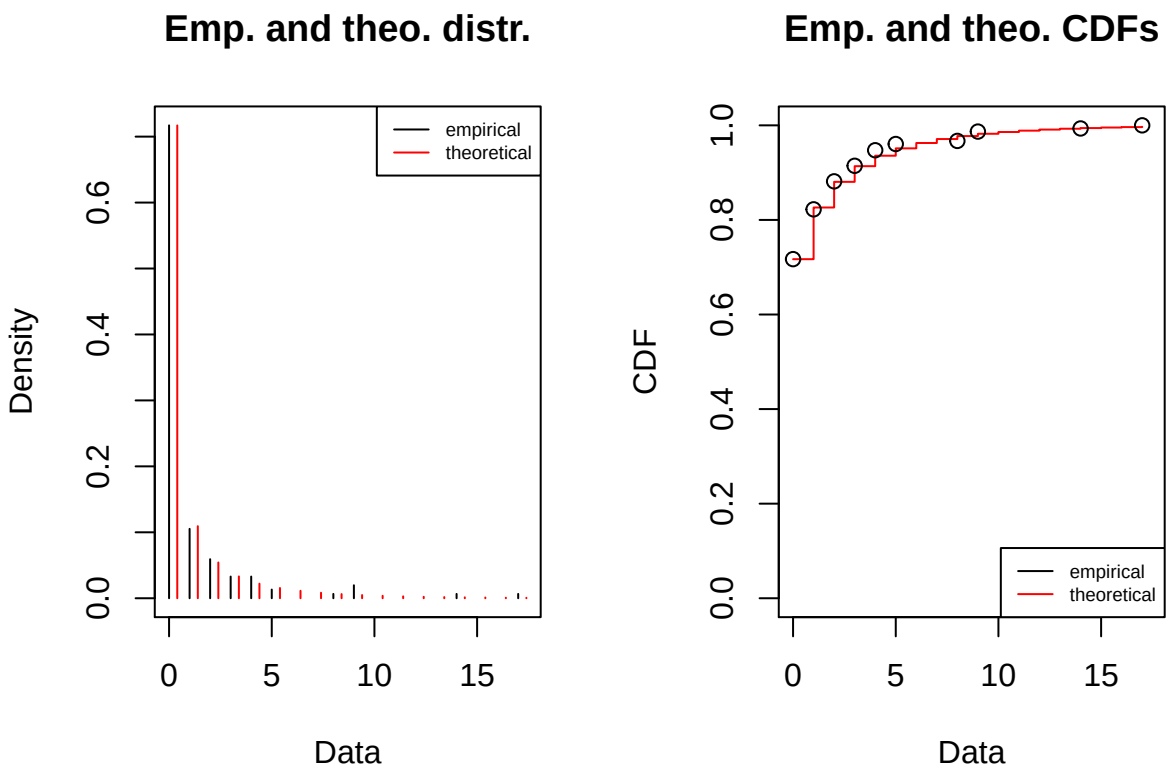

Figure 1: Empirial and theoretical density distribution and the cumulative density function (CDF) of the number of secondary cases.

Serial interval distribution

The serial interval distribution of EVD can be obtained from figure 3E of the study by the WHO Ebola Response Team.[3]

```
# Range of reported serial intervals
days <- 0:43
# Observed intervals for each day
frequency <- c(0,1,3,1,4,1,6,1,2,2,11,6,0,1,10,3,5,8,4,3,3,1,
               0,2,0,2,0,3,1,1,1,0,0,0,0,0,2,0,1,0,1,1,0,1)
d <- rep(days,frequency)
# Fitting a gamma distribution to the serial interval
fit.serial <- fitdist(d,"gamma")
summary(fit.serial)
```

```
## Fitting of the distribution ' gamma ' by maximum likelihood
## Parameters :
##       estimate Std. Error
## shape   2.5931    0.36042
## rate    0.1697    0.02602
## Loglikelihood:  -324.6   AIC:  653.1   BIC:  658.2
## Correlation matrix:
##        shape   rate
## shape 1.0000 0.9065
## rate  0.9065 1.0000
```

Fitting a gamma distribution to the data provides maximum likelihood estimates of the mean serial interval (15.28 days) and the shape parameter (2.59).

```
plot(fit.serial)
```

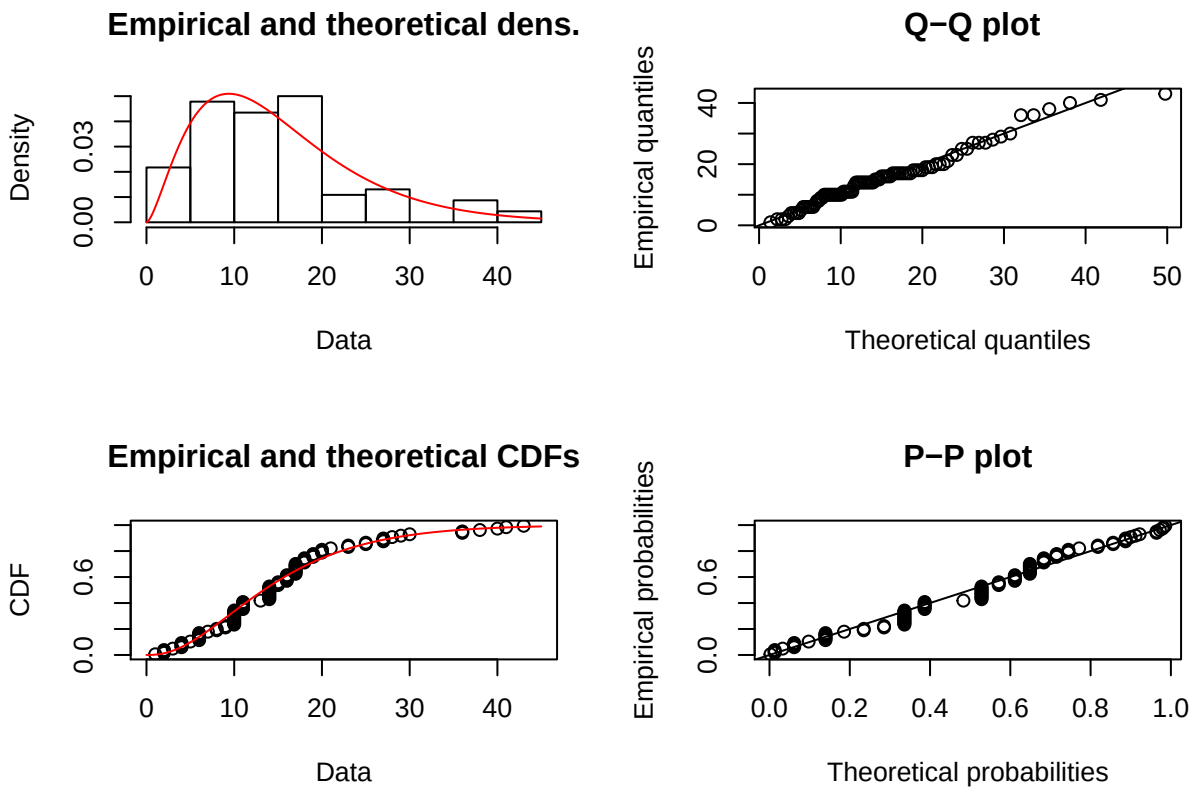

Figure 2: Empirial and theoretical density distribution and the cumulative density function (CDF) of the serial interval.

Simulating outbreaks

Based on the serial interval distribution and the number of secondary cases, one can simulate stochastic trajectories of EVD outbreaks starting from a single infected index case.

```
# Set seed for random number generator
set.seed(645)
# Number of simulation runs
runs <- 1e2
# Number of initial cases
seed <- 1
# Initialize plot
plot(NA,xlim=c(0,100),ylim=c(0,100),xlab="Time (days)",
     ylab="Cumulative number of EVD cases",frame=FALSE)
# Set color scheme for different trajectories
cols <- sample(terrain.colors(runs))
# Simulate outbreak trajectories
for(i in 1:runs) {
    cases <- seed
```

```
    t <- rep(0,seed)
    times <- t
    while(cases > 0) {
        secondary <- rnbinom(cases,size=fit.cases$estimate[1],mu=fit.cases$estimate[2])
        t.new <- numeric()
        for(j in 1:length(secondary)) {
            t.new <- c(t.new,t[j] + rgamma(secondary[j],shape=fit.serial$estimate[1],
                       rate=fit.serial$estimate[2]))
        }
        cases <- length(t.new)
        t <- t.new
        times <- c(times,t.new)
    }
    lines(sort(times),1:length(times),col=cols[i],lwd=1)
    points(max(times),length(times),col=cols[i],pch=16)
}
```

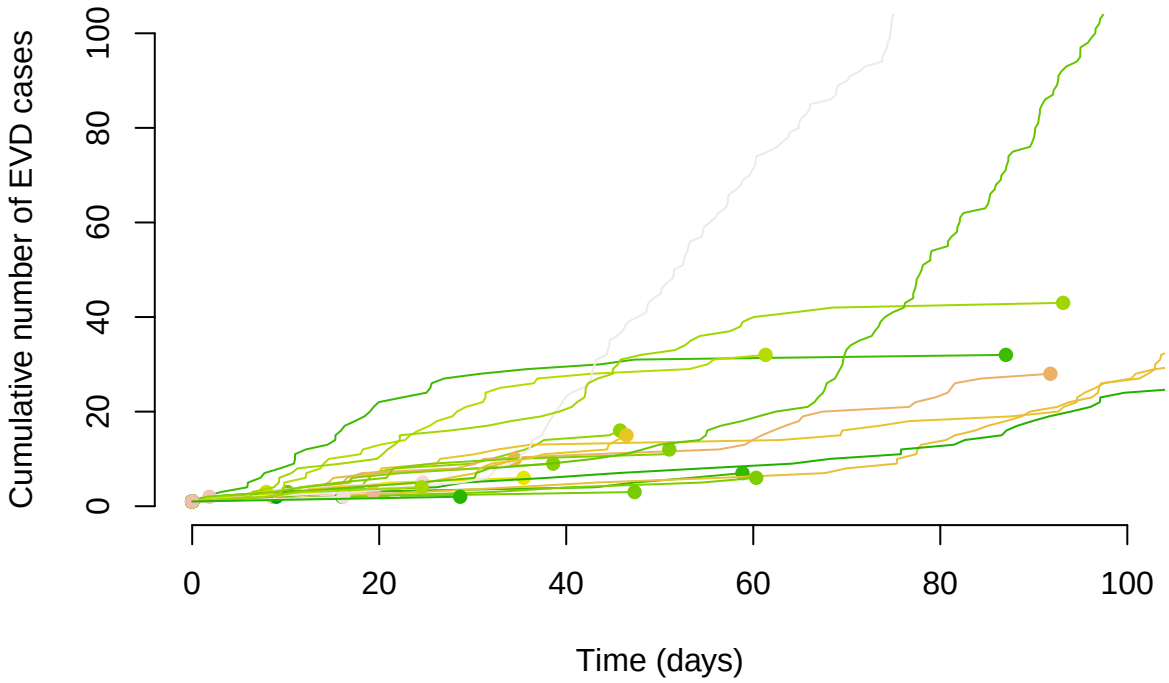

Figure 3: Simulated outbreaks of EVD. Each line represents one of 200 stochastic realizations of epidemic trajectories. Dots indicate that the outbreak goes extinct.

References

1R Development Core Team *et al.* R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing, 2014.

2Faye O, Boëlle P-Y, Heleze E*et al.* Chains of transmission and control of Ebola virus disease in Conakry, Guinea, in 2014: an observational study. *Lancet Infect Dis* 2015; published online Jan. DOI:10.1016/S1473-3099(14)71075-8.

3WHO Ebola Response Team *et al.* Ebola virus disease in West Africa–the first 9 months of the epidemic and forward projections. *N Engl J Med* 2014; **371**: 1481–95.