Average time in developmental sequences

H G Solari and M A Natiello

June 27, 2014

Abstract

We revise the relation between average developmental time and developmental rates in stochastic and deterministic developmental sequences where death is a likely event. We show that what has come to be known as the \textit{Gamma trick} introduces an undesirable shortening of the developmental time when it is performed according to the established habits. We show the proper way of introducing intermediate developmental stages and how this correction accounts for most (if not all) observations in simulations of dengue epidemics where it was claimed that “the time from epidemic onset to peak increases with decreasing variance in the distribution of epidemiological parameters”.

1 Introduction

Gamma distributed maturation or developmental models have been introduced in several biological contexts, quite often associated to compartmental models. They were proposed for insect development [Manetsch, 1976], modeling the exposed and viremic periods of diseases [Keeling and Grenfell, 1997, Lloyd, 2001, Lloyd et al., 2007, Chowell et al., 2007, 2013], bacteria development in chemostats [Wolkowicz et al., 1997] and viral production [Mittler et al., 1998]. The use of Gamma distributions has been justified frequently in the name of realism. However, the realism of the Gamma distribution is more often than not a class of idealism.

The cumulative Gamma probability function reads

$$P(0 < t < T) = \frac{1}{\Gamma(K)} \int_0^{T/\tau} t^{K-1} \exp(-t) dt$$

where $K$ is the shape parameter and $\tau$ is the scale parameter.

The two qualities of the Gamma distribution associated to realism are:

- The probability $P(t \leq 0)$ is zero, meaning that for example an infected person cannot become contagious at a time prior to infection.
- For large shape factors the Gamma distribution converges to a normal distribution $N(K\tau, K\tau^2)$ [Abramowitz and Stegun, 1970]
The realism implied in this sense is rather limited, it actually means that the Gamma distribution can provide a distribution of times which are close to the normal distribution without incorporating small, but unbearable, probabilities for negative times. Thus, the realism incorporated is in line with (and most likely based upon) the prejudice that the phenomena must be described as determinism (contained in the mean) and a nuisance called noise. The Gamma distribution is a distinguished member of the family of (sums of) exponentially distributed random variables.

Recently, the relation between developmental times and exponentially distributed random variables in the mosquito *Aedes (Stegomyia) aegypti* has been explored in relation to experimental data [Aznar et al., 2014]. Most of the medical data available for the duration of the different biologically identified phases of a disease (exposed, contagious, ...) is not in the form of statistics but rather in terms of perceived time-intervals. Statements like: “the asymptomatic period extends between $N$ and $N + E$ days”, are the rule (see e.g., [Newton and Reiter, 1992, WHO, 2008, Lee et al., 2013]). Such numbers are then interpreted by modelers in terms of an average time: in this example we would propose an average time $<T>$ lying somewhere in $[N, N + E]$.

Note that how to translate an observed time interval into a constant input data for a model is itself a modeling choice. One may choose to compute the mean time, the median, the middle point of the interval (i.e., $N + E/2$), etc. Any of these choices will influence the final description in its own way. Since the distributions of times are not known in general, it is sensible to ask whether they are very relevant or not for the overall description. The changes introduced in the dynamics of epidemics when the distribution of times are changed have been the subject of a controversy [Keeling and Grenfell, 1997, Lloyd, 2001, Conlan et al., 2010].

Recently [Chowell et al., 2013] has come back to the subject claiming that there is a relation such that “the time from epidemic onset to peak increases with decreasing variance in the distribution of epidemiological parameters” a conclusion reached for the case of dengue fever using the *Gamma trick* (see the paragraph below discussing the idea).

The goal of this manuscript is to gain control of the “Gamma trick” beyond the habits currently in practice, examining the procedure in relation with the description of epidemic progress. It is perhaps opportune to recall Burks’ study of Peirce’s philosophy of science. According to Burks [Burks, 1946]: An instinctive or habitual reaction cannot be an inference, (Burks quotes Peirce) “If one does not at all know how one’s belief comes about, it cannot be called even by the name of inference”.

2 Exponential races and developmental sequences

Consider an individual that undergoes a process modeled by a sequence of internal stages (i.e., auxiliary states -not yet biologically identifiable- labeling internal steps in the evolution of a phase of the process) and evolving from
stage to stage by jumps taking place with two different rates: (a) the evolution within the process at a rate \( W \) and (b) the cessation of the process (by death) at a rate \( D \). To fix ideas, let us think of an infective illness requiring a certain incubation phase. Let \( p_i(t) \), \( i = 1, \cdots, K \) indicate the probability of death, regardless of where in the cycle it took place (note that there is no “stage zero”, it is just a label to collect all death events). The process ends at label \( K + 1 \) where neither further incubation evolution nor death occurs. \( p_{K+1}(t) \) is then the probability of having completed the process by time \( t \), having moved then to the next phase. In particular, we may think of having completed the incubation period, emerging with an infected individual now capable of provoking contagion.

The Kolmogorov Forward Equation ([Kolmogoroff, 1931]) for the probabilities reads

\[
\begin{align*}
\frac{\partial p_0}{\partial t} &= D \sum_{i=1}^{K} p_i \\
\frac{\partial p_1}{\partial t} &= -(W + D)p_1 \\
\frac{\partial p_i}{\partial t} &= Wp_{i-1} - (W + D)p_i, \quad i = 2 \cdots K \\
\frac{\partial p_{K+1}}{\partial t} &= Wp_K.
\end{align*}
\]

Expressions for \( N \) individuals and solutions to the general equation can be found in [Solari and Natiello, 2013].

The system (2) is at the core of the Gamma trick. For \( K = 1 \) it represents exponentially distributed times of completing the process. Each step in the process results either in death or progress to the next stage, and as such, there is an exponential race [Durrett, 2001] between death and progress. At each stage, the race is won with probability \( D/(W + D) \) by death and with probability \( Q = W/(W + D) \) by progress. The time to the next event given that the individual is in stage \( i \) is exponentially distributed with rate \( W + D \).

The solution of (2) for the initial condition \( p_1(0) = 1, p_0(0) = 0 = p_i(0), i = 2, \cdots, K + 1 \) reads [Solari and Natiello, 2013, Aznar et al., 2014]

\[
\begin{align*}
p_0(t) &= \int_0^t D \sum_{i=1}^{K} p_i(s) \, ds \\
&= \frac{D}{W + D} \sum_{i=1}^{K} \left( \frac{W}{W + D} \right)^{i-1} \frac{1}{(i-1)!} \int_0^{(W+D)t} s^{i-1} \exp(-s) \, ds \\
p_1(t) &= \left( \frac{W}{W + D} \right)^{i-1} \frac{(W + D)^{i-1}}{(i-1)!} t^{i-1} \exp(-(D + W)t) \\
p_{K+1}(t) &= \left( \frac{W}{W + D} \right)^K \frac{1}{(K-1)!} \int_0^{(W+D)t} s^{K-1} \exp(-s) \, ds.
\end{align*}
\]
The time to complete the cycle is 
\[ T = \sum_{i=1}^{K} \tau_i \]
where \( \tau_i \) are independent random variables exponentially distributed with rate \( W + D \). Then, \( p_{K+1}(t) \) is the product of having survived to \( K \) races with individual probability \( Q \), times the cumulative distribution for the total time of the process, conditioned to having reached all the way to the final stage.

**The Gamma trick** The Gamma trick is usually presented in the absence of mortality. In such a case, the average time for the process is
\[
<T> = \int_0^\infty t \frac{dp_{K+1}(t)}{dt} \ dt = \frac{1}{W(K-1)!} \int_0^\infty s^K \exp(-s) \ ds = \frac{K}{W} = K <\tau>.
\]
The variance of \(<T>\) is \( \sigma^2 = K/W^2 \) while \( \sigma/<T> = 1/\sqrt{K} \) is the coefficient of variation. This coefficient estimates the width of the distribution. At this point, it is argued that the only biological input is the perceived time of completion for the phase in question, while the stages \( K \) are just a modeling instrument. Clearly, any value of \( K \) can in the end fit the observed completion time by properly adjusting \( W \). Indeed, through \( 1/\Theta = <T> = K/W \) we obtain the relation \( W(K) = K\Theta \) (the new notation stresses the approach in which the constant \( \Theta \) comes from the biological input, while \( K \) and \( W(K) \) are modeling instruments). It is further argued that the researcher still has the “liberty” of adjusting \( K \) (thereof the word trick). In this way, some of the influence of choosing different shape parameters can be explored for a given average time taken from medical or biological information.

While the above discussion is standard, care must be taken to extend it to the case where mortality is not negligible. When mortality (or other event competing with progress) is present, the average time in a compartment becomes \( 1/(D + W) \) and consequently \( <T> = K/(W + D) \). Hence, if \( <T> = 1/\Theta \) is taken to be constant (experimental input), the rate \( W \) changes when changing the number of steps \( K \) as
\[
W(K) = K\Theta - D. \tag{4}
\]
In other words, the experimental input of average evolution times for a phase of the disease is not represented by \( W \) in terms of exponential races, but rather by a mixture of disease evolution rate and death rate.

In a similar way, another model may claim that the constant quantity taken from experimental data corresponds to the median time rather than the mean. This claim forces a different relation between the rates. Let \( m_K \) be the ratio between median time and average time for a Gamma distribution with shape parameter \( K \) (0.69, 0.84, 0.96, 1. for \( K = 1, 2, 8, 50 \) respectively). The claim then is that \( m_K <T> = 1/\Theta \) is the observed median, fixed from experimental input. Together with the relation \(<T> = K/(W + D)\), we obtain
\[
W(K) = m_K K\Theta - D. \tag{5}
\]
Notice that when \( \Theta < D \) (or \( m_K \Theta < D \) in the case of preserving the median) not any positive integer value of \( K \) is allowed, since \( W \) has to be nonnegative.
In other words, when the average life expectancy of the bearing individual is smaller than the average time of the process, a minimal number of steps, $K_{\text{min}}$, is needed to match the empirical observations.

Since the mosquito needs some time to effectively transmit the disease (see [Nishiura and Halstead, 2007]), let us say 12.1 days [Focks et al., 1993], and the change of the mosquito from exposed to infective occurs in about one day, this is, the transition between exposed and infective is fast relative to the duration of the exposed period, while the mortality rate has been often considered to be $1/(10.5\text{days})$ [Focks et al., 1993], a considerable error is introduced using the approximated expression $W(K) = K\Theta$, instead of Eq.(4) or (5) for the dynamics of exposition/infectivity within the mosquito.

For future use we define the quantity $Q$ noticing that:

$$Q \geq \lim_{K \to \infty} \left( \frac{K\Theta - D}{K\Theta} \right)^K = \exp(-12.1/10.5) = 0.316 = Q.$$

3 Application to a dengue model

In the case of a vectorial disease such as dengue fever, the life expectancy of human beings is measured in years while the evolution times of the disease are counted in days. Hence, for the time-range of a dengue epidemics, it is safe to assume fixed human population (no births, no deaths). For a situation where the climate is roughly uniform along the year, it is possible to consider also the mosquito population constant. However, birth and death cannot be neglected in this case, since the life span of a mosquito is considerably smaller than the evolution times of the disease.

Let us consider a SEIR-SEI model for dengue. The human population is divided as usual in (the fractions of) susceptible individuals, $S_h$, exposed $E_h$, infective, $I_h$, and recovered, $R_h$. Mosquitoes (more precisely the fractions of mosquito population) can belong to the classes of susceptible mosquitoes, $S_m$, $K$ stages of exposed mosquitoes, $E_{mk}$ and the infective mosquitoes, $I_m$. Disregarding birth and mortality of human population, we have

$$\frac{dS_h}{dt} = -\beta S_h I_m$$

$$\frac{dE_h}{dt} = \beta S_h I_m - aE_h$$

$$\frac{dI_h}{dt} = aE_h - rI_h$$

$$\frac{dR_h}{dt} = rI_h$$

$$\frac{dS_m}{dt} = D - \beta I_h S_m - DS_m$$

$$\frac{dE_{m1}}{dt} = \beta I_h S_m - (W(K) + D)E_{m1}$$

$$\frac{dE_{mj}}{dt} = W(K)E_{m(j-1)} - (W(K) + D)E_{mj}$$
\[
\frac{dI_m}{dt} = W(K)E_{mK} - DI_m
\]

(13)

with rates \(W(K)\) given by Equations (4) or (5) and parameter values (taken from [Chowell et al., 2013] to facilitate the comparison of results) \(\Theta = 1/(12d),\ D = 1/(10.5d),\ a = 1/(5.5d),\ r = 1/(5d),\ \beta = M\beta',\ \beta' = 0.7/d\) (inverse time unit is missing in [Chowell et al., 2013]) and \(M = 3\). We assume that birth and death rates in mosquitoes match each other, so that mosquito population is also constant. \(M\) is the number of mosquitoes per human (the related quantity \(M'\) will be used below).

The way in which the extrinsic exposed period affects the evolution of the epidemic can be grasped as follows. Eq. (6) can be formally integrated as

\[
S_h = \exp \left(-\int_0^t \beta I_m(s)ds\right).
\]

(14)

In turn, the probability of a mosquito to be infective at time \(t\) having acquired the virus at time 0 (integrating (11:13)) is nothing but the process described in Equation (2) except that the last compartment presents mortality. Using Eqs. (3),

\[
I_m(t) = E_{m1}(0)\frac{\exp(-Dt)}{\Gamma(K)} \int_0^{Wt} \exp(-s)s^{K-1}ds.
\]

(15)

Recalling that \(\beta = M\beta'\), the asymptotic fraction of humans infected by the originally infected mosquito is

\[
IH = 1 - \exp \left(-\int_0^\infty \beta I_m(s)ds\right) = 1 - \exp \left(-\frac{E_{m1}(0)M\beta'}{D} \left(\frac{W}{W+D}\right)^K\right)
\]

\[
\approx E_{m1}(0)\frac{M\beta'}{D} \left(\frac{W}{W+D}\right)^K.
\]

(16)

Thus, changing the rate \(W\) will not only change the time-distribution of the extrinsic cycle, but it will, as well, modify the probability of a mosquito transmitting the disease. We can find a work-around for this unwanted side-effect of adopting the Gamma trick. Usually, the total number of mosquitoes - or in this situation \(M\), the number of mosquitoes per human - is not well estimated and it is also used as an adjustable parameter ([Fernández et al., 2013, Chowell et al., 2013]). We then disaggregate the \(K\)-dependency in \(IH\) by modifying the value of \(M\) so that Eq. (16) is kept constant for all \(K\) values and all methods (preserving mean, median and the raw gamma trick). For the present situation we proceed as follows:

\[
\mathcal{M} = \frac{M}{Q} \left(\frac{W}{W+D}\right)^K,
\]

(17)

\[
IH \approx E_{m1}(0)\frac{M\beta'}{D} \left(\frac{W}{W+D}\right)^K = E_{m1}(0)\frac{\mathcal{M}\beta' Q}{D}.
\]

(18)
Discussing how $I_h(t)$ changes when some parameter is changed, is analogous to the discussion on partial derivatives, one needs to specify what objects are changing and what objects are kept constant. The decision can only be taken considering the meaning to be attributed to the objects (or partial derivatives). In this case, changing the number of steps $K$ while keeping $M$ constant speaks of a situation in which the number of mosquitoes is known by some direct method. In contrast, keeping $M$ constant we examine the situation in which the basic reproductive number $R_0$ of the epidemic is used to infer the mosquito population, which is the case more often than not. The constant $M$ is the mosquito contribution to $R_0$.

![Figure 1: Comparison of raw (red), constant mean (blue) and constant median (black) Gamma tricks for $K = 2, 8, 50$. The basic reproductive number has been kept constant in all cases by keeping $M = 3$ constant, cf. Eq.(17).](image)

In Figure 1 we show the results of applying the Gamma trick to the model of Equations (6:13) in different ways, where in all cases Eqs.(17,18) are used, setting the value $M = 3$. The plots display $I_h(t)$, the fraction of infected humans as a function of time. The red curves correspond to letting $W(K) = K \Theta$, disregarding the fact that the mortality $D$ is not negligible (raw results). Note that the red curves are flawed by inconsistency since $D$ cannot be neglected with
this choice of parameters. The blue curves correspond to Eq.(4) while the black curves display Eq.(5). Blue and black curves become indistinguishable at the scale of the plot for $K$ around 50. Recall also that in this example $K_{\text{min}} = 2$.

4 Conclusions

In conclusion, the correct matching between experimental values and the model requires to properly adjust the rates to the experimental values, taking particular care when the characteristic time for death is comparable to the exposed period. This can be done by keeping the average time or the median time (as discussed here), or any other significant quantity, constant. When this is done, the effect of faster or slower epidemic growth associated to changing the number of steps in the Gamma trick is sensibly reduced (in the Figure, peak position moves less than 2% between $K = 8$ and $K = 50$ fixing the median and slightly more fixing the mean, as compared with almost 8.6% for the incorrect -raw- calculation). The $K$-dependency of the peak position, especially for the median, is comparable to the errors introduced by replacing observed experimental time-intervals with some indicative figure (e.g., the median).

To claim that “the time from epidemic onset to peak increases with decreasing variance in the distribution of epidemiological parameters” appears now as an exaggerated claim, stemming from a misunderstanding. Using the median to represent observed experimental time-intervals, while keeping $M$ constant, the influence of $K$ in the final epidemic peak description is negligible in as much the number of mosquitoes remains a fitting parameter.

References


Victoria Romeo Aznar, María Sol De Majo ans Sylvia Fischer, Mario A Natillo, and Hernán G Solari. A model for the development of *Aedes (Stegomyia) aegypti* (and other insects) as a function of the available food. Available from the authors, 2014.


Gerardo Chowell, R. Fuentes, A. Olea, X. Aguilera, H. Nesse, and J. M. Hyman. The basic reproduction number $r_0$ and effectiveness of reactive interventions


