

On a link between a species survival time in an evolution model and the Bessel distributions

Hervé Guiol^a, Fábio P. Machado^b and Rinaldo Schinazi^c

^a*Université Joseph Fourier—Grenoble I*

^b*University of São Paulo*

^c*University of Colorado at Colorado Springs*

Abstract. We consider a stochastic model for species evolution. A new species is born at rate λ and a species dies at rate μ . A random number, sampled from a given distribution F , is associated with each new species and assumed as its fitness, at the time of birth. Every time there is a death event, the species that is killed is the one with the smallest fitness. We consider the (random) survival time of a species with a given fitness f . We show that the survival time distribution depends crucially on whether $f < f_c$, $f = f_c$ or $f > f_c$ where f_c is a critical fitness that is computed explicitly.

1 Introduction

Consider a stochastic model for species evolution in which a new species is born at rate λ and an existing species dies at rate μ . A random number, sampled from a given distribution F , is associated with each new species at the time of birth. We think of the random number associated with a given species as being the *fitness* of the species. These fitnesses are independent of each other and of everything else in the process. Every time there is a death event, the species that is killed is the one with the smallest fitness. We assume F to be an absolute continuous distribution function. In this paper we study the survival time of a given species with fitness f . We show that there is a critical fitness f_c and a sharp phase transition for the survival time of the species. Our analysis is based on a closed connection between our model and random walks.

A similar model, meant to build phylogenetic trees, was introduced in Liggett and Schinazi (2009). A discrete version of this model is studied in Guiol et al. (2011) where a phase transition is shown.

Key words and phrases. Bessel distributions, phase transition, stochastic model for species evolution, survival time.

Received February 2011; accepted August 2011.

2 Main results

Recall we assume that F is an absolute continuous distribution function. This implies that there exists a probability density φ on \mathbb{R} such that

$$F(x) = \int_{-\infty}^x \varphi(u) du.$$

Denote $\text{Supp}(F)$ the support of distribution F :

$$\text{Supp}(F) = \{x \in \mathbb{R} : \varphi(x) > 0\}.$$

Assume that originally there are k species in the process, with associate fitness $f_1 < f_2 < \dots < f_k = f$ in the support of F . Denote τ_f^k the r.v. corresponding to the survival time of the species with fitness f in that context.

Theorem 2.1. *The survival time τ_f^k has a Bessel distribution*

$$\mathbb{P}(\tau_f^k > t) = 1 - \left(\frac{\mu}{\lambda_f}\right)^{k/2} \int_0^t e^{-(\mu+\lambda_f)u} \frac{k}{u} I_k(2\sqrt{\mu\lambda_f}u) du \quad (2.1)$$

with $\lambda_f := \lambda F(f) < \lambda$ and where I_k is the modified Bessel function of the first kind with index k defined by

$$I_k(x) = \sum_{\ell=0}^{+\infty} \frac{1}{(\ell+k)! \ell!} \left(\frac{x}{2}\right)^{2\ell+k}. \quad (2.2)$$

Remark 2.2. Devroye (1986, Chapter IX, Section 7, page 470) includes distribution (2.1) among the class of Bessel function distributions. One can also find it in Feller (1968, Chapter II, Section 9, Problem 15, page 65) without a specific name.

Remark 2.3. Whenever $\lambda_f \leq \mu$, $\tau := \tau_f^k$ has probability density

$$\varphi_\tau(t) = \left(\frac{\mu}{\lambda_f}\right)^{k/2} e^{-ct} \frac{k}{t} I_k(2\sqrt{\mu\lambda_f}t) \quad \text{for all } t > 0$$

and some $c > 0$. In the case $\lambda_f > \mu$ the previous function is not a density probability since

$$\int_0^\infty \left(\frac{\mu}{\lambda_f}\right)^{k/2} e^{-ct} \frac{k}{t} I_k(2\sqrt{\mu\lambda_f}t) dt = \left(\frac{\mu}{\lambda_f}\right)^{k/2} \left(\frac{\mu}{\lambda_f}\right)^{k/2} = \left(\frac{\mu}{\lambda_f}\right)^k < 1.$$

Remark 2.4. The survival time τ_f^k is not affected by living species with fitness above f . Besides, we have

$$\tau_f^k = \sum_{j=1}^k \tau_f^j - \tau_f^{j-1},$$

being $\tau_f^0 = 0$. Observe that τ_f^k are a.s. finite stopping times and from the Strong Markov property $(\tau_f^j - \tau_f^{j-1})_{1 \leq j \leq k}$ is an i.i.d. sequence of r.v. with the distribution of $\tau_f := \tau_f^1$.

Though formula (2.1) gives the exact distribution of the survival time, it is not straightforward to come to a conclusion from it. The next result sheds light in the phase transition property of our model.

Corollary 2.5. *With the previous notation,*

(a) *If $\lambda_f < \mu$, then*

$$\mathbb{P}(\tau_f^k > t) \sim C_k \frac{e^{-\gamma t}}{t^{3/2}}$$

with $C_k = \frac{k}{2\sqrt{\pi}} \left(\frac{\mu}{\lambda_f}\right)^{k/2} (\mu\lambda_f)^{-1/4} (\sqrt{\mu} - \sqrt{\lambda_f})^{-2}$ and $\gamma = (\sqrt{\mu} - \sqrt{\lambda_f})^2$.

(b) *If $\lambda_f > \mu$, then*

$$\mathbb{P}(\tau_f^k = +\infty) = 1 - \left(\frac{\mu}{\lambda_f}\right)^k;$$

$$\mathbb{P}(t < \tau_f^k < +\infty) \sim C_k \frac{e^{-\gamma t}}{t^{3/2}}.$$

(c) *If $\lambda_f = \mu$, then*

$$\mathbb{P}(\tau_f^k > t) \sim k(\pi\mu t)^{-1/2}.$$

Remark 2.6. Note, from Corollary 2.5, that if $\lambda > \mu$, there is a phase transition in f . A species born with a fitness lower than

$$f_c := F^{-1}(\mu/\lambda) \quad (2.3)$$

dies out exponentially fast, while a species with a fitness greater than f_c has a positive probability of surviving forever. The larger λ/μ (recall that F^{-1} is non-decreasing) the more welcoming the environment is to new species. If $\lambda/\mu < 1$, all species will die exponentially fast. On the other hand, if λ/μ is large, then even species with relatively low fitness will make it.

Up to now we have discussed the survival of a species with a given fitness f . It is particularly relevant to derive some information about the distribution of these surviving species. Suppose that $\lambda > \mu$ and let L_t and R_t be the sets of species alive at time t whose fitness is respectively lower than f_c and higher than f_c . Since each fitness that has appeared up to time t will not show up again a.s., we can identify each species with its fitness and think of L_t and R_t as sets of points in $(-\infty, f_c)$ and (f_c, ∞) , respectively. The next result states a straightforward application of the main result of Guiol et al. (2011).

Proposition 2.7. *Suppose that $\lambda > \mu$. Then*

- (a) *The number $|L_t|$ of species whose fitness is below f_c is a null recurrent birth and death process. In particular, the set L_t is empty infinitely often with probability one.*
- (b) *Let $f_c < a < b$, then*

$$\lim_{t \rightarrow \infty} \frac{1}{t} |R_t \cap (a, b)| = \frac{\lambda(F(b) - F(a))}{\lambda + \mu} \quad a.s.$$

Remark 2.8. Observe that from Ben-Ari et al. (2011) it would be also possible to get a Central Limit Theorem and a Law of the Iterate logarithm for R_t .

3 Proofs

3.1 Construction of the process

The construction uses ideas from the Harris Graphical method for Markov Processes and basically takes advantage from projection properties of a bidimensional Poisson process with rate 1. In the sequel we construct a bivariate process $Z_t = (Z_t^1, Z_t^2)$ in which Z_t^1 will represent the number of living species at time t , and Z_t^2 will be the set of associated living fitness: In particular, $|Z_t^2| = Z_t^1$.

Let M be a two dimensional Poisson process with rate 1 on $\mathbb{R}^+ \times \mathbb{R}$. For notational convenience we will identify the x -line of the plane as the time line.

Suppose we start the process with $k \geq 1$ species, let f_1, \dots, f_k be k independent random variables with F distribution, independent from M .

Let $T_0 = 0$ and $Z_0 = (Z_0^1, Z_0^2) = (k, \{f_1, \dots, f_k\}) \in \mathbb{N} \times \mathbf{S}$, where \mathbf{S} is the set of finite subsets of real numbers in $[0, 1]^{\mathbb{N}}$.

Define

$$T_1 = \inf\{t > 0 : M([0, t] \times [0, \lambda + \mu]) > 0\}, \quad (3.1)$$

that is, the first time $t \in \mathbb{R}^+$ that a Poisson mark falls into the strip $\mathbb{R}^+ \times [0, \lambda + \mu]$. Denote by (T_1, Y_1) the coordinate of the Poisson mark realizing the infimum in (3.1). Observe that from the Poisson process properties Y_1 is a uniform $[0, 1]$ r.v. independent of T_1 .

- Whenever $Y_1 \in [0, \lambda]$, then let $f_{k+1} = F^{-1}(Y_1/\lambda)$ (observe that f_{k+1} is also a r.v. with F law and independent of T_1) and let

$$Z_{T_1} = (Z_0^1 + 1, Z_0^2 \cup \{f_{k+1}\}) = (k + 1, \{f_0, \dots, f_{k+1}\});$$

this will represent the birth of a new species;

- else, whenever $Y_1 \in]\lambda, \lambda + \mu]$, let

$$Z_{T_1} = (Z_0^1 - 1, Z_0^2 \setminus \{\min\{Z_0^2\}\}) = (k - 1, \{f_0, \dots, f_k\} \setminus \{\min\{f_i : 1 \leq i \leq k\}\});$$

this will represent the death of the weakest species.

For all $t \in [0, T_1[$ denote $Z_t = Z_0$. We have thus constructed the process Z_t until time T_1 (included).

For all $n \geq 1$ denote by T_n the time of the n th mark of the Poisson process M in the strip $\mathbb{R}^+ \times [0, \lambda + \mu]$, that is,

$$T_n = \inf\{t > T_{n-1} : M([0, t] \times [0, \lambda + \mu]) > 0\}.$$

Suppose the process Z_t is constructed up to time T_n , $n \geq 1$. As before, denote by (T_{n+1}, Y_{n+1}) the coordinate of the $n + 1$ st Poisson mark.

- Whenever $Y_{n+1} \in [0, \lambda]$, then let $f_{k+n+1} = F^{-1}(Y_{n+1}/\lambda)$ and define

$$Z_{T_{n+1}} = (Z_{T_n}^1 + 1, Z_{T_n}^2 \cup \{f_{k+n+1}\}),$$

- else

$$Z_{T_{n+1}} = (Z_{T_n}^1 - \mathbf{1}_{\{Z_{T_n}^1 > 0\}}, Z_{T_n}^2 \setminus \{\min\{Z_{T_n}^2\}\}),$$

with the convention $\min \emptyset = \emptyset$;

then for all $t \in [T_n, T_{n+1}[$ let $Z_t = Z_{T_n}$.

So by induction one can construct the process $(Z_t)_{t \geq 0}$ so that the second coordinate of Z_t , that is, Z_t^2 , represents our fitness process starting with k species.

3.2 An useful coupling

We are interested in the distribution of τ_f^k . One can have this from X_t , a process coupled to the process Z_t , being X_t the number of species whose fitnesses are smaller than f at time t . Observe that, by construction, $X_0 = Z_0^1$. Moreover, at time T_1 ,

- if $Y_1 \in [0, \lambda_f] \cup]\lambda, \lambda + \mu]$, then let $X_{T_1} = Z_{T_1}^1$. Observe that this corresponds to a simultaneous death or to a simultaneous birth with associate fitness less than f for the Z process;
- else (when $Y_1 \in]\lambda_f, \lambda]$) then let $X_{T_1} = X_0$. In this case there is a birth on the Z process with associate fitness bigger than f and nothing for the X process.

As before, define $X_t = X_0$ for all $t \in [0, T_1[$.

For all set A of numbers in $[0, 1]$ denote by

$$\phi_f(A) = \{x \in A : x \leq f\}, \quad (3.2)$$

that is, the set of numbers in A less or equal to f .

Observe that $X_t = Z_t^1$ on $[0, T_1[$, $X_{T_1} = |\phi_f(Z_{T_1}^2)| \leq Z_{T_1}^1$.

For $n \geq 1$ suppose that X_t is constructed up to time T_n .

If $X_{T_n} \neq 0$, then $|\phi_f(Z_{T_n}^2)| = X_{T_n}$.

- If $Y_{n+1} \in [0, \lambda_f]$ (recall that $f_{k+n+1} := F^{-1}(Y_{n+1}/\lambda) \leq f$), define

$$X_{T_{n+1}} = X_{T_n} + 1,$$

- if $Y_{n+1} \in [\lambda, \lambda + \mu]$, then define

$$X_{T_{n+1}} = X_{T_n} - 1,$$

- else let $X_{T_{n+1}} = X_{T_n}$;

In case $X_{T_n} = 0$ let $X_{T_{n+1}} = 0$.

This defines a random sequence $(T_n, X_{T_n})_{n \geq 0}$; we define the process $(X_t)_{t \geq 0}$ as $X_t = X_{T_n}$ for all $t \in [T_n, T_{n+1}[$.

The proof of Theorem 2.1 relies on the following lemma.

Lemma 3.1. *For any $k \geq 1$,*

$$\{\tau_f^k > t\} = \{X_t > 0\}, \quad (3.3)$$

that is, τ_f^k has the same law as the first passage time to 0 of X_t the simple Bernoulli random walk starting at k with rate $c = \lambda_f + \mu$ and individual steps equal to 1 or -1 with respective probability $p = \lambda_f/c$ and $q = \mu/c$.

Proof. We have $X_0 = Z_t^1 = |\phi_f(Z_t^2)| = k > 0$. From the construction for all $t < \tau_f^k$ we have $\min Z_t^2 \leq f$; this implies $X_t > 0$. Conversely, if $\tau_f^k \leq t$ as $\min Z_{\tau_f^k}^2 > f$, this implies $X_{\tau_f^k} = 0$ and, thus, $X_t^1 = 0$. \square

Proof of Theorem 2.1. Let $(J_n)_{n \geq 1}$ denote the jump times of the process $(X_t)_{t \geq 0}$ and set $J_0 = 0$. The random sequence $(X_{J_n})_{n \geq 0}$ is a simple discrete time random walk on \mathbb{N} with individual steps equal to 1 or -1 with respective probability p and q . Denote by H_0 the first hitting time of 0 of this walk. A standard computation (see, for instance, Grimmett-Stirzaker (2001, (15), page 79)) gives

$$\mathbb{P}(H_0 = n | X_0 = k) = \frac{k}{n} \binom{n}{(n+k)/2} q^{(n+k)/2} p^{(n-k)/2}$$

whenever $n + k$ is even, 0 otherwise. As J_n has a Gamma (see Grimmett-Stirzaker (2001, Chapter 6)) distribution with parameters c and n , this implies that

$$\begin{aligned} \mathbb{P}(X_t = 0) &= \int_0^t \sum_{n=k}^{\infty} \frac{c^n}{(n-1)!} u^{n-1} e^{-cu} \mathbb{P}(H_0 = n | X_0 = k) du \\ &= \int_0^t e^{-cu} \frac{k}{u} \sum_{n=k}^{\infty} \frac{1}{n!(n-k)!} (cu)^{2n-k} q^n p^{n-k} du \\ &= \int_0^t e^{-cu} \frac{k}{u} \sum_{\ell=0}^{\infty} \frac{1}{(\ell+k)! \ell!} (cu)^{2\ell+k} q^{\ell+k} p^{\ell} du \\ &= \int_0^t e^{-cu} \frac{k}{u} \left(\frac{q}{p}\right)^{k/2} \sum_{\ell=0}^{\infty} \frac{1}{(\ell+k)! \ell!} (cu \sqrt{qp})^{2\ell+k} du, \end{aligned}$$

and from the definition of the Bessel function (2.2),

$$\mathbb{P}(X_t = 0) = \int_0^t e^{-cu} \frac{k}{u} \left(\frac{q}{p}\right)^{k/2} I_k(2cu\sqrt{pq}) du. \quad \square$$

As observed in Remark 2.3, whenever $\lambda_f < \mu$, the expression (2.1) gives the density probability of τ_f :

$$\varphi_{\tau_f}(t) = \sqrt{\frac{\mu}{\lambda_f}} e^{-(\mu+\lambda_f)t} \frac{1}{t} I_1(2\sqrt{\mu\lambda_f}t)$$

for $t > 0$, which allows to compute its Moment Generating Function:

$$M(s) = \mathbf{E}(e^{-\tau_f s}) = \frac{2\mu}{\sqrt{(s + \mu + \lambda_f)^2 - 4\mu\lambda_f} + s + \mu + \lambda_f}.$$

This in turns allows us to compute $\mathbf{E}(\tau_f) = \frac{2\mu}{\mu - \lambda_f}$, so that one can see easily that

$$\mathbf{E}(\tau_f^k) = k \frac{2\mu}{\mu - \lambda_f}$$

for all $k \geq 0$.

Proof of Corollary 2.5(a). When $\lambda_f < \mu$ (2.1) reads

$$\mathbb{P}(\tau_f^k > t) = \left(\frac{\mu}{\lambda_f}\right)^{k/2} \int_t^{+\infty} e^{-(\mu+\lambda_f)u} \frac{k}{u} I_k(2\sqrt{\mu\lambda_f}u) du.$$

From Abramowitz and Stegun (1992, page 377, Equation 9.7.1) for x large enough and all fixed v ,

$$I_v(x) \approx \frac{e^x}{\sqrt{2\pi x}} \left(1 - \frac{4v^2 - 1}{8x} + \frac{(4v^2 - 1)(4v^2 - 9)}{2!(8x)^2} - \frac{(4v^2 - 1)(4v^2 - 9)(4v^2 - 25)}{3!(8x)^3} + \dots \right),$$

so that for all $k \geq 1$

$$\frac{e^x}{\sqrt{2\pi x}} \left(1 - \frac{4k^2 - 1}{8x} \right) \leq I_k(x) \leq \frac{e^x}{\sqrt{2\pi x}} \quad (3.4)$$

for x large enough,

$$\begin{aligned} & \frac{1}{2\sqrt{\pi}(\mu\lambda_f)^{1/4}} \frac{e^{-(\sqrt{\mu}-\sqrt{\lambda_f})^2 u}}{u^{3/2}} \left(1 - \frac{4k^2 - 1}{16\sqrt{\mu\lambda_f}u} \right) \\ & \leq \frac{e^{-(\mu+\lambda_f)u}}{u} I_k(2\sqrt{\mu\lambda_f}u) \leq \frac{1}{2\sqrt{\pi}(\mu\lambda_f)^{1/4}} \frac{e^{-(\sqrt{\mu}-\sqrt{\lambda_f})^2 u}}{u^{3/2}}, \end{aligned} \quad (3.5)$$

also for x large enough. Denoting $\gamma = (\sqrt{\mu} - \sqrt{\lambda_f})^2$, observe that

$$\left(\frac{1}{\gamma} \frac{1}{t^{3/2}} - \frac{4k^2 - 1}{2\gamma^2} \frac{1}{t^{5/2}} \right) e^{-\gamma t} \leq \int_t^{+\infty} \frac{e^{-\gamma u}}{u^{3/2}} du \leq \frac{1}{\gamma} \frac{e^{-\gamma t}}{t^{3/2}}. \quad (3.6)$$

Thus,

$$\int_t^{+\infty} \frac{e^{-\gamma u}}{u^{3/2}} du \sim \frac{1}{\gamma} \frac{e^{-\gamma t}}{t^{3/2}},$$

so we have a sharp asymptotic estimate for the integral of the upper bound in (3.5).

For the integral of the lower bound, denoting $\alpha = (4k^2 - 1)/(16\sqrt{\mu\lambda_f})$, just observe that

$$\int_t^{+\infty} \left(1 - \frac{\alpha}{u} \right) \frac{e^{-\gamma u}}{u^{3/2}} du \geq \int_t^{+\infty} \frac{e^{-\gamma u}}{u^{3/2}} du - \frac{\alpha}{\gamma} \frac{e^{-\gamma t}}{t^{5/2}} \quad (3.7)$$

to see that we also have a sharp asymptotic estimate for the integral of the lower bound in (3.5). Besides, both asymptotic estimates agree.

Plugging (3.7) and (3.6) into (3.5) and then into (2.1), we finally conclude that for t large enough

$$\mathbb{P}(\tau_f^k > t) \sim \left(\frac{\mu}{\lambda_f} \right)^{k/2} \frac{1}{2\sqrt{\pi}(\mu\lambda_f)^{1/4}} \frac{k}{(\sqrt{\mu} - \sqrt{\lambda_f})^2} \frac{e^{-(\sqrt{\mu} - \sqrt{\lambda_f})^2 t}}{t^{3/2}}. \quad \square$$

Proof of Corollary 2.5(b). This is immediate from the preceding computations and Remark 2.3. \square

Proof of Corollary 2.5(c). When $\lambda_f = \mu$ (2.1) reads

$$\mathbb{P}(\tau_f^k > t) = \int_t^{+\infty} e^{-2\mu u} \frac{k}{u} I_k(2\mu u) du,$$

using in turn inequalities (3.4) leads directly to the result. \square

Proof of Proposition 2.7. To see this observe that the embedded discrete Markov chain for our process is the *stochastic model of evolution* defined in Guiol et al. (2011) in such a way that $p = \lambda/(\lambda + \mu)$. Consider that whenever the total number of species is 0, the death marks in the construction of the process are ignored so the total number of species stays 0 with probability $\mu/(\lambda + \mu)$. \square

Acknowledgments

Research supported by CNPq (306927/2007-1), Fapesp (2009/18253-7), ANR BLAN07-2184264 and PICS 5470. We thank the anonymous referees for remarks that improved the presentation of the paper.

References

- Abramowitz, M. and Stegun, I. A. (eds.) (1992). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. New York: Dover. Reprint of the 1972 edition. [MR1225604](#)
- Ben-Ari, I., Matzavinos, A. and Roitershtein, A. (2011). On a species survival model. *Elect. Com. in Probab.* **16**, 226–233. [MR2788894](#)
- Devroye, L. (1986). *Non-Uniform Random Variate Generation*. New York: Springer. [MR0836973](#)
- Feller, W. (1968). *An Introduction to Probability Theory and Its Applications, Vol. 2*. New York: Wiley. [MR0228020](#)
- Grimmett, G. and Stirzaker, D. (2001). *Probability and Random Processes*, 3rd ed. New York: Oxford Univ. Press. [MR2059709](#)
- Guiol, H., Machado, F. P. and Schinazi, R. B. (2011). A stochastic model of evolution. *Markov Processes Relat. Fields* **17**, 253–258. [MR2856242](#)
- Liggett, T. M. and Schinazi, R. B. (2009). A stochastic model for phylogenetic trees. *J. Appl. Probab.* **46**(2), 601–607. [MR2535836](#)

H. Guiol
UJF—Grenoble 1/CNRS/Grenoble INP
TIMC-IMAG UMR 5525
Grenoble, F-38041
France
E-mail: Herve.Guiol@imag.fr

F. P. Machado
University of São Paulo
Rua do Matao, 1010
Sao Paulo
Sao Paulo, 05508-090
Brazil
E-mail: fmachado@ime.usp.br

R. Schinazi
University of Colorado at Colorado Springs
1420 Austin Bluffs
Colorado Springs, Colorado 80933
USA
E-mail: rschinaz@uccs.edu