Syracuse University
SURFACE

Mathematics Faculty Scholarship | Mathematics

9-23-2011

A Branching Process for Virus Survival

J. Theodore Cox
Syracuse University

Rinaldo B. Schinazi
University of Colorado at Colorado Springs

Follow this and additional works at: http://surface.syr.edu/mat

Part of the Mathematics Commons

Recommended Citation
Cox, J. Theodore and Schinazi, Rinaldo B., "A Branching Process for Virus Survival" (2011). *Mathematics Faculty Scholarship.* Paper 46.
http://surface.syr.edu/mat/46

This Article is brought to you for free and open access by the Mathematics at SURFACE. It has been accepted for inclusion in Mathematics Faculty Scholarship by an authorized administrator of SURFACE. For more information, please contact surface@syr.edu.

arXiv:1109.5124v1 [math.PR] 23 Sep 2011

A Branching Process for Virus Survival

J. Theodore Cox*
Syracuse University

Rinaldo B. Schinazi
University of Colorado, Colorado Springs

Abstract. Quasispecies theory predicts that there is a critical mutation probability above which a viral population will go extinct. Above this threshold the virus loses the ability to replicate the best adapted genotype, leading to a population composed of low replicating mutants that is eventually doomed. We propose a new branching model that shows that this is not necessarily so. That is, a population composed of ever changing mutants may survive.

Key words: quasispecies, branching process, random environment, evolution.
AMS Classification: Primary: 60K37 Secondary: 92D25

1 Introduction.

Compared to other species an RNA virus has a very high mutation rate and a great deal of genomic diversity. Hence, a virus population can be thought of as an ensemble of related genotypes called quasispecies, see Eigen (1971) and Eigen and Schuster (1977). From the virus point of view a high mutation rate is advantageous because it may create rather diverse virus genomes, this may overwhelm the immune system of the host and ensure survival of the virus population, see Vignuzzi et al. (2006). On the other hand, a high mutation rate may result in many nonviable individuals and hurt the quasispecies, see Sanjuan et al. (2004) and Elena and Moya (1999). It seems therefore that mutation rates should be high but not too high. A simple mathematical model makes this point. Consider a virus population having genomes 1 and 2, where genome 1 has a higher replication rate a_1 and genome 2 has a lower replication rate a_2. We suppose that when type 1 individuals replicate, the new individual has a type 1 genome with probability $1-r$ and a type 2 genome with probability r. Type 2 genome individuals do not mutate. The model is then

$$\begin{aligned} \frac{dv_1}{dt} &= a_1(1-r)v_1 \\ \frac{dv_2}{dt} &= a_1 r v_1 + a_2 v_2 \end{aligned} \tag{1.1}$$

where v_i is the number of type i genomes for $i = 1, 2$. This is a variation of a model in Section 8.5 of Nowak and May (2000). A slightly different but perhaps better interpretation of this model is to think of genome 1 as being a specific (high performing) genome and genome 2 as the collection of all the other genomes in the population.

This system of differential equations is easily solved, and one can check that the ratio v_1/v_2 converges as t goes to infinity. It turns out that the limit is strictly positive if and only if $r < 1 - a_2/a_1$. That is, in order for type 1 to be maintained in the population the mutation r needs to be below the threshold $1 - a_2/a_1$. Hence, this model predicts that above a certain mutation threshold faithful replication of the best adapted genotype is compromised. Moreover, there seems to be general agreement in the biology literature that above this threshold the virus population will go extinct, see Eigen (2002) and Manrubia et al. (2010). We propose here a simple stochastic model that shows that this is not necessarily so. In our model the population may survive, even if faithful replication of the best adapted genotype is compromised, with the population being composed of ever changing mutants.

*Supported in part by NSF Grant No. 0803517

Our results may be biologically relevant for the following reason. An important current strategy to fight HIV and other viruses is to try to increase the mutation probability of the virus, see Eigen (2002) and Manrubia et al (2010). This assumes that above a certain mutation threshold the virus will die out. Our model suggests that at least in theory this strategy may not work.

We now describe our continuous time evolution process. Let μ be a probability distribution with support contained in $[0, \infty)$ and which is absolutely continuous with respect to Lebesgue measure, and let $r \in [0, 1]$. Start with one individual at time 0, and sample a birth rate λ from the distribution μ. The individual gives birth at rate λ and dies at rate 1. Every time there is a birth the new individual: (i) with probability $1 - r$ keeps the same birth rate λ as its parent, and (ii) with probability r is given a new birth rate λ', sampled independently of everything else from the distribution μ. We think of r as the mutation probability and the birth rate of an individual as representing the fitness or genotype of the individual. Since μ is assumed to be continuous, a genotype cannot appear more than once in the evolution of the population. For convenience we label the genotypes in the order of their appearance.

Let $Z(t)$ denote the number of individuals alive at time t. We say that the evolution process *survives* if $Z(t) > 0 \ \forall \ t \geq 0$ and otherwise *dies out*. Our main interest is in determining whether survival with positive probability is possible and by what mechanism can survival be achieved.

Theorem 1. *For $0 \leq r \leq 1$ and probability distributions μ on $[0, \infty)$, the evolution process survives with positive probability if and only at least one of the following survival conditions holds:*

$$\mu(\{\lambda : \lambda(1 - r) > 1\}) > 0, \tag{I}$$

$$\int_{\{\lambda : \lambda(1-r) \leq 1\}} \frac{\lambda r}{1 - \lambda(1 - r)} d\mu(\lambda) > 1. \tag{II}$$

The two extreme cases $r = 0$ and $r = 1$ are easy to understand. If $r = 0$ then (II) cannot hold and (I) reduces to $\mu((1, \infty)) > 0$. In this case, no new types are ever produced, the initial branching rate is used forever by all individuals. Conditional on the initial branching rate λ, $Z(t)$ is a linear birth-death process which survives iff $\lambda > 1$. Thus (I) is equivalent to positive probability of survival. When $r = 1$, (I) cannot hold and (II) reduces to $\int \lambda d\mu(\lambda) > 1$. Now each new individual is a new genotype. It is not hard to see that conditional on a given individual's branching rate λ, the total number of offspring of that individual is k with probability

$$\frac{1}{1 + \lambda} \Big(\frac{\lambda}{1 + \lambda} \Big)^k, \quad k = 0, 1, \ldots,$$

with mean λ. Thus the unconditional mean number of offspring of the first individual is $\int \lambda \mu(d\lambda)$, and the total number of individuals that ever live in the evolutionary process is the same as the total progeny in a Galton-Watson process with an offspring distribution which has this mean. The total progeny is infinite with positive probability if and only if this mean is larger than 1, so (II) is equivalent to positive probability of survival.

Condition (I) corresponds to the prediction of the differential equation model (1.1). That is, below a certain threshold for the mutation probability the virus can survive because a well adapted (i.e. high λ) fixed genotype can survive. However, if (I) fails it is still possible to have survival by (II). In this case survival holds because of a growing "cloud" of ever changing mutants of low replicative ability.

Observe that for any $\epsilon > 0$ and r in $[0, 1)$ there are distributions μ for which (I) holds but $\int \lambda d\mu(\lambda) < \epsilon$. This shows that the behavior of our evolution process is drastically different from the classical Galton-Watson process in homogeneous or random environments. For these processes survival is possible if and only if the expected offspring (or a closely related expectation) is large enough (see Harris (1989) for homogeneous environments and Smith and Wilkinson (1969) for random environments).

It is clear that if the support of μ is unbounded then (I) holds for all $r < 1$, so for interesting examples we consider distributions with compact support. Among these distributions a natural family to consider is the uniform distribution on $[0, a]$, $a > 0$. As the following shows, this class exhibits all possible types of survival behavior depending on the exact values of a and r.

Corollary 1. *Let μ be the uniform distribution on $[0,a]$, $a>0$. If $0<a\le 1$ then the evolution process dies out a.s. for all $r\in[0,1]$, while if $a>2$ the evolution process survives with positive probability for all $r\in[0,1]$. If $a=2$ then the evolution process dies out a.s. for $r=1$ and survives with positive probability for all $r\in[0,1)$. If $1<a<2$ then there exists $r_c\in(1-\frac{1}{a},1)$ such that*

(a) *If $r<1-\frac{1}{a}$ then* (I) *holds and the evolution process survives with positive probability.*

(b) *If $1-\frac{1}{a}\le r<r_c$ then* (II) *holds and the evolution process survives with positive probability.*

(c) *If $r\ge r_c$ then the evolution process dies out a.s.*

In words, whether the population goes extinct when the mutation rate is above a certain threshold depends crucially on the value of a. If $a>2$ there is no such threshold: the population survives for any mutation probability r. Note also that for $1<a<2$ there are two distinct thresholds: $1-\frac{1}{a}$ and r_c. If $r<1-\frac{1}{a}$ a well adapted genome may survive forever while if $1-\frac{1}{a}\le r<r_c$ no fixed genome can survive forever. In this regime the population survives as a growing cloud of ever changing mutants. Finally, if $r\ge r_c$ the population goes extinct.

2 Proof of Theorem 1

Proof of Theorem 1. Recall that we start with a single genotype 1 individual at time 0. Let X_t be the number of type 1 individuals alive at time t. Conditional on the initial branching rate λ, X_t is a birth-death process with individual birth rate $\lambda(1-r)$ and death rate 1. In particular, it is well known (see Chapter 4 of Karlin and Taylor (1975)) that it survives with positive probability if and only if $\lambda(1-r)>1$, and that

$$E(X_t|\lambda)=\exp((\lambda(1-r)-1)t). \tag{2.1}$$

Integration of the condition $\lambda(1-r)>1$ with respect to μ gives

$$P(X_t\ge 1\ \forall\ t>0)>0 \text{ iff } \mu(\{\lambda:\lambda(1-r)>1\})>0.$$

Now let Y_t be the number of different genotypes born up to time t that are offspring of genotype 1 individuals. Then $Y_t\uparrow Y_\infty$ as $t\to\infty$, the total number of different genotypes ever produced by genotype 1 individuals. Note that if $r>0$ then $Y_\infty<\infty$ if and only if $X_t=0$ eventually. For $h>0$ it is easy to see that

$$E(Y_{t+h}-Y_t|\lambda,X_t)=\lambda rhX(t)+o(h) \text{ as } h\downarrow 0,$$

from which it follows that

$$\frac{d}{dt}E(Y_t|\lambda)=\lambda rE(X_t|\lambda)$$

and therefore, using (2.1),

$$E(Y_t|\lambda)=r\lambda\int_0^t E(X_s|\lambda)ds=r\lambda\int_0^t \exp((\lambda(1-r)-1)s)ds.$$

Integration with respect to the measure μ now yields

$$E(Y_t)=\int_0^{+\infty}\int_0^t r\lambda\exp((\lambda(1-r)-1)s)ds\,d\mu(\lambda).$$

By the monotone convergence theorem, $E(Y_t)\uparrow E(Y_\infty)$ as $t\to\infty$. Letting $m(r)=E(Y_\infty)$, it is easy to show using the above that

$$m(r)=\begin{cases}+\infty & \text{if } \mu(\{\lambda:\lambda(1-r)>1\})>0\\ \displaystyle\int_0^{1/(1-r)}\frac{r\lambda}{1-\lambda(1-r)}d\mu(\lambda) & \text{if } \mu(\{\lambda:\lambda(1-r)>1\})=0.\end{cases}$$

We now define the tree of genotypes first introduced by Schinazi and Schweinsberg (2008) for a different model. Assume (I) does not hold, and thus $Y_\infty < \infty$ a.s. Each vertex in the tree will be labeled by a positive integer. There will be a vertex labeled k if and only if an individual of genotype k is born at some time. We draw a directed edge from j to k if the first individual of genotype k to be born had an individual of genotype j as its parent. This construction gives a tree whose root is labeled 1 because all individuals are descendants of the individual of genotype 1 that is present at time zero. The tree of genotypes is a (discrete time) Galton-Watson tree with offspring distribution $p_k = P(Y_\infty = k)$. The mean of the offspring distribution is $m(r)$, and hence, the tree of genotypes is infinite with positive probability if and only if $m(r) > 1$.

To finish the proof, we claim that there are only two ways for the evolution process to survive: either a fixed genotype survives forever with positive probability ((I) holds), or the tree of genotypes is infinite with positive probability ((II) holds). It is clear that if either of these occur then the evolution process survives with positive probability. Suppose now that both (I) and (II) fail. Then with probability one each genotype that ever appears gives birth to only finitely many individuals and also the tree of types is finite a.s. This means that the total number of individuals that ever appear is finite. □

3 Proof of Corollary 1

Let μ be the uniform distribution on $[0, a]$ where $a > 0$. Then (I) is equivalent to $a(1-r) > 1$. If $a(1-r) \le 1$ then

$$m(r) = \frac{1}{a}\int_0^a \frac{r\lambda}{1-(1-r)\lambda}d\lambda. \tag{3.1}$$

The case $0 < a \le 1$. Here $a(1-r) \le 1$ for all $r \in [0,1]$, so (I) does not hold. Furthermore, the fact that $a \le 1$ implies that the integrand in (3.1) is an increasing function of r. Thus for all $r \in [0,1]$,

$$m(r) \le m(1) = a/2 < 1,$$

and hence (II) also fails. For every r the evolution process dies out a.s.

The case $a > 1$. A little calculus shows that

$$m(r) = -\frac{r}{1-r} - \frac{1}{a}\frac{r}{(1-r)^2}\ln(1-a(1-r)), \quad r \in (1-\frac{1}{a}, 1). \tag{3.2}$$

To complete the proof of Corollary 1 we will need the following properties of $m(r)$.

(P1) $m(r)$ is continuous on $(1-1/a, 1]$, $\lim_{r\downarrow 1-1/a} m(r) = \infty$ and $\lim_{r\uparrow 1} m(r) = a/2$.

(P2) If $a \ge 3/2$ then $m(r)$ is strictly decreasing on $(1-\frac{1}{a}, 1)$

(P3) If $1 < a < 3/2$ then there exists $r_a \in (1-\frac{1}{a}, 1)$ such that $m(r)$ is strictly decreasing on $(1-\frac{1}{a}, r_a)$ and strictly increasing on $(r_a, 1)$.

The proof of (P1) is simple and we will omit it. The proofs of (P2) and (P3) require some work, so we will postpone them for now and complete the proof of Corollary 1 assuming (P2) and (P3) have been established. We consider three cases.

(i) If $a \ge 2$ and $r < 1$, then by (P2) $m(r) > m(1) = a/2 \ge 1$, so (II) holds for all $r \in (1-1/a, 1)$. Also, $m(1) = a/2$ implies (II) holds for $a > 2$ but fails for $a = 2$.

(ii) If $3/2 \le a < 2$ then by $m(1) < 1$, and hence by (P1) and (P2) there exists a unique $r_c \in (1-1/a, r_a)$ such that $m(r_c) = 1$. By (P2), (II) holds for $r < r_c$ but fails for $r \ge r_c$.

(iii) If $1 < a < 3/2$ then by (P1) and (P3) $m(r_a) < 1$. It follows that there exists a unique $r_c \in (1-1/a, r_a)$ such that $m(r_c) = 1$, $m(r) > 1$ on $(1-1/a, r_c)$ and $m(r) < 1$ on $(r_c, 1]$.

The proof of Corollary 1 is now complete except for the proofs of (P2) and (P3). At this point it is convenient to change variables. If we define the function

$$g(x) = 1 - x + \frac{1}{a}(x - x^2)\ln(1 - \frac{a}{x}), \quad x \in (a, \infty),$$

then

$$m(r) = g(\frac{1}{1-r}).$$

Moreover, m is increasing (decreasing) on the interval (r_1, r_2) iff g is increasing (decreasing) on the interval $((1-r_1)^{-1}, (1-r_2)^{-1})$. A little calculation gives the first three derivatives of g,

$$\begin{aligned} g'(x) &= -1 - \frac{x-1}{x-a} - \frac{1}{a}(2x-1)\ln(1-\frac{a}{x}) \\ g''(x) &= -\frac{a - 3ax + 2x^2}{x(x-a)^2} - \frac{2}{a}\ln(1-\frac{a}{x}) \\ g'''(x) &= -\frac{a^2 + ax(2a-3)}{x^2(x-a)^3}. \end{aligned}$$

With some additional calculation one can explicitly check that

$$\lim_{x \downarrow a} g'(x) = -\infty, \qquad \lim_{x \to +\infty} g'(x) = 0, \tag{3.3}$$

$$\lim_{x \downarrow a} g''(x) = +\infty, \qquad \lim_{x \to +\infty} g''(x) = 0. \tag{3.4}$$

We also note that by (P1),

$$\lim_{x \downarrow a} g(x) = \infty, \qquad \lim_{x \to +\infty} g(x) = a/2. \tag{3.5}$$

Suppose $a \geq 3/2$. Then $g'''(x) < 0$ for all $x > a$, and hence the function g'' is strictly decreasing on (a, ∞). In view of (3.4), g'' must be positive on $(a, +\infty)$, which implies g' is strictly increasing on $(a, +\infty)$. In view of (3.3), g' must be negative on $(a, +\infty)$, which implies g is strictly decreasing on $(a, +\infty)$. This means that $m(r)$ is strictly decreasing on $(1 - 1/a, 1)$, so (P2) is proved.

Finally, suppose that $1 < a < 3/2$, and put $b = a/(3-2a)$. Then $b > a$, $g''' < 0$ on (a, b) and $g''' > 0$ on (b, ∞). As a consequence, g'' is strictly decreasing on (a, b) and strictly increasing on (b, ∞). In view of (3.4) there must exist a unique $c \in (a, b)$ such that $g'' > 0$ on (a, c) and $g'' < 0$ on (c, ∞). This implies g' is strictly increasing on (a, c) and strictly decreasing on (c, ∞). In view of (3.3) there must exist a unique $x_a \in (a, c)$ such that $g' < 0$ on (a, x_a) and $g' > 0$ on (x_a, ∞). This implies g is strictly decreasing on (a, x_a) and strictly increasing on (x_a, ∞). By setting $r_a = 1 - 1/x_a$ and using the correspondence between the functions m and g we obtain (P3).

References.

M. Eigen (1971) Selforganization of matter and the evolution of biological macromolecules. Naturwissenschaften 58 465-523.

M. Eigen (2002) Error catastrophe and antiviral strategy. PNAS 99 13374-13376.

M. Eigen and P. Schuster (1977) The hypercycle. A principle of self-organization. Part A: emergence of the hypercycle. Naturwissenschaften 64 541-565.

S. F. Elena and A. Moya (1999) Rate of deleterious mutation and the distribution of its effects on fitness in vesicular stomatitis virus. J. Evol. Biol. 12 1078-1088.

T.E.Harris (1989) *The Theory of Branching Processes.* Dover, New York.

S. Karlin and H.M. Taylor (1975) *A First Course in Stochastic Processes, 2nd Ed.* Academic Press, New York.

S.C. Manrubia, E. Domingo and E. Lazaro (2010) Pathways to extinction: beyond the error threshold. Phil. Trans. R. Soc. B 365, 1943-1952.

M. A. Nowak and R.M.May (2000) *Virus dynamics.* Oxford University Press.

R. Sanjuan, A. Moya and S. F. Elena (2004). The distribution of fitness effects caused by single-nucleotide substitutions in an RNA virus. PNAS 1018396-8401.

R.B.Schinazi and J. Schweinsberg (2008) Spatial and non spatial stochastic models for immune response. Markov Processes and Related Fields 14 255-276.

W.L.Smith and W.E.Wilkinson (1969) On branching processes in random environments. The Annals of Mathematical Statistics 40, 814-827.

M. Vignuzzi, J.K. Stone, J.J. Arnold, C.E.Cameron and R. Andino (2006). Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. Nature 439, 344-348