



Mathematical Statistics
Stockholm University

**Who was the infector – probabilities in
the presence of individual variability**

Åke Svensson

Research Report 2012:8

ISSN 1650-0377

Postal address:

Mathematical Statistics
Dept. of Mathematics
Stockholm University
SE-106 91 Stockholm
Sweden

Internet:

<http://www.math.su.se/matstat>



Mathematical Statistics
Stockholm University
Research Report **2012:8**,
<http://www.math.su.se/matstat>

Who was the infector – probabilities in the presence of individual variability

Åke Svensson

May 2012

Abstract

The probability that an observed infection has been transmitted from a particular member of a set of potential infectors is calculated. The calculations only use knowledge of the infection times. It is shown that the probabilities depend on the variability of individual infectivity. The analyses are based on different background information and different assumptions on the progress of infectivity. The results are illustrated by numerical calculations and simulations.

Who was the infector – probabilities in the presence of individual variability

Åke Svensson

Mathematical Department, Stockholm University

Summary. The probability that an observed infection has been transmitted from a particular member of a set of potential infectors is calculated. The calculations only use knowledge of the infection times. It is shown that the probabilities depend on the variability of individual infectivity. The analysis are based on different background information and different assumptions on the progress of infectivity. The results are illustrated by numerical calculations and simulations.

Keywords: Epidemic models, Generation times, Transmission chains

1. Introduction

When analyzing the spread of an infectious disease it is often of interest to identify the infector of an infected person. This can seldom be done with complete certainty. However, it may be possible to calculate the probability that a certain member of a set of potential infectors is the true transmitter. How this can be done, using only observations of the times of infection, is the topic of the present paper.

One reason to calculate probabilities is to better understand the transmission dynamics. In a study of SARS outbreaks Wallinga and Teunis (2004) analysed the possible impact of control measures by following the expected number of secondary cases during the course of the epidemic. Given times of infection, they could by adding the probabilities that a certain infector was responsible for future infections, estimate the expected number of secondary infections.

As in all statistical and probabilistic analysis we have to carefully consider how the observed data are generated. In section 3 we assume that the histories of the potential infectors are unrelated, and in section 4 that the candidates form a transmission tree, i.e., the set of possible infectors consists of one original infected and a sequence of persons that have infected each other. Cauchemez and Ferguson (2012) also study the related, and more complex, problem of how to find the most probable transmission chain. In their analysis variations in infectivity of potential infectors were related to observable quantities.

The probabilities, that we are interested in will depend on generation times, i.e. the times between a primary infection and its corresponding secondary infections. Individual random variations in infectivity will have substantial impact. In the examples used in this paper the variations are assumed to be generated by a SEIR model with random latent and infectious periods. Basic assumptions also concern homogeneous mixing, and constant infectivity during the infectious period. Of course, it is possible to analyze more complicated models, but our main purpose here is to illustrate the problem and to highlight possibilities to perform the calculations. For this reason we use a simple setting. The models and the notation are presented in section 2.

E-mail: akes@math.su.se

Section 5 gives numerical examples that illustrate that the assumptions underlying the analysis are crucial. In the simulated examples, as will certainly be the case for observations from real epidemics, the probabilities found, do in general not give a very precise indication of who was the infector for a specific case. Thus other information than the times of infection is required to precisely indicate the infector.

It is difficult to find general analytic expressions for the interplay between infection times and the underlying stochastic model. In section 6 it is discussed how the findings in the numerical examples may be generalized.

2. Basic model and notation

We will study a situation where v infections are observed to occur at times $\tau_1 < \tau_2 < \dots < \tau_v$. Without loss of generality we may assume that $\tau_1 = 0$.

2.1. Infectivity and generation times

It is assumed that an infected person spreads the infection according to an intensity process that depends on the time after the infection, i.e. the age of the infection. The intensity processes may be individual and random, but the random intensity functions for different individuals are assumed to be independent. We consider spread in a closed population that is assumed to be homogeneously mixing. The victim for a new infection is a randomly chosen member of the population.

The intensity process for the i 'th infected is denoted by κ_i . The interpretation is that the infected person has potentially infectious contacts according to a Poisson process with time-varying intensity $\kappa_i(a)$, where a is the age of the infection. The contact will lead to a transmission if the contacted is susceptible. If there exists immunity in the population the occurrence of new infections will be influenced by this. Let $s(t)$ be the proportion of susceptible persons in the population at time t . If the i 'th infected is infected at time τ_i secondary infections will occur according to a Poisson process with intensity $s(t)\kappa_i(t - \tau_i)$.

In this model the total infectivity that individual i spreads is

$$\lambda_i = \int_0^{\infty} \kappa_i(a) da \quad (1)$$

The parameter λ_i can be interpreted as the mean number of possible infectious contacts this infected individual takes.

The basic reproduction number, i.e. the expected number of secondary infections in a totally susceptible population, is then

$$R_0 = E(\lambda_i) = \int_0^{\infty} g(a) da, \quad (2)$$

where

$$g(a) = E(\kappa_i(a)). \quad (3)$$

The cohort (or basic) generation time density, k (cf Svensson (2007) and Tomba et al. (2010)), corresponds to g normalized to have total mass 1, i.e.,

$$k(a) = \frac{E(\kappa_i(a))}{R_0} = \frac{g(a)}{R_0}. \quad (4)$$

The mean generation time is defined as

$$\int_0^{\infty} ak(a)da. \quad (5)$$

2.2. Models for individual variations

In order to illustrate the effects of individual variability we will use a model of the type generally referred to as a SEIR model. An infection is assumed to be followed by a latent period, during which the infection is not transmitted. After the latent period follows an infectious period. We will here, for simplicity, assume that a person takes infectious contacts with a constant non-random rate γ , that is the same for all infected. If an infectious person takes such a contact with a susceptible person the infection is transmitted.

The duration of the latent and infectious periods may vary between individuals. Let L_i be the duration of the latent period and X_i the duration of the infectious period. Furthermore let $I_i(a)$ be the indicator function that the i 'th infected is infectious at time a after the infection, i.e.,

$$I_i(a) = \begin{cases} 1 & \text{if } L_i \leq a \leq L_i + X_i, \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$\kappa_i(a) = \gamma I_i(a). \quad (6)$$

Obviously $\lambda_i = \gamma X_i$ and $R_0 = \gamma E(X)$.

Assume that the pairs L_i and X_i are independent and that their respective distribution functions are H_L and H_I , with densities h_L and h_I , then

$$k(a) = \frac{H_L(a) - \int_0^a H_L(a-s)h_I(s)ds}{E(X)}. \quad (7)$$

If there is no latent period, i.e. when the infectious period starts immediately after infection,

$$k(a) = \frac{1 - H_I(a)}{E(X)}. \quad (8)$$

Furthermore

$$p(a) = \Pr(I_i(a) = 1) = E(X)k(a), \quad (9)$$

and the probability that the i 'th infected is infectious at time t is

$$p(t - \tau_i) = E(X)k(t - \tau_i).$$

As examples we will use three different cases where the latent and infectious periods are gamma-distributed. To make the cases comparable we will choose parameter values so that the mean length of the latent and the mean generation times in each case are approximately what is assumed for seasonal influenza (cf Carrat et al. (2008)). In the first case both the latent and the infectious periods are assumed to be exponentially distributed. This implies a large individual variation. In the second case the variation is smaller, and in the third case the times are assumed to be constant.

Case 1: L_i is exponential distributed with intensity μ_L , and X_i is exponential distributed with intensity μ_I . If $\mu_L \neq \mu_I$

$$k(a) = \frac{\mu_L \mu_I}{\mu_I - \mu_L} (\exp(-\mu_L a) - \exp(-\mu_I a)),$$

and

$$p(a) = \frac{\mu_L}{\mu_I - \mu_L} (\exp(-\mu_L a) - \exp(-\mu_I a)).$$

If $\mu = \mu_L = \mu_I$ then

$$k(a) = \mu^2 a \exp(-\mu a),$$

and

$$p(a) = \mu a \exp(-\mu a).$$

The mean generation time is $1/\mu_L + 1/\mu_I$. In the calculations we have chosen $\mu_L = 1$, and $\mu_I = 1/2$. This gives the mean generation time 3.

Case 2: L_i is gamma distributed with shape parameter α and rate parameter $\mu_L \alpha$, and X_i is also gamma distributed with shape parameter β and rate parameter $\mu_I \beta$. The means of the latent and infectious periods are $1/\mu_L$ and $1/\mu_I$. The mean generation time is $1/\mu_L + 1/\mu_I \frac{1+\beta}{2\beta}$ (cf Svensson (2007)). In the calculations we have chosen $\alpha = \beta = 8$, $\mu_L = 1$, and $\mu_I = 9/32$. This gives the mean generation time 3.

Case 3: L_i and X_i are constant. This corresponds to $\alpha = \beta = \infty$. To obtain the same mean latent and generation times as in the two previous cases we chose $L_i \equiv 1$ and $X_i \equiv 4$.

Figure 2.2 illustrates the cohort generation time density for the three cases.

3. Unrelated potential infectors

In this section we will assume that the first $\nu - 1$ infections are unrelated and derive the probability that the i 'th infected infects the v 'th. Let $t = \tau_\nu$.

3.1. Non-random infectivity

Suppose that $\nu - 1$ Poisson processes are running in parallel and that they have non-random intensity functions $v_1(t), \dots, v_{\nu-1}(t)$. Given that an event happens in one of these processes at time t the probability that it occurs in the i 'th process is

$$\frac{v_i(t)}{\sum_j v_j(t)}.$$

(The summation is for $j = 1, \dots, \nu - 1$).

In case all individuals have the same, non-random, infectivity intensity the probability that the i 'th infector is the responsible for the ν 'th infection, at time $t = \tau_\nu$, can be calculated as

$$P_i^a = \frac{s(t)k(t - \tau_i)}{\sum_j s(t)k(t - \tau_j)}. \quad (10)$$

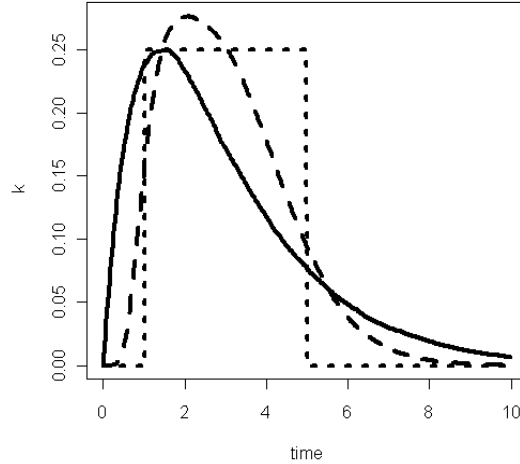


Fig. 1. $k(a)$, for case 1 (solid), case 2 (dashed), and case 3 (dotted).

or equivalently (according to (9))

$$P_i^a = \frac{p(t - \tau_i)}{\sum_j p(t - \tau_j)} = wp(t - \tau_i), \quad (11)$$

where w is a constant such that $\sum P_i^a = 1$.

As will be clear from the following discussion the expression is not valid if the intensity functions are random. In such cases the probabilities (10) and (11) can, at best, be regarded as approximations.

3.2. Random infectivity

We will now consider the possibility that the infectious processes are random.

Due to the assumption of homogeneous mixing and the assumption that all infectious individuals are equally infectious during their infectious periods

$$Z_i = \frac{I_i(t - \tau_i)}{\sum_j I_j(t - \tau_j)} \quad (12)$$

is, conditional on the latent and infectious times, the probability that the infection is transmitted from infector i . Without the conditioning Z_i should be regarded as a random variable.

Since there has to be at least one infector it is a necessary restriction that

$$\sum_j I_j(t - \tau_j) > 0.$$

If there is no other relation between the potential infectors this is the only restriction that has to be considered.

The probability that the i 'th infected is the infector can be expressed as:

$$\begin{aligned}
P_i^u &= \mathbb{E} \left(Z_i \mid \sum_j I_j(t - \tau_j) > 0 \right) \\
&= \frac{\sum_{r=1}^{\nu-1} \frac{1}{r} \Pr \left(I_i(t - \tau_i) = 1, \sum_{j \neq i} I_j(t - \tau_j) = r - 1 \right)}{P(\sum_j I_j(t - \tau_j) > 0)} \\
&= p(t - \tau_i) \frac{\mathbb{E} \left(\frac{1}{1 + \sum_{j \neq i} I_j(t - \tau_j)} \right)}{\Pr(\sum_j I_j(t - \tau_j) > 0)} \\
&= p(t - \tau_i) w_i.
\end{aligned} \tag{13}$$

Comparing the probabilities P_i^u and P_i^a we find that

$$\frac{P_i^u}{P_i^a} \sim w_i. \tag{14}$$

Since the factors w_i depend on i the probabilities will differ from those given by (11). Note that $\sum_{j \neq i} I_j(t - \tau_j)$ is the sum of $\nu - 2$ independent random variables. These random variables are stochastically ordered according to the probabilities $p(t - \tau_j)$, which are proportional to $k(t - \tau_j)$. Thus, the sum tends to be large when $k(t - \tau_i)$ is small.

If there is no latent time the ordering of the τ_i 's imply that $k(t - \tau_i)$ decreases and w_i increases with i (cf (8)). As a consequence the probabilities P_i^a are, in this case, too large for long generation times.

There exist an explicit version of the expression (13):

$$P_i^u = p(t - \tau_i) \frac{\left(\sum_{j=0}^{\nu-2} (-1)^j \frac{1}{1+j} \sum_{y_i=0, |y|=j} \prod_{r=1}^{\nu-1} (p(t - \tau_r))^{y_r} \right)}{1 - \prod_{j=1}^{\nu-1} (1 - p(t - \tau_j))}, \tag{15}$$

where $y = (y_1, \dots, y_\nu)$ is a vector of zeros and ones and $|y| = \sum y_j$. (See appendix for a proof).

4. Spread in a transmission tree

A possible scenario is that we know that the infected and the transmission links form a tree with its root at the first infected. This can be the case if the observations come from an study of infectious spread within a family, a school class, or some other small closed population.

If we observe a transmission tree we know that there is at least one of the candidates that is infectious at any time of infection τ_2, \dots, τ_ν . This means that

$$\sum_{j < m} I_j(\tau_m - \tau_j) > 0.$$

for all $2 \leq m \leq \nu$.

A transmission tree may be a part of a larger transmission chain. In order to have a well defined situation we will assume that the observed infections are the first emerging from an initial infector. We will then have another restriction, namely that the potential infectors represent all infections before time t . If this is the case we can also calculate the susceptible proportion of the population. If the population has n susceptible members at the time of the initial infection the assumption of that contacts are taken at random in the population leads to that $s(t) = 1 - (i - 1)/n$ when $\tau_i \leq t < \tau_{i+1}$.

In this setting it is of interest to calculate the probability that the i 'th infected infects the j 'th. Let

$$Z_{ij} = \frac{I_i(\tau_j - \tau_i)}{\sum_m I_m(\tau_j - \tau_m)}. \quad (16)$$

The probabilities are

$$P_{ij}^c = \mathbb{E}(Z_{ij} \mid \sum_{j < m} I_j(\tau_m - \tau_j) > 0 \text{ if } m \leq \nu \text{ and only } \nu \text{ infected up to } \tau_\nu). \quad (17)$$

We have not been able to find any simple closed version. In the following section we suggest a simulation procedure to do the calculations. It turns out that the probabilities (17) will depend on R_0 (via γ). Since the actual infectivity in a population also depends on the proportion, $s(t)$, of susceptible individuals the probabilities will also depend on the population size.

4.1. A simulation procedure

We first assume that the infectivity functions, $(\kappa_1, \dots, \kappa_{\nu-1})$ for the first $\nu - 1$ infected in the chain are known. The total infectious force at time $\tau \leq \tau_\nu$ is

$$U(\tau) = \gamma s(\tau) \sum_{i=1}^{\nu-1} \kappa_i(\tau - \tau_i).$$

We start by deriving a density for $\xi = (\tau_2, \dots, \tau_\nu)$. Infections occur according to a Poisson process with the random intensity process U . It is random because it depends dynamically on when previous infections occurred. In fact, the process that counts the number of infections from the initiation and forwards is a martingale. From general theory, (cf Bremaud (1981), pg 226) it follows that the density (related to a standard Poisson process) of the random vector ξ is

$$\prod_{i=2}^{\nu} U(\tau_i) \exp\left(\int_0^{\tau_\nu} (1 - U(\tau)) d\tau\right).$$

The probability that it is the i 'th infector that causes the j 'th infection is

$$\frac{\gamma s(\tau_j) \kappa_i(\tau_j - \tau_i)}{U(\tau_j)} = \frac{\kappa_i(\tau_j - \tau_i)}{\sum_m \kappa_m(\tau_j - \tau_m)}.$$

If we take into consideration that the infectivity functions are random we find that the conditional expectations can be expressed as

$$P_{ij}^c = \frac{\mathbb{E} \left(\frac{\gamma s(\tau_j) \kappa_i(\tau_j - \tau_i)}{U(\tau_j)} \prod_{k=2}^{\nu} U(\tau_k) \exp\left(-\int_0^{\tau_\nu} U(\tau) d\tau\right) \right)}{\mathbb{E} \left(\prod_{i=2}^{\nu} U(\tau_i) \exp\left(-\int_0^{\tau_\nu} U(\tau) d\tau\right) \right)}.$$

Let $(\kappa_1^r, \dots, \kappa_{\nu-1}^r)$, $r = 1, \dots, m$ be m sets of simulated infectivity force functions, and let U^r be the corresponding total infectivity functions. We can estimate P_{ij}^c with

$$\hat{P}_{ij}^c = \frac{\sum_{r=1}^m \frac{\gamma s(\tau_j) \kappa_i^r(\tau_j - \tau_i)}{U^r(\tau_j)} \prod_{k=2}^{\nu} U^r(\tau_k) \exp\left(-\int_0^{\tau_\nu} U^r(\tau) d\tau\right)}{\sum_{r=1}^m \prod_{i=2}^{\nu} U^r(\tau_i) \exp\left(-\int_0^{\tau_\nu} U^r(\tau) d\tau\right)}.$$

Observe that if some $U^r(\tau_j) = 0$, $j = 2, \dots, \nu$ then there can not (with probability 1) exist a chain with the given times of infection. Simulated values of the infectivity force function that leads to this will not give any contribution to the estimate.

For the models described in section 2 it is enough to know the latent and infectious times to find the infectivity functions. Let $W = (L_1, X_1, \dots, L_{\nu-1}, X_{\nu-1})$. Furthermore let $J_i(\tau, W)$ equal 1 if the i 'th infected is infectious at time τ , which it is when $\tau_i + L_i \leq \tau < \tau_i + L_i + X_i$, and 0 otherwise. The number of the potential infectors that are infectious at time τ is $J(\tau, W) = \sum J_i(\tau, W)$, and $U(\tau) = \gamma s(\tau) J(\tau, W)$.

Let a_{ij} be the total time the i 'th infector is infectious before the j 'th infection, then

$$a_{ij} = \max(\min(\tau_j - L_i - \tau_i, X_i), 0).$$

Of course a_{ij} can only be positive if $i < j$. Since $s(\tau_j) = 1 - (j-2)/n$ it follows that the sum of all infectious times up till time τ_ν can be calculated as

$$T(W) = \sum_{i=1}^{\nu-1} \sum_{j=1}^{\nu-1} \left(1 - \frac{j-1}{n}\right) (a_{i,j+1} - a_{i,j}).$$

Now let W^1, \dots, W^m be a sequence of independent simulated random elements reflecting $\nu - 1$ latent and infectious times. An estimate of the probability that the i 'th infector is the one who infects the j 'th at time τ_j is then estimated by

$$\frac{\sum_{r=1}^m \frac{J_i(\tau_j, W^r)}{J(\tau_j, W^r)} \prod_{k=2}^{\nu} J(\tau_k, W^r) \exp(-\gamma T(W^r))}{\sum_{r=1}^m \prod_{k=2}^{\nu} J(\tau_k, W^r) \exp(-\gamma T(W^r))}.$$

In the following calculations we have chosen to make sufficiently many simulation so that

$$\sum_{i=1}^m \mathbb{I}(\prod_{j=2}^{\nu} J(\tau_j) > 0) = k_m,$$

where k_m is a predesigned number. Thus the simulations have to produce k_m possible chains.

5. Numerical examples

The expressions P_i^a , P_i^u and P_{ij}^c depend on ν and $(\tau_1, \dots, \tau_\nu)$ as well as on the assumed cohort generation time. P_{ij}^c also depends on γ or equivalently on R_0 .

To illustrate this we first consider a situation where there are only three infected. The second example illustrates a more complicated situation with a longer observed transmission tree.

5.1. Two possible infectors

The probability that the first infected is the true infected will depend both on $\tau_3 = t$, i.e., the time when the third infected was infected and $\tau_2 = s$, the time for the second infection.

The calculations are simple if there is no latency time and the infectious times are exponentially distributed with mean μ_I . In this case explicit expression of the probabilities that the initial infected infects the third are

$$P_1^a = \frac{1}{1 + \exp(\mu_I s)},$$

$$P_1^u = \frac{1 - \exp(-\mu_I(t - s))/2}{1 + \exp(\mu_I s) - \exp(-\mu_I(t - s))},$$

and

$$P_{13}^c = 1/2.$$

These three probabilities are concerned with the same event but calculated under different assumptions. Observe that they are all different and ordered as $P_1^u \leq P_1^a \leq P_{13}^c$.

If there is latent periods the calculations are more complicated. We will consider the three cases defined in section 2. We assume that the population is large, so that the fact that one person is infected do not influence the probability of further infections. The probabilities, P_1^a and P_1^u are illustrated in figure 2. The difference between P_1^a and P_1^u is relatively small but depends heavily on the generation time model used. If there is no individual variability $P_1^a = P_1^u$.

Figures 3 and 4 illustrate the probabilities in case the infections form a tree. It is seen that the influence of R_0 is considerable in cases where there is a substantial individual variability. As can be expected it is less in case 2 than in case 1. In case 2 the relation is illustrated only for $\tau_2 > 0.4$. The probability for smaller time distance between the first and second infection will be small. If there is no individual variability, as in case 3, $P_1^a = P_1^u = P_{13}^c$.

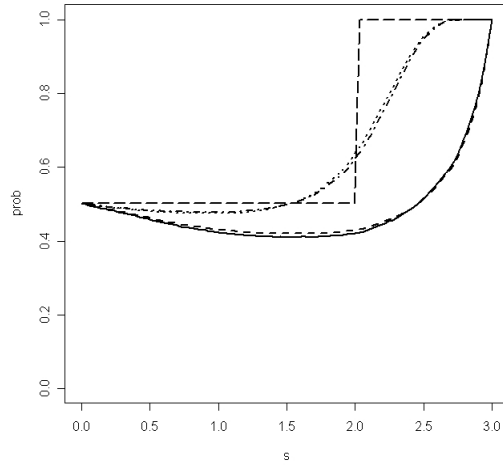


Fig. 2. Probability that the first infected has infected the third infected as a function of the time of infection for the second infected when the third infection occurs at time $t = 3$. P_1^a case 1 (solid) and case 2 (dashed). P_1^u case 1 (dotted) and case 2 (dotdashed). For case 3 $P_1^a = P_1^u$ (longdash).

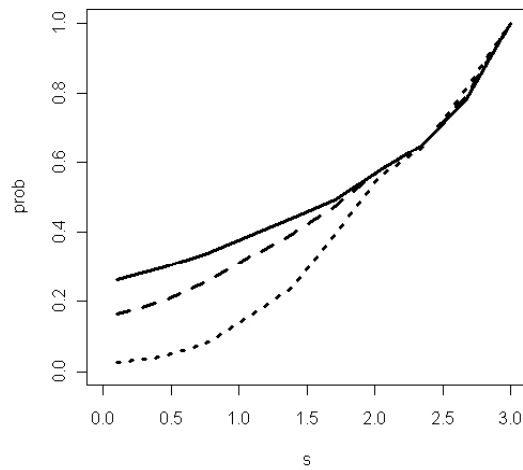


Fig. 3. Probability that the first infected has infected the third infected as a function of the time of infection for the second infected ($s > 0.4$) when the third infection occurs at time $t = 3$. P_{13}^c for $R_0 = 1.2$ (solid), 2 (dashed) and 4 (dotted). Model as in case 1.

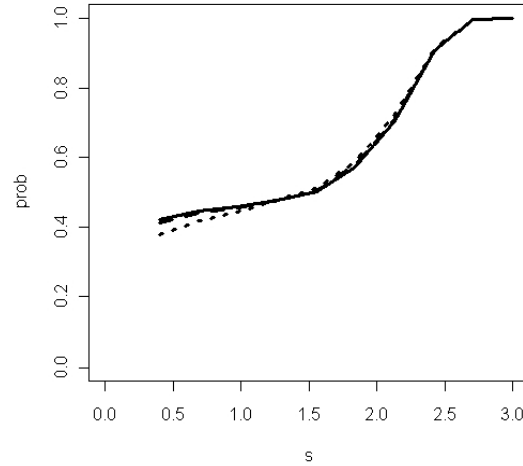


Fig. 4. Probability that the first infected has infected the third infected as a function of the time of infection for the second infected ($s > 0.4$) when the third infection occurs at time $t = 3$. P_{13}^c for $R_0 = 1.2$ (solid), 2 (dashed) and 4 (dotted). Model as in case 2.

5.2. Observation of a transmission tree

To illustrate the use of the calculations in a more complicated situation we will use a simulated example. The simulated epidemic takes place in a large population. The spread is simulated with a model as in case 2 described above and with $R_0 = 2$. The times of infection are

$$\begin{aligned} & (\tau_1, \dots, \tau_{10}, \tau_{11}) \\ & = (0, 0.751, 1.622, 3.009, 3.248, 3.290, 3.527, 3.963, 4.158, 4.371, 4.425). \end{aligned}$$

In sequence the infectors were (the first infection comes from outside the population)

$$(-, 1, 2, 3, 3, 3, 4, 7, 4, 8, 4).$$

This means that the eleventh infection occurred at $t = 4.425$ and that the real infector in this simulated example was the fourth infected.

We can now calculate the probabilities that the i 'th infected infected the j 'th infected ($i < j$) given that the eleven observed infections form a tree. The probabilities for case 1 are presented in table 1, for case 2 in table 2, and for case 3 in table 3. The values of P_{ij}^c are derived from simulations as described above with $k_m = 5.000$ and with $R_0 = 2$.

In order to illustrate the importance of R_0 for the probabilities table 4 gives the estimates of the expected number of infections due to the different infected under the different models and for different values of R_0 . These numbers are the sums of the probabilities related to the i 'th infected.

Table 4. Estimated expected number of secondary infections due to i 'th infected infected

model	R_0	infector									
		1	2	3	4	5	6	7	8	9	10
1	1.2	3.13	2.01	1.86	1.02	0.61	0.61	0.44	0.20	0.10	0.01
1	2	2.88	2.03	1.94	1.08	0.64	0.61	0.46	0.22	0.10	0.01
1	4	2.23	2.02	1.40	1.18	0.66	0.66	0.48	0.25	0.11	0.02
2	1.2	3.59	2.39	2.20	0.68	0.48	0.43	0.22	0.01	0.00	0.00
2	2	3.54	2.40	2.23	0.68	0.48	0.43	0.22	0.01	0.00	0.00
2	4	3.38	2.49	2.34	0.68	0.46	0.43	0.20	0.01	0.00	0.00
3	-	4.25	2.25	2.25	0.58	0.33	0.33	0	0	0	0

Table 5. Estimated probabilities that the i 'th infected infected the j 'th infected, using the model as in case 1 with $R_0 = 2$

i	j		
	2	3	4
1	1	0.61	0.24
2	-	0.39	0.34
3	-	-	0.41

5.2.1. The importance of the length of the observed chain

It should be observed that we when calculating the probabilities P_{ij}^v have condition on the event that there is a transmission chain of at least length ν . If we only wish to calculate the probabilities for the $\nu_1 < \nu$ first infections and only condition on the event that at least ν_1 infections have been observed we will get other probabilities. The same is of course true if we know more infections occur after time τ_ν . To illustrate this we will reanalyze the same situation as in section 5.2 using only the four first infections. The estimated probabilities are given in table 5

If there is no individual variations, as in case 3, the probabilities can be read from the upper left-hand corner of table 3.

5.2.2. Simultaneous probabilities

We have calculated marginal probabilities that the i 'th infected infects the j 'th. The events that the i 'th infected infects both the j 'th and the i^* 'th infected infects the j^* 'th are not independent. Thus the simultaneous probability is in general not equal to the product of the corresponding marginal probabilities. However, this will be true if there is no individual variation, as in case 3. To illustrate this we have calculated the probabilities for the six possible chains in the example treated in the previous section with four infected in the tree. These probabilities are given in table 6

For trees with more observations the number of possibilities will soon get overwhelmingly large. For the chain with $\nu = 11$ there are $10!$ possible trees to consider, most of them with very small probabilities.

Observe that the expected number of secondary infections depends on the marginal probabilities only.

Table 6. Estimated probabilities for possible chains with four infected using the model as in case 1 with $R_0 = 2$

chain	prob
(-,1,1,1)	0.20
(-,1,1,2)	0.18
(-,1,1,3)	0.23
(-,1,2,1)	0.05
(-,1,2,2)	0.16
(-,1,2,3)	0.18

6. Discussion

There are several sources of randomness involved in a model for epidemic spread. It is necessary to understand how different sources can influence the probability that an infection is transmitted from a particular infective person.

An important issue is how the observed infections are obtained. We have considered two very different situations. In the first we investigate from where an infection was transmitted under the assumption that there is no connection between the potential infectors. In the second situation we consider that we have a tree of infections and try to calculate the (marginal) probabilities for different transmission links. In both situations we have conditioned on the assumed observation scheme. The conditioning have consequences. It is shown in the numerical examples that if we that the length of the observed transmission tree influences the probability. This is reasonable, since if we know that many individuals in a population eventually are infected we can conclude the early infectors will probably have been powerful transmitters. If they were not there should have been a large chance that the infection tree stopped early.

We have throughout assumed homogeneous mixing, in the sense that there are no prior differentiation between transmission links, i.e., all possible pairs of persons can be involved in a transmission. This also implies that the only information that can be used is the times of infections. In real cases other information may be available, such as family connections or spatial closeness.

Since the analysis is based only on infection times generation times will be essential. By analyzing a special and very simple model we have illustrated that individual variations in how infectivity varies over time play an important part. This will also be the case in more complicated and realistic models. The effect will be larger the more individual variation there is.

A more surprising results may be that in a transmission tree not only the timing of infections but also the strength of infectivity measured by e.g. R_0 is important. This is the case if there is individual variations in infectivity. Heuristically we can understand this as an effect of the conditioning. If the infections are sparse and R_0 is high it can be explained by short infectious times.

References

- Bremaud, P., 1981. Point Processes and Queues, Martingale Dynamics. Springer-Verlag, New York Heidelberg Berlin.

- Carrat, F., Vergu E., Ferguson, N. M., Lemaître M., Cauchemez S., Leach S., Valleron A.-J., 2008 Time lines of infection and disease in human influenza: A review of volunteer challenge studies *American Journal of Epidemiology* 167 (7),775-785.
- Cauchemez, S. and Ferguson, N. M., 2012 Methods to infer transmission risk factors in complex outbreak data. *Journal of the Royal Society Interface* 9 (68),456-469.
- Svensson, Å., 2007. A note on generation times in epidemic models. *Mathematical Biosciences* 208 (1), 300–311.
- Tomba, G. S., Svensson, Å., Asikainen, T., Giesecke, J., 2010. Some model based considerations on observing generation times for communicable diseases. *Mathematical Biosciences* 223 (1), 24–31.
- Wallinga, J., Teunis, P., 2004. Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. *American Journal of Epidemiology* 160 (6), 509–516.

Appendix A

Let J_1, \dots, J_v be independent Bernoulli variables with $P(J_i = 1) = z_i$, $i = 1, \dots, v$, then

$$\mathbb{E} \left(\frac{1}{1 + \sum J_i} \right) = \sum_{r=0}^v \frac{1}{1+r} \sum_{|y|=r} \prod_{i=1}^v (z_i)^{y_i} (1-z_i)^{1-y_i}.$$

The right-hand side is a first degree polynomial in z_1, \dots, z_v . Elementary algebra shows that the expression can be rewritten as

$$\mathbb{E} \left(\frac{1}{1 + \sum J_i} \right) = \sum_{r=0}^v m_r \sum_{|y|=r} \prod_{i=1}^v z_i^{y_i}.$$

In this formula the coefficients, m_r , may be identified by sequentially inserting $Z_i = 1$ when $i \leq r$ and 0 otherwise for $r = 0, \dots, v$. It is found that

$$m_r = \frac{(-1)^r}{1+r}.$$

Appendix B

In the model referred to as case 1 the progress from infection to end of latent time and to the end of infectious time is described by independent exponential distributed random variables. Also time between infections are exponentially distributed. If we describe the progress of the epidemic in terms of such events we find that the inter event times are the smallest of a number of exponentially distributed times and thus also exponentially distributed with intensities that is the sum of the intensities of the ongoing processes.

We will need some notation. Let E denote the event of a new infection, L_i the end of the latent period for infector i , and I_i the end of the infectious period for infector i . Suppose that we now that the first infector infects the second person at time τ_2 . It follows that the first infector is in an infectious period and the remaining distribution of that period is exponentially distributed. The time till a new possible infectious contact is exponentially distributed with intensity γ . We also know that the second infector is in a latent state. The time till the end of the latency period is exponentially distributed with mean $1/\mu_L$ and the following infectious period is exponentially distributed with mean $1/\mu_I$.

In order that a third infection occurs at time $\tau_3 = t + \tau_2$ one of the following sequence of events will have take place: (E) , (L_2, E) , (L_2, I_1, E) , (L_2, I_2, E) , or (I_1, L_2, E) .

Simple, but tedious, calculations yield that the likelihoods of these events are (in order)

$$\begin{aligned} G_1(t) &= \gamma \exp(-(\gamma + \mu_L + \mu_I)t), \\ G_2(t) &= \frac{\gamma \mu_L}{\mu_L - \gamma - \mu_I} (\exp(-2(\gamma + \mu_I)t) - \exp(-(\gamma + \mu_L + \mu_I)t)), \\ G_3(t) &= G_4(t) = \frac{1}{2} \left(\frac{\gamma \mu_I}{\gamma + \mu_I} \exp(-(\gamma + \mu_I)t) \right. \\ &\quad - \frac{\gamma \mu_I \mu_L}{(\gamma + \mu_I)(\mu_L - \gamma - \mu_I)} \exp(-2(\gamma + \mu_I)t) \\ &\quad \left. + \frac{\gamma \mu_I}{\mu_L - \gamma - \mu_I} \exp(-(\gamma + \mu_L + \mu_I)t) \right), \end{aligned}$$

and

$$G_5(t) = \frac{\gamma\mu_I}{\mu_L - \gamma - \mu_I} \exp(-(\gamma + \mu_I)t) - \frac{\gamma\mu_I\mu_L}{(\gamma + \mu_I)(\mu_L - \gamma - \mu_I)} \exp(-\mu_L t) \\ + \frac{\gamma\mu_I}{\gamma + \mu_I} \exp(-(\gamma + \mu_L + \mu_I)t).$$

The first infected person causes the third infections with probability 1 in the first and fourth cases and with probability 1/2 in the second case. Thus the probability that the first infected infects the third if this infection happens time t after τ_2 can be calculated as

$$Z(t) = \frac{G_1(t) + G_2(t)/2 + G_4(t)}{G_1(t) + G_2(t) + G_3(t) + G_4(t) + G_5(t)}.$$

The probabilities P_i^c illustrated in figures 6.1-6.4 can be calculated as $Z(\tau_3 - \tau_2)$.

This kind of argument can be generalized to more than 3 infections in the chain and to models with gamma distributed latent and infectious times. However, the complexity of the calculations soon gets overwhelming.