The fitness of the strongest individual in the subcritical GMS model[*]

Carolina Grejo[†]   Fábio Machado[†]   Alejandro Roldán-Correa[†]

Abstract

We derive the strongest individual fitness distribution on a variation for a species survival model proposed by Guiol, Machado and Schinazi [5]. We point out to the fact that this distribution relies on the Gauss hypergeometric function and when $p = \frac{1}{2}$ on the hypergeometric function type I distribution.

Keywords: GMS model; random walk; Gauss hypergeometric function.
AMS MSC 2010: Primary 60J20; 60G50, Secondary 33C05.
Submitted to ECP on September 21, 2015, final version accepted on February 2, 2016.
Supersedes arXiv:1509.04686.

1 Introduction

We consider a discrete time model beginning from an empty set. At each time $n \geq 1$, a new species is born with probability p or there is a death (if the system is not empty) with probability $q = 1 - p$. Let X_n be the total number of species at time n. X_n is a random walk on $\mathbb{Z}_+$ that jumps to right with probability p and jumps to left with probability q. When X_n is at 0 the process jumps to 1 with probability p or stays at 0 with probability $1 - p$. We assign a random number to each new species. This number has a uniform distribution on $[0, 1]$. We think of this number as a fitness associated to each species. These random numbers are independent to each other. When a death occurs, the individual with lowest fitness dies. This model, latter denominated GMS model, was first proposed and studied in Guiol *et al* [5]. Some interesting variations were further studied in Guiol *et al* [6], Ben Ari *et al* [2] and Skevi and Volkov [10].

In Guiol *et al* [5] it is shown that there is a sharp phase transition for $p > 1/2$. For R_n, the set of species with fitness higher than $f_c = \frac{1-p}{p}$ at time n approachs an uniform distribution in the following sense. For $f_c < a < b < 1$

$$\lim_{n \to \infty} \frac{|R_n \cap (a, b)|}{n} = p(b - a) \ \text{a.s.}$$

On the other hand every specie born with fitness less than f_c disappear after a finite (random) time. The set of species present in the system whose fitness is smaller than f_c becomes empty infinitely many times.

[†]Instituto de Matemática e Estatística - Universidade de São Paulo, Brazil.
 E-mail: carolina@ime.usp.br, fmachado@ime.usp.br, roldan.alejo@gmail.com
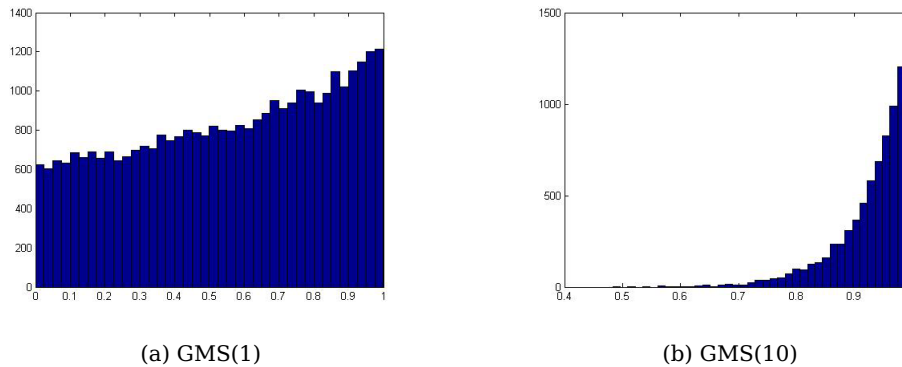
(a) GMS(1)

(b) GMS(10)

Figure 1: Histograms of the fitnesses of the strongest individual on GMS(m) after 200,000 births and deaths for $p = 1/4$.

Here we focus on the case $p \leq 1/2$ in order to understand better the dynamics of this model. In this case, the process X_n is recurrent and the system becomes empty infinitely many times. Therefore it is not interesting to study the distribution of the fitness of the species which are alive on the system in the long run. An interesting point is to study the distribution of the fitness of the strongest individual on each excursion between the epochs when the system becomes empty.

We propose a variation for the GMS model by considering that each time the system becomes empty, a set of m individuals are introduced with independent set of fitness. This variation is meant to reinforce competition among species before the system becomes empty again.

2 Results

We deduce explicitly the distribution of the fitness of the strongest individual on excursions between the epochs when the system becomes empty. The last individual to die before the system becomes empty is the strongest on that excursion because the first ones to die are those individuals with the smallest fitness.

Observe that some excursions may have length 2. When this happens, the individual who is born, dies right away without competing with any other individual. To ensure that each excursion has competition among individuals in a sort of natural selection process, we introduce a change-over on the model: Each time after the system becomes empty, m independent new species are placed on the system (instead of just 1) with probability p, or the system stays empty with probability $1 - p$. We denote this variation by GMS(m). In this set up GMS(1) is the original model.

Figures 1a and 1b show the role of the competition on the distribution of the fitness of the strongest individual on each excursion. Short excursions are more commom on GMS(1) than on GMS(10). That behaviour favors individuals with lower fitnesses to be the strongest ones. Competition introduced in GMS(10) avoids that.

The next result computes the fitness distribution of the strongest individual to die right before the system becomes empty on GMS(m) model. It is shown in terms of the hypergeometric function of Gauss (see Luke [8]). This function is denoted by $_2F_1(a, b; c; z)$, namely,

$$_2F_1(a, b; c; z) = \sum_{k \geq 0} \frac{(a)_k (b)_k}{(c)_k} \frac{z^k}{k!}, \quad |z| < 1, \tag{2.1}$$

where a, b, c, are real numbers with $c \neq 0, -1, -2, \ldots$, and $(a)_k$ is the coefficient Pocham-

mer, namely,

$$(a)_k = a(a+1)\cdots(a+k-1) \quad (a)_0 = 1.$$

Theorem 2.1. *Let $p \leq 1/2$ and Z_m be the fitness of the strongest individual before the system becomes empty on GMS(m) model. Then Z_m is a random variable with distribution*

$$\mathbb{P}[Z_m \leq t] = (qt)^m {}_2F_1\left(\frac{m}{2}, \frac{m+1}{2}; m+1; 4pqt\right) \qquad 0 \leq t < 1.$$

Corollary 2.2. *Let $p \leq 1/2$ and Z be the fitness of the strongest individual before the system becomes empty on GMS(1) model. Then*

$$\mathbb{P}[Z \leq t] = \frac{1 - \sqrt{1 - 4pqt}}{2p}, \qquad 0 \leq t < 1.$$

For $p = 1/2$, Z follows a Beta distribution $B(1, 1/2)$.

Remark 2.3. By Theorem 2.1 we have Z_m density probability function is

$$
\begin{aligned}
f_m(t) &= \frac{d}{dt}\left[(qt)^m {}_2F_1\left(\frac{m}{2}, \frac{m+1}{2}; m+1; 4pqt\right)\right] \\
&= mq^m t^{m-1} {}_2F_1\left(\frac{m}{2}, \frac{m+1}{2}; m; 4pqt\right)
\end{aligned}
\tag{2.2}
$$

where the last line have been obtained by using Abramowitz and Stegun [1, Eq. 15.2.4]. When $p = q = 1/2$, the distribution of $1 - Z_m$ is known as *hypergeometric function type I distribution* (see Gupta and Nagar [7, p. 298]).

Corollary 2.4. $\mathbb{E}[Z_m] = 1 - \dfrac{q^m}{m+1} {}_2F_1\left(\dfrac{m}{2}, \dfrac{m+1}{2}; m+1; 4pq\right)$

Remark 2.5. Considering Corollary 2.4 when $p = 1/2$, by using the following equality (see Gradshteyn and Ryzhik [4, Eq. 7.512.11])

$$
{}_2F_1(\alpha, \beta; \gamma; 1) = \frac{\Gamma(\gamma)\Gamma(\gamma - \alpha - \beta)}{\Gamma(\gamma - \alpha)\Gamma(\gamma - \beta)}
$$

we have that

$$
\begin{aligned}
\mathbb{E}[Z_m] &= 1 - \frac{q^m}{m+1} {}_2F_1\left(\frac{m}{2}, \frac{m+1}{2}; m+2; 1\right) \\
&= 1 - \frac{1}{2^m(m+1)}\left[\frac{\Gamma(m+2)\Gamma(3/2)}{\Gamma(\frac{m+3}{2} + \frac{1}{2})\Gamma(\frac{m+3}{2})}\right] \\
&= 1 - \frac{\sqrt{\pi}\, m!}{2^{m+1}\Gamma(\frac{m+3}{2} + \frac{1}{2})\Gamma(\frac{m+3}{2})}
\end{aligned}
$$

where the last line has been obtained by using $\Gamma(3/2) = \sqrt{\pi}/2$. Now, using the duplication formula, namely,

$$
\Gamma(2z) = \frac{\Gamma(z + \frac{1}{2})\Gamma(z)}{2^{1-2z}\sqrt{\pi}}
$$

we get

$$
\mathbb{E}[Z_m] = 1 - \frac{2}{(m+1)(m+2)}
$$

3 Proofs

Proof Theorem 2.1. For $n = 0, 1, \ldots$ we define

$$\tau_n = \inf\{l \geq 1 : X_{n+l} = 0, X_n = 0\}.$$

In words τ_n is the length of a excursion from 0 to 0. As the process X_n is homogeneous, the distribution of τ_n does not depend on n so we consider the random variable $\tau := \tau_0$. Besides, as $p \leq 1/2$ we have that $\mathbb{P}[\tau < \infty] = 1$ and

$$\mathbb{P}[\tau = k + 1] = \mathbb{P}[T_{-m} = k] = \frac{m}{k}\binom{k}{\frac{k-m}{2}}p^{(k-m)/2}q^{(k+m)/2}, \ k \geq m, \ k+m \text{ even},$$

where T_{-m} is the time of the first visit to $-m$ for a random walk on $\mathbb{Z}$ beginning at 0. (See Bhattacharya and Waymire [3])

If $\tau = k + 1$, we see along that excursion, extra $\frac{k-m}{2}$ births and $\frac{k+m}{2}$ deaths. The last death corresponds to the individual with the strongest fitness among all $\frac{k+m}{2}$ that were born. Hence,

$$\mathbb{P}[Z_m \leq t] = \sum_{k=m}^{\infty} \mathbb{P}[\tau = k+1]\mathbb{P}[\max(Y_1, ..., Y_{\frac{m+k}{2}}) \leq t],$$

where $Y_1, ..., Y_{\frac{m+k}{2}}$ are i.i.d. uniform random variables on $[0, 1]$. Therefore,

$$
\begin{aligned}
\mathbb{P}[Z_m \leq t] &= \sum_{k=m}^{\infty} \frac{m}{k}\binom{k}{\frac{k-m}{2}}p^{(k-m)/2}q^{(k+m)/2}t^{(k+m)/2}\mathbf{1}_{\{m+k \text{ even}\}} \\
&= \sum_{l=m}^{\infty} \frac{m}{2l-m}\binom{2l-m}{l}p^{l-m}q^l t^l \quad (k+m = 2l, l \geq m) \\
&= \sum_{j=0}^{\infty} \frac{m}{m+2j}\binom{m+2j}{m+j}p^j q^{m+j}t^{m+j} \quad (l = m+j) \\
&= (qt)^m \sum_{j=0}^{\infty} \frac{m}{m+2j}\binom{m+2j}{m+j}(pqt)^j \\
&= (qt)^m \sum_{j=0}^{\infty} \frac{(m)_{2k}}{(m+1)_k}\frac{(pqt)^j}{k!} \\
&= (qt)^m \, {}_2F_1\left(\frac{m}{2}, \frac{m+1}{2}; m+1; 4pqt\right)
\end{aligned}
$$

where the last line has been obtained by using $(a)_{2k} = \left(\frac{a}{2}\right)_k\left(\frac{a+1}{2}\right)_k 2^{2k}$ and the definition of Gauss hypergeometric function. $\qquad\square$

Proof Corollary 2.2. It is a particular case of Theorem 2.1 when $m = 1$. In this situation

$$
\begin{aligned}
qt \, {}_2F_1\left(\frac{1}{2}, 1; 2; 4pqt\right) &= qt \sum_{k \geq 0} \frac{(1/2)_k(1)_k}{(2)_k}\frac{(4pqt)^k}{k!} \\
&= \frac{1}{p}\sum_{k \geq 0} \frac{(2k)!}{(k+1)! \, k!}(pqt)^{k+1} \\
&= \frac{1 - \sqrt{1 - 4pqt}}{2p}
\end{aligned}
$$

where the last line has been obtained by using $(1)_k(1/2)_k = 2^{-2k}(2k)!$ and the result given in Prudnikov *et al* [9, Eq. 5.2.13.8]. □

Proof Corollary 2.4.

$$
\begin{aligned}
\mathbb{E}[Z_m] &= \int_0^1 \mathbb{P}[Z_m > t]\, dt \\
&= 1 - q^m \int_0^1 t^m\, {}_2F_1\left(\tfrac{m}{2}, \tfrac{m+1}{2}; m+1; 4pqt\right)\, dt \\
&= 1 - \tfrac{q^m}{m+1}\, {}_2F_1\left(\tfrac{m}{2}, \tfrac{m+1}{2}; m+2; 4pq\right)
\end{aligned}
$$

where the last line has been obtained by using the result given in Gradshteyn and Ryzhik [4, Eq. 7.512.11]. □

References

[1] Abramowitz, M. and Stegun, I.: Handbook of Mathematical Functions. *National Bureau of Standards.* (1970).

[2] Ben-Ari, I.; Matzavinos, A. and Roitershtein, A.: On a species survival model. *Electronic Communications in Probability,* **16**, (2011), 226-233. MR-2788894

[3] Bhattacharya, R.N. and Waymire, E.C.: Stochastic Processes With Applications. *Society for Industrial & Applied, New York.* (2009). MR-3396216

[4] Gradshteyn, I.S. and Ryzhik, I.M.: Table of integrals, series, and products. seventh edition. *Elsevier/Academic Press, Amsterdam. Translated from the Russian, Translation edited and with a preface by Alan Jeffrey and Daniel Zwillinger.* (2007). MR-2360010

[5] Guiol, H.; Machado, F. and Schinazi, R.: A Stochastic model of evolution. *Markov Process. Related Fields*, **17**, no. 2, (2011), 253-258. MR-2856242

[6] Guiol, H.; Machado, F. and Schinazi, R.: On a link between a species survival time in an evolution model and the Bessel distributions. *Brazilian Journal of Probability and Statistics*, **27**, no. 2, (2013), 201-209. MR-3028804

[7] Gupta, A.K. and Nagar, D.K.: Matrix Variate Distributions. *Chapman & Hall/CRC, Boca Raton* (2000). MR-1738933

[8] Luke, Y.L.:The Special Functions and Their Approximations.*Academic Press, New York* vol. 1. (1969)

[9] Prudnikov, A.P.; Brychkov, Yu.A. and Marichev, O.I.: Integrals and Series. Volume 1. Elementary Functions. *Taylor & Francis, London. Translated from the Russian by N.M. Queen* (2002). MR-0874986

[10] Skevi, M. and Volkov, S.E.: On the generalization of the GMS evolutionary model. *Markov Process. and Related Fields*, **18**, no. 2, (2012), 311-322. MR-2985139

Electronic Journal of Probability
Electronic Communications in Probability

Advantages of publishing in EJP-ECP

- Very high standards

- Free for authors, free for readers

- Quick publication (no backlog)

- Secure publication (LOCKSS[1])

- Easy interface (EJMS[2])

Economical model of EJP-ECP

- Non profit, sponsored by IMS[3], BS[4] , ProjectEuclid[5]

- Purely electronic

Help keep the journal free and vigorous

- Donate to the IMS open access fund[6] (click here to donate!)

- Submit your best articles to EJP-ECP

- Choose EJP-ECP over for-profit journals

---

[1] LOCKSS: Lots of Copies Keep Stuff Safe http://www.lockss.org/
[2] EJMS: Electronic Journal Management System http://www.vtex.lt/en/ejms.html
[3] IMS: Institute of Mathematical Statistics http://www.imstat.org/
[4] BS: Bernoulli Society http://www.bernoulli-society.org/
[5] Project Euclid: https://projecteuclid.org/
[6] IMS Open Access Fund: http://www.imstat.org/publications/open.htm