

RESEARCH

Open Access

Extinction in a branching process: why some of the fittest strategies cannot guarantee survival

Sterling Sawaya^{1,3*} and Steffen Klaere²

*Correspondence:

sterlingsawaya@gmail.com¹Institute for Behavioral Genetics,
University of Colorado, Boulder,
Colorado, USA³Formerly: Department of
Anatomy, and Allan Wilson Centre
for Molecular Ecology and
Evolution, University of Otago,
Dunedin, New ZealandFull list of author information is
available at the end of the article

Abstract

Biological fitness is typically measured by the expected rate of reproduction, but strategies with high fitness can also have high probabilities of extinction. Likewise, gambling strategies with a high expected payoff can also have a high risk of ruin. We take inspiration from the gambler's ruin problem to examine how extinction is related to population growth. Using moment theory we demonstrate how higher moments can impact the probability of extinction and how the first few moments can be used to find bounds on the extinction probability, focusing on s -convex ordering of random variables. This approach generates "best case" and "worst case" scenarios to provide upper and lower bounds on the probability of extinction.

MSC Codes: 92D15, 60J80, 60E15**Keywords:** Extinction; Branching process; S -convex

1 Introduction

Reproduction is necessary for the survival of populations. However, a population can have a high expected reproductive rate but nevertheless go extinct with near certainty (Lewontin and Cohen 1969). For example, populations with large variation in reproductive success can sometimes have a high probability of extinction, even if they have a high expected growth (Tuljapurkar and Orzack 1980).

Similarly, investors and gamblers can avoid Gambler's Ruin through growth of capital. However, a gambler should not simply apply the strategy with the highest expected growth rate as it may also have a high risk of ruin. For example, investors can use the Kelly ratio (Kelly 1956) to maximize expected geometric growth of their capital but strict adherence to this ratio can be risky, and playing a more conservative strategy is often recommended (MacLean et al. 2010).

To estimate the probability of Gambler's Ruin, one can use approximations based on moments (Ethier and Khoshnevisan 2002; Canjar 2007; Hürlimann 2005). Here we apply these approaches to estimate the probability of extinction in a branching process. The mathematics of Gambler's Ruin are very similar to that of extinction in a branching process (Courtois et al. 2006). Both statistical models involve a random variable (pay-off/offspring number), resulting in a random walk (change in capital/change in population size), and an absorbing state (ruin/extinction). Moreover, both processes are assumed to

be Markovian, and finding the probability of ruin/extinction involves solving for the root of a convex function.

Here we examine the random variable representing the number of offspring, and investigate how the moments of this random variable are related to the probability of extinction. We demonstrate an important relationship between these moments and extinction: odd moments favor survival and even moments favor extinction. The first moment of the offspring distribution, its mean, has the biggest influence on extinction. However, the first moment alone is not usually informative about extinction probabilities. In fact, strategies with arbitrarily large first moments can nevertheless go extinct with near certainty. Some of the “fittest” strategies can be highly unlikely to survive.

Using the first few moments of the offspring distribution, one can obtain bounds on the ultimate probability of extinction (Courtois et al. 2006; Daley and Narayan 1980). These bounds provide “best case” and “worst case” distributions. We present these bounds, termed s -convex extremal random variables, adapted from actuarial science and research on the gambler’s ruin problem (Denuit and Lefevre 1997; Hürlimann 2005; Courtois et al. 2006). The extremal distributions for discrete processes have been developed previously, using up to four moments (Hürlimann 2005). Here we find the conditions under which these extremal distributions provide non-trivial bounds. Using some simple examples, we demonstrate how these methods can be used to rank distributions using only their moments. We then discuss how these bounds can be used to better understand the evolutionary process.

2 Methods

2.1 Extinction in the Galton-Watson branching process

To investigate biological extinction, we use a Galton-Watson branching process in which, at each discrete time interval, every individual generates i discrete offspring with probability p_i , and zero offspring with p_0 . Without loss of generality we assume that an individual produces its offspring and then dies, so that each individual in a population is restricted to a single generation. The offspring number is a random variable, which we denote by X . Let n be the maximum value of X so that X takes values in the state space $\mathcal{D}_n = \{0, 1, 2, \dots, n\}$

At any given time t , the size of a population (Z_t) is the number of individuals in the branching process. We set $Z_0 \equiv 1$ unless otherwise specified. The probability of extinction of a branching process is $q \equiv \lim_{t \rightarrow \infty} P(Z_t = 0 | Z_0 = 1)$. If the starting size of the population is greater than one, then the overall probability of extinction can be defined as

$$\lim_{t \rightarrow \infty} P(Z_t = 0 | Z_0 = N) = q^N.$$

So we can solve for extinction in the case of $Z_0 = 1$ and extend the results to larger starting populations if necessary.

The recursive formula for finding q can be found through a first step analysis (Kimmel and Axelrod 2002). The probability that the lineage of a single individual eventually goes extinct is the probability that it dies without offspring (p_0) plus the probability that it produces a single offspring whose lineage dies out (p_1q) plus the probability that it produces two offspring whose joint lineages die out (p_2q^2), and so on.

This leads to the formal definition of the probability generating function:

$$f(q) = \mathbb{E}[q^X] = p_0 + p_1q + p_2q^2 + p_3q^3 \dots p_nq^n = \sum_{k=0}^n p_k q^k. \quad (1)$$

The probability of extinction of a branching process starting with a single individual is the smallest root of the equation $f(q) = q$ for $q \in [0, 1]$. The solution $q = 1$ is always a root of (1) and is not necessarily the smallest positive root. In some cases, the probability of extinction is trivially obvious. For instance, if $p_0 = 0$ individuals always produces at least one offspring, therefore $q = 0$. Furthermore, cases where $\mathbb{E}[X] \leq 1$ always yield $q = 1$ (Kimmel and Axelrod 2002).

Inferring the probability of extinction analytically for branching processes with $p_0 > 0$ and $\mathbb{E}[X] > 1$ can be difficult because (1) has n complex-valued roots according to the fundamental law of algebra. In the following we illustrate how (1) can be seen in terms of moments of the offspring distribution, and discuss how this approach can be used to estimate q .

2.2 Moments of the branching process

Let $m_k \equiv \mathbb{E}[X^k]$ denote the k th moment of the branching process generator X . The first moment, m_1 , is equivalent to the average offspring number. Higher moments can be used to obtain other summary statistics of the distribution, such as the variance $\sigma^2 = m_2 - m_1^2$.

The Laplace transform of (1) can be used to (recursively) express extinction in terms of the moments of the branching process

$$\begin{aligned} f(q) &= \mathbb{E}[q^X] = \mathbb{E}[e^{X \log q}] \\ &= 1 + m_1 \log q + m_2 \frac{(\log q)^2}{2} + m_3 \frac{(\log q)^3}{6} + \dots \\ &= \sum_{k=0}^{\infty} m_k \frac{(\log q)^k}{k!} \end{aligned}$$

where $m_0 = 1$. Note that $m_k > 0$ for all $k \geq 0$. Furthermore, with $q \in (0, 1)$ we have $\log q < 0$. Therefore, even moments increase the probability of extinction while odd moments decrease it. Additionally, if $q \in (e^{-1}, 1)$ then $\log q \in (-1, 0)$ and the series converges with $\log q$. Thus, approximations, $f^*(q)$, which take the form

$$f^*(q) = \sum_{k=0}^{s-1} m_k \frac{(\log q)^k}{k!} + o((\log q)^s)$$

for $s \geq 3$ are only accurate when q is large and the moments are small. As $q \downarrow 0$, the series requires more and more terms to provide accurate approximation. Therefore, when q is small the first few moments are not necessarily informative about the probability of extinction.

2.3 s-Convex orderings of random variables

Here we demonstrate how the first few moments of the offspring distribution can be used to find bounds on the probability of extinction. The random variable X is bound by zero and its maximum, n , conveniently allowing for s -convex ordering (Denuit and Lefevre 1997; Hürlimann 2005; Courtois et al. 2006). Following (Hürlimann 2005) denote by Δ

the forward difference operator for $g : \mathcal{D}_n \rightarrow \mathbb{R}$ by $\Delta g(i) = g(i+1) - g(i)$ for all $i \in \mathcal{D}_{n-1}$. Analogously for $k \in \mathcal{D}_n$ the k -th order forward difference operator is defined recursively by $\Delta^0 g = g$ and for $k \geq 1$ by $\Delta^k g(i) = \Delta^{k-1} g(i-1) - \Delta^{k-1} g(i)$ for all $i \in \mathcal{D}_{n-k}$. Then, for two random variables X and Y valued in \mathcal{D}_n we say X precedes Y in the s -convex order, written $X \leq_{s-cx}^{\mathcal{D}_n} Y$ if $\mathbb{E}[g(X)] \leq \mathbb{E}[g(Y)]$ for all s -convex real functions g on \mathcal{D}_n . A convenient consequence is that if $X \leq_{s-cx}^{\mathcal{D}_n} Y$ then

$$\begin{aligned} \mathbb{E}(X^k) &= \mathbb{E}(Y^k) \quad \text{for } k = 1, 2, \dots, s-1 \\ \mathbb{E}(X^k) &\leq \mathbb{E}(Y^k) \quad \text{for } k \geq s. \end{aligned}$$

Define the *moment space* for all random variables with state set \mathcal{D}_n and fixed first $s-1$ moments m_1, \dots, m_{s-1} by

$$\mathfrak{B}_{s,n}^{\bar{m}} \equiv \mathfrak{B}(\mathcal{D}_n, m_1, m_2, \dots, m_{s-1}).$$

Since the random variable X is strictly positive, its moment space only contains positive elements. Further, we are only interested in cases where the mean is greater than 1 so that extinction is not certain. This provides a moment space with well behaved properties. The study of the moment problem (see e.g., (Karlin and McGregor 1957; Prékopa 1990)) yields an important relationship between consecutive moments on $\mathfrak{B}_{s,n}^{\bar{m}}$ conditional on $m_1 \geq 1$

$$(m_i)^{\frac{i+1}{i}} \leq m_{i+1} \leq nm_i. \tag{2}$$

Minimum and maximum extrema distributions on $\mathfrak{B}_{s,n}^{\bar{m}}$ can be found for any distribution on \mathcal{D}_n , with fixed first s moments m_1, m_2, \dots, m_s (Denuit and Lefevre 1997). The random variables for these distributions are denoted $X_{\min}^{(s)}$ and $X_{\max}^{(s)}$ such that

$$X_{\min}^{(s)} \leq_{s-cx}^{\mathcal{D}_n} X \leq_{s-cx}^{\mathcal{D}_n} X_{\max}^{(s)} \quad \text{for all } X \in \mathcal{D}_n.$$

Extrema have been derived for $s = 2, 3, 4, 5$ (Denuit and Lefevre 1997; Denuit et al. 1999; Hürlimann 2005). Here, we reiterate these results providing the inferred distributions and their utility when obtaining bounds on the probability of extinction. We begin on $\mathfrak{B}_{2,n}^{\bar{m}}$ with the maximal random variable, $X_{\max}^{(2)}$, defined as:

$$X_{\max}^{(2)} = \begin{cases} 0 & \text{with } p_0 = 1 - \frac{m_1}{n} \\ n & \text{with } p_n = \frac{m_1}{n}. \end{cases}$$

For $X_{\max}^{(2)}$ we observe $m_{i+1} = nm_i$, so by (2) this can clearly be seen as the maximum extrema. Intuitively, this is the “long shot” distribution on \mathcal{D}_n , a worst case scenario. Because the values and respective probabilities of $X_{\max}^{(2)}$ are known, q can be solved explicitly by finding the least positive root of the generating function:

$$f(q) = p_0 + p_n q^n.$$

This provides an upper limit on extinction because this generating function will be greater than or equal to the generating function for all other random variables with the same m_1 and n , on $q \in [0, 1]$.

$\mathfrak{B}_{2,n}^{\bar{m}}$ is a very general moment space and the first moment does not often provide much information about an unknown distribution. Therefore, $X_{\max}^{(2)}$ is not likely to be a tight upper bound when n is large or unknown. However, if m_1 is near n , then the distribution can be fairly well approximated by $X_{\max}^{(2)}$.

Unlike $X_{\max}^{(2)}, X_{\min}^{(2)}$ does not provide a useful bound on the probability of extinction. $X_{\min}^{(2)}$ is defined as:

$$X_{\min}^{(2)} = \begin{cases} \alpha & \text{with } p_\alpha = \alpha + 1 - m_1 \\ \alpha + 1 & \text{with } p_{\alpha+1} = m_1 - \alpha \end{cases} \quad (3)$$

where α is the integer on \mathcal{D}_n such that

$$\alpha < m_1 \leq \alpha + 1.$$

This extremal random variable represents a best case scenario. However, since $m_1 > 1$, α must be larger than zero and this branching process has no chance of death (i.e. $p_0 = 0$) and consequently no chance of extinction ($q = 0$). Therefore $X_{\min}^{(2)}$ does not provide a useful bound on the probability of extinction as the bound $q \geq 0$ is obvious.

This bound and all other bounds examined here can be found using discrete Chebyshev systems (Denuit and Lefevre 1997). However, extremal bounds are perhaps more intuitive for continuous random variables, to which the discrete cases can be seen as similar (Shaked and Shanthikumar 2007; Hürlimann 2005; Denuit et al. 1999). For example, $X_{\min}^{(2)}$ in the continuous case has only one possible value, m_1 with $p_{m_1} = 1$. By (2) this is clearly an extrema because $(m_i)^{(i+1)/i} = m_{i+1} = (m_1)^{i+1}$. In comparison, the discrete case (3) has similar properties.

The following notation helps extending these calculations to higher order systems (Denuit et al. 1999). Let $w, x, y, z \in \mathcal{D}_n$, and set $m_0 = 1$. Then:

$$\begin{aligned} m_{j,z} &:= z \cdot m_{j-1} - m_j, & j = 1, 2, \dots; \\ m_{j,z,y} &:= y \cdot m_{j-1,z} - m_{j,z}, & j = 2, 3, \dots; \\ m_{j,z,y,x} &:= x \cdot m_{j-1,z,y} - m_{j,z,y}, & j = 3, 4, \dots; \\ m_{j,z,y,x,w} &:= w \cdot m_{j-1,z,y,x} - m_{j,z,y,x}, & j = 4, 5, \dots \end{aligned}$$

The reader should recognize this notation as it is simply the iterative forward difference operator Δ^k for moments.

If the first two moments are known, then a tighter upper bound can be found. On $\mathfrak{B}_{3,n}^{\bar{m}}$ the minimal distribution in the 3-convex sense is given by:

$$X_{\min}^{(3)} = \begin{cases} 0 & \text{with } p_0 = 1 - p_\alpha - p_{\alpha+1} \\ \alpha & \text{with } p_\alpha = \frac{m_{2,\alpha+1}}{\alpha} \\ \alpha + 1 & \text{with } p_{\alpha+1} = \frac{-m_{2,\alpha}}{\alpha + 1} \end{cases}$$

where

$$\alpha < \frac{m_2}{m_1} \leq \alpha + 1.$$

This bound is already known in the branching process literature (Daley and Narayan 1980). Similar to $X_{\max}^{(2)}$, the extremal random variable $X_{\min}^{(3)}$ represents a worst case scenario, this time using two moments. The root of the equation

$$f(q) = q = p_0 + p_\alpha q^\alpha + p_{\alpha+1} q^{\alpha+1} \quad (4)$$

provides an upper bound to the probability of extinction, so that (4) has greater values at any $q \in [0, 1)$ than the probability generating functions of any other random variable in $\mathfrak{B}_{3,n}^{\bar{m}}$.

In contrast to $X_{\max}^{(2)}$, the minimum extrema on $\mathfrak{B}_{3,n}^{\bar{m}}$ yields the upper limit for the probability of extinction. The alternation between minimum and maximum for the worst case scenarios is due to the convexity of (1). Again, this extrema is perhaps more intuitive in the continuous sense, in which

$$X_{\min, \text{cont.}}^{(3)} = \begin{cases} 0 & \text{with } p_0 = 1 - p_{m_2/m_1} \\ \frac{m_2}{m_1} & \text{with } p_{m_2/m_1} = \frac{(m_1)^2}{m_2}. \end{cases}$$

In this case, successive moments simply grow by m_2/m_1 , so that $m_{i+1} = m_i(m_2/m_1)$, providing a minimum on $\mathfrak{B}_{3,n}^{\bar{m}}$. And, as was the case for the minimum on $\mathfrak{B}_{2,n}^{\bar{m}}$, the discrete minimum extrema on $\mathfrak{B}_{3,n}^{\bar{m}}$ has similar properties to the continuous minimum extrema.

For both $\mathfrak{B}_{2,n}^{\bar{m}}$ and $\mathfrak{B}_{3,n}^{\bar{m}}$ the discrete cases are simply discretization of the continuous case. However, this is not necessarily the case for higher moment spaces (Courtois et al. 2006). While the continuous cases provide more intuitive extrema, derivation of the discrete case for higher moments is not as simple as deriving the continuous case and discretizing.

Next, we examine the maximum extrema on $\mathfrak{B}_{3,n}^{\bar{m}}$:

$$X_{\max}^{(3)} = \begin{cases} \alpha & \text{with } p_\alpha = \frac{m_{2,n,\alpha+1}}{n - \alpha} \\ \alpha + 1 & \text{with } p_{\alpha+1} = \frac{-m_{2,n,\alpha}}{n - \alpha - 1} \\ n & \text{with } p_n = 1 - p_\alpha - p_{\alpha+1} \end{cases}$$

where

$$\alpha < \frac{nm_1 - m_2}{n - m_1} \leq \alpha + 1.$$

Since $X_{\max}^{(3)}$ can only provide non-trivial information about q if $p_0 > 0$, this extremal distribution is only informative about extinction when $\alpha = 0$ and $p_\alpha > 0$, which is the case whenever $nm_1 - m_2 < n - m_1$. Although this requirement may appear restrictive, some classes of distributions have simple rules under which $X_{\max}^{(3)}$ is informative. For example, for binomial distributions, $B_{n,p}$, $X_{\max}^{(3)}$ will provide a non-zero lower bound if $1/n < p \leq 1/(n - 1)$.

We move on to $\mathfrak{B}_{4,n}^{\bar{m}}$. The use of three moments can improve bounds on the probability of extinction, but as with all of the maximal random variables, $X_{\max}^{(4)}$ requires the knowledge of the maximum, n . $X_{\max}^{(4)}$ is defined as:

$$X_{\max}^{(4)} = \begin{cases} 0 & \text{with } p_0 = 1 - p_\alpha - p_{\alpha+1} - p_n \\ \alpha & \text{with } p_\alpha = \frac{m_{3,n,\alpha+1}}{\alpha(n - \alpha)} \\ \alpha + 1 & \text{with } p_{\alpha+1} = \frac{-m_{3,n,\alpha}}{(\alpha + 1)(n - \alpha - 1)} \\ n & \text{with } p_n = \frac{m_{3,\alpha,\alpha+1}}{n(n - \alpha)(n - \alpha - 1)} \end{cases}$$

where

$$\alpha < \frac{m_2n - m_3}{m_1n - m_2} \leq \alpha + 1.$$

While this is a potential improvement to the lower bound given by $X_{\min}^{(3)}$, the improvement is sometimes negligible. As $n \rightarrow \infty$, the difference between $X_{\max}^{(4)}$ and $X_{\min}^{(3)}$ vanishes because

$$\lim_{n \rightarrow \infty} \frac{m_2 n - m_3}{m_1 n - m_2} = \frac{m_2}{m_1}$$

and furthermore, if $n \rightarrow \infty$ then $p_n \rightarrow 0$. Therefore, the resulting generating function for $X_{\max}^{(4)}$ is identical to (4) if the maximal value is unknown. So, like the first moment, the third moment is uninformative about extinction when n is unknown, unless assumptions are made about the distribution (see e.g., (Daley and Narayan 1980; Ethier and Khoshnevisan 2002)).

The minimal extrema for $\mathfrak{B}_{4,n}^{\bar{m}}$, $X_{\min}^{(4)}$ is given by

$$X_{\min}^{(4)} = \begin{cases} \alpha, & \text{with } p_\alpha = \frac{m_{3,\beta,\beta+1,\alpha+1}}{(\beta - \alpha)(\beta + 1 - \alpha)} \\ \alpha + 1, & \text{with } p_{\alpha+1} = \frac{-m_{3,\beta,\beta+1,\alpha}}{(\beta - \alpha)(\beta - 1 - \alpha)} \\ \beta, & \text{with } p_\beta = \frac{m_{3,\alpha,\alpha+1,\beta+1}}{(\beta - \alpha)(\beta - 1 - \alpha)} \\ \beta + 1, & \text{with } p_{\beta+1} = \frac{-m_{3,\alpha,\alpha+1,\beta}}{(\beta - \alpha)(\beta + 1 - \alpha)} \end{cases}$$

where α and β are given by

$$\alpha < \frac{m_{3,\beta,\beta+1}}{m_{2,\beta,\beta+1}} \leq \alpha + 1, \quad \beta < \frac{m_{3,\alpha,\alpha+1}}{m_{2,\alpha,\alpha+1}} \leq \beta + 1.$$

Again, this bound is only useful if $p_0 > 0$. Unfortunately there is no short form equation to identify which spaces $\mathfrak{B}_{4,n}^{\bar{m}}$ fit this requirement. However, one can easily determine if a given $\mathfrak{B}_{4,n}^{\bar{m}}$ has a useful $X_{\min}^{(4)}$. Assuming $\alpha = 0$, $\hat{\beta}$ is simply bound by

$$\hat{\beta} < \frac{m_3 - m_2}{m_2 - m_1} \leq \hat{\beta} + 1.$$

And if $m_{3,\hat{\beta},\hat{\beta}+1} < m_{2,\hat{\beta},\hat{\beta}+1}$, then the bound is useful because the resulting $X_{\min}^{(4)}$ has $p_0 > 0$. Alternatively, if $m_{3,\hat{\beta},\hat{\beta}+1} \geq m_{2,\hat{\beta},\hat{\beta}+1}$ the supports for $X_{\min}^{(4)}$ have $p_0 = 0$ and consequently $q = 0$.

If the first four moments are known, the extremal variable $X_{\min}^{(5)}$ can be obtained. Its distribution takes a simple form, but the equations used to find its values and relative probabilities are relatively large. From (Hürlimann 2005), $X_{\min}^{(5)}$ is defined as:

$$X_{\min}^{(5)} = \begin{cases} 0 & \text{with } p_0 = 1 - p_\alpha - p_{\alpha+1} - p_\beta - p_{\beta+1} \\ \alpha & \text{with } p_\alpha = \frac{m_{4,\beta,\beta+1,\alpha+1}}{\alpha(\beta - \alpha)(\beta + 1 - \alpha)} \\ \alpha + 1 & \text{with } p_{\alpha+1} = \frac{-m_{4,\beta,\beta+1,\alpha}}{(\alpha + 1)(\beta - \alpha)(\beta - 1 - \alpha)} \\ \beta & \text{with } p_\beta = \frac{m_{4,\alpha,\alpha+1,\beta+1}}{\beta(\beta - \alpha)(\beta - 1 - \alpha)} \\ \beta + 1 & \text{with } p_{\beta+1} = \frac{-m_{4,\alpha,\alpha+1,\beta}}{(\beta + 1)(\beta - \alpha)(\beta + 1 - \alpha)} \end{cases}$$

where

$$\alpha < \frac{m_{4,\beta,\beta+1}}{m_{3,\beta,\beta+1}} \leq \alpha + 1, \quad \beta < \frac{m_{4,\alpha,\alpha+1}}{m_{3,\alpha,\alpha+1}} \leq \beta + 1. \tag{5}$$

Courtois et al. (Courtois et al. 2006) explained that there is no analytic form to directly obtain α and β for $X_{\min}^{(5)}$. They showed this by disproving the intuitive idea that the discrete support encloses the continuous support. To find α and β , we iteratively search all possible supports on D_n until both inequalities are satisfied. This exhaustive method for finding the supports for this extrema is not ideal, especially if D_n is dense. Linear programming can be used to easily find the extremal supports and their probabilities (Prékopa 1990), but such approaches are not necessary when D_n is sparse (e.g., when n is relatively small).

Hürlimann (Hürlimann 2005) also presents a form for the upper extremal variable in $\mathfrak{B}_{5,n}^{\bar{m}}$. The process $X_{\max}^{(5)}$ is defined as:

$$X_{\max}^{(5)} = \begin{cases} \alpha, & \text{with } p_{\alpha} = \frac{m_{4,n,\beta,\beta+1,\alpha+1}}{(\beta - \alpha)(\beta + 1 - \alpha)(n - \alpha)} \\ \alpha + 1, & \text{with } p_{\alpha+1} = \frac{-m_{4,n,\beta,\beta+1,\alpha}}{(\beta - \alpha)(\beta - \alpha - 1)(n - \alpha - 1)} \\ \beta, & \text{with } p_{\beta} = \frac{m_{4,n,\alpha,\alpha+1,\beta+1}}{(\beta - \alpha)(\beta - \alpha - 1)(n - \beta)} \\ \beta + 1, & \text{with } p_{\beta+1} = \frac{-m_{4,n,\alpha,\alpha+1,\beta}}{(\beta - \alpha)(\beta + 1 - \alpha)(n - \beta - 1)} \\ n, & \text{with } p_n = 1 - p_{\alpha} - p_{\alpha+1} - p_{\beta} - p_{\beta+1} \end{cases}$$

where

$$\alpha < \frac{m_{4,n,\beta,\beta+1}}{m_{3,n,\beta,\beta+1}} \leq \alpha + 1, \quad \beta < \frac{m_{4,n,\alpha,\alpha+1}}{m_{3,n,\alpha,\alpha+1}} \leq \beta + 1.$$

As was the case for $X_{\min}^{(4)}$, one can determine if $X_{\max}^{(5)}$ has $p_0 > 0$ by assuming $\alpha = 0$ and solving for $\hat{\beta}$ with

$$\hat{\beta} < \frac{m_{4,n,0,1}}{m_{3,n,0,1}} \leq \hat{\beta} + 1.$$

If the resulting $\hat{\beta}$ in the inequality $m_{4,n,\hat{\beta},\hat{\beta}+1} < m_{4,n,\hat{\beta},\hat{\beta}+1}$ holds, the bound for $X_{\max}^{(5)}$ is informative.

All $X_{\max}^{(j)}$ extrema rely on the maximum offspring number, n . Similar to $X_{\max}^{(4)}$, when n is unknown or infinity $X_{\max}^{(5)}$ goes to the minimum on the lower moment space, here $X_{\min}^{(4)}$. Thus if n is unknown, $X_{\max}^{(j)}$ goes to $X_{\min}^{(j-1)}$, at least for the cases examined here.

The Chebychev approach can be used to extend this approach to higher moments (Hürlimann 2005). However, moments above the fourth are rarely used, and higher moments can be difficult to estimate from small samples. Further, the equations for the supports and probabilities for moments above the fourth become increasingly complex.

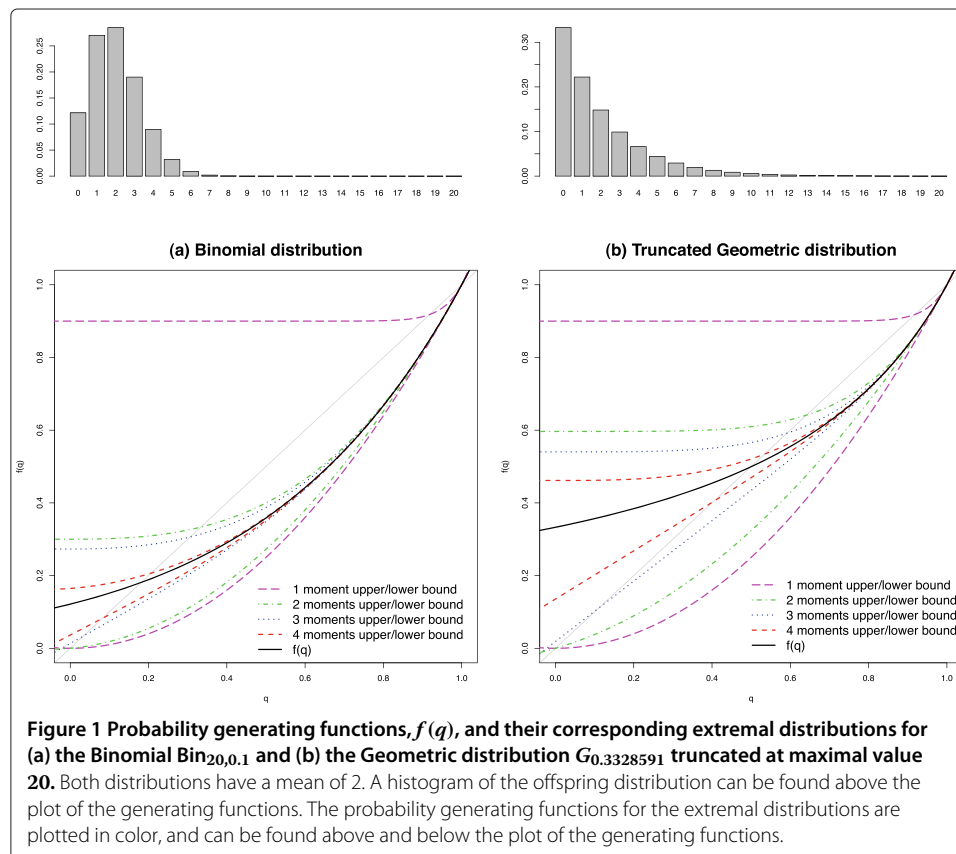
3 Results and discussion

Here we discuss some example distributions, graph their generating functions, and also graph generating functions for the extremal distributions. The plot of the probability generating function, $f(q)$, on $q \in (0, 1)$ is a useful way to visualize how the moments are related to extinction. The probability generating function takes the value p_0 at $q = 0$. At small q , $f(q)$ has a slope of approximately p_1 . In this part of the function, when q is small, there can be a weak relationship between $f(q)$ and moments. In comparison,

when q is close to 1, the moments are closely related to $f(q)$. For example $f'(1) = m_1$. Higher moments begin to influence the function as q moves away from 1.

The probability of extinction of a process is found when $f(q) = q$, i.e. at the intersect between its probability generating function $f(q)$ and the diagonal q . Thus, processes with a high probability of extinction will cross the diagonal near $q = 1$, in the domain of q in which the probability generating function is often closely related to its first few moments.

Plotting the probability generating functions for the extremal distributions helps demonstrate why they act as bounds on extinction. In these examples (Figure 1), we compare two distributions with identical first moment and maximum ($m_1 = 2$, $n = 20$), i.e. both distributions are in $\mathfrak{B}_{2,20}^2$. In particular, we look at a binomial distribution and a truncated geometric distribution. For each of these plots we also plot the generating functions for some of the extremal distributions. The extremal distributions provide clear bounds: best case extrema are found below the plot of the generating function, worst case extrema are found above. For example, the extremal distribution based on one moment, $X_{\max}^{(2)}$, provides an upper bound on the probability of extinction, and can be seen as the upper line in both plots. Because they share an identical first moment and maximum, $X_{\max}^{(2)}$ is the same for both distributions. Clearly, one moment does not provide a good bound in these examples. As more moments are used, the bounds become tighter. The extrema using four moments provide relatively accurate upper and lower bounds for both examples. The lower bounds provide the best case extrema, which are useful in both cases only when three or four moments are known. The lower bound using two moments is



not useful here in either case, as its probability generating function crosses the diagonal at zero so its probability of extinction is zero. The lower bound using only one moment was not included because its generating function is trivial and always uninformative about extinction.

Importantly, these examples demonstrate why higher moments are often necessary to compare strategies. These two distributions have identical first moments ($m_1 = 2$) so classically their fitness value would be equal. However, the binomial example is more likely to survive. If entire distributions are known, then extinction probabilities can be calculated explicitly using (1). This requires solving a polynomial of degree 20 for the examples shown here, which were solved in R (R Development Core Team 2011) with the package “rootSolve” (Soetaert and Herman 2009).

If instead the moments are known, the extremal distributions can be found and the roots of their generating functions can be solved to find the bounds on extinction. These roots can again be solved in R (R Development Core Team 2011) with the package “rootSolve” (Soetaert and Herman 2009). However, some of these extremal generating functions are relatively simple and can be solved by hand. For example, our binomial distribution ($\text{Bin}_{20,0.1}$) has $m_1 = 2$ and $m_2 = 5.8$. The resulting $X_{\min}^{(3)}$ has supports at 0, 2 and 3 with respective probabilities of 0.3, 0.1 and 0.6, leading to its generating function

$$q = 0.3 + 0.1q^2 + 0.6q^3.$$

Using the knowledge that the generating function has a root at $q = 1$, this equation can be factorized as:

$$0 = (q - 1)(3q - 1)(2q + 3).$$

The probability of extinction is the smallest positive root of the above equation, $1/3$ (Table 1), providing an upper bound on extinction.

If four moments are known, one can conclude that the truncated geometric distribution has a higher probability of extinction. Compare the extremal distributions when four moments are known, paying attention to where they cross the diagonal. The value at the intersect is the probability of extinction for the extrema, which we display in Table 1 and Table 2, respectively for the binomial example and the truncated geometric example. Using four moments, the best case for the truncated geometric example (0.404, Table 2) is worse than the worst case for binomial example (0.207, Table 1). In fact, the worst case for the binomial example using two moments (0.333) is already better than the best case for the truncated geometric using four moments (0.404). These examples highlight how moment spaces can be used to rank branching processes by their extinction probabilities when only moments of their distributions are known.

Table 1 Extinction probabilities and supports for the extremal distributions of the Binomial example $B_{20,0.1}$

	1 moment	2 moments	3 moments	4 moments
Supports	{2}	{1,2,20}	{0,1,3,4}	{0,1,3,4,20}
Lower bound (best)	0.000	0.000	0.034	0.083
Supports	{0,20}	{0,2,3}	{0,2,3,20}	{0,1,2,4,5}
Upper bound (worst)	0.918	0.333	0.306	0.207

The actual probability of extinction for this process is 0.181.

Table 2 Extinction probabilities and supports for the extremal distributions of the truncated Geometric example

	1 moment	2 moments	3 moments	4 moments
Supports	{2}	{1,2,20}	{0,1,7,8}	{0,1,6,7,20}
Lower bound (best)	0.000	0.000	0.110	0.404
Supports	{0,20}	{0,4,5}	{0,4,5,20}	{0,3,4,11,12}
Upper bound (worst)	0.918	0.641	0.592	0.534

The actual probability of extinction for this process is 0.499.

And finally, these examples can be used to better understand how ranking distributions using their *s*-convex extrema can be useful in investing and gambling (Canjar 2007; Courtois et al. 2006; Denuit and Lefevre 1997; Ethier and Khoshnevisan 2002; Hürlimann 2005). If these distributions were returns on an investment or gamble, then by comparing their moments an investor could determine that the binomial distribution is a superior investment model. Both distributions would provide the same expected growth on capital, but the geometric distribution would have a higher probability of gambler's ruin. Being wary of gambler's ruin is especially important for an investor with limited initial funds for their investment.

4 Conclusion

The work here is intended to highlight the relationship between the moments of the offspring distribution and the probability of extinction. Extinction can be defined in terms of moments, but the first few moments are only informative about extinction under certain conditions. Nevertheless, for all offspring distributions there exists an interesting relationship with even and odd moments: high even moments favor extinction, high odd moments favor survival. This relationship between even and odd moments is also seen in the stochastic Price equation, where relative growth rates increase with increasing odd moments, and decrease with increasing even moments (Rice 2008).

The relationship between moments and extinction can provide insight into the evolutionary process. A high first moment can favor survival, but worst case extrema ("long shots") represent the strategies that are least likely to survive. Strategies with a relatively low second moment (low variance) will always have a lower probability of extinction than their corresponding "long shot" extrema. When two moments are known, the worst case distributions have the lowest third moment (strongest right skew). Therefore, strategies with identical first and second moments and relatively high third moments (strong left skew) will always have a better chance at survival than the extrema with the lowest third moment. Worst case extrema using three moments have the highest possible fourth moment (excessive kurtosis). The relative importance of higher moments depends on the distribution, and in some cases higher moments can have a big influence on extinction.

Strategies with a high probability of extinction are unlikely to be found in natural populations, even if their expected reproductive rate is high (Tuljapurkar and Orzack 1980). New alleles will often arrive in a population as a singlet, and extinction is permanent unless the same mutation occurs more than once. In such cases, survival is more important than the average rate of reproduction. Using moments of the offspring distribution one can find bounds on extinction using their *s*-convex extrema. If the best case extrema

for a set of moments has a high probability of extinction, then strategies with these moments will be evolutionarily unlikely, regardless of how fit these strategy would be if they avoided extinction.

Gamblers can avoid strategies with a high risk of ruin by calculating their odds. In natural populations, such calculations are not required to prevent the occurrence of high risk strategies. Instead, risky strategies will be naturally unlikely, especially considering that many arrive as a single allele with one chance at survival. Similarly, gamblers and investors who begin with limited funds and chose risky strategies are likely to “go extinct” through gambler’s ruin. Risk is not solely determined by mean growth, and strategies with a high mean can sometimes have high risk. Unfortunately, these high risk and high reward strategies are unlikely to return anything without sufficient investment, so natural avoidance of risk can result in missed opportunity for growth.

Competing interests

The authors declare that they have no competing interests.

Authors’ contributions

SS and SK both conceived, designed and developed the methods and results, which were then interpreted by both authors. The manuscript was primarily written by SS, with help from SK. Both authors read and approved the final manuscript.

Acknowledgement

The authors like to thank Ninh Anh for a fruitful discussion on calculating the supports and probabilities for the extremal distributions, and the many constructive comments by the editor and two anonymous referees.

Author details

¹Institute for Behavioral Genetics, University of Colorado, Boulder, Colorado, USA. ²Department of Statistics and School of Biological Sciences, University of Auckland, Auckland, New Zealand. ³Formerly: Department of Anatomy, and Allan Wilson Centre for Molecular Ecology and Evolution, University of Otago, Dunedin, New Zealand.

Received: 22 December 2013 Accepted: 6 May 2014

Published: 16 June 2014

References

- Canjar, RM: Gambler’s Ruin revisited: The effects of skew and large jackpots. In: Ethier, S, Eadington, W (eds.) *Optimal Play: Mathematical Studies of Games and Gambling*. Institute for the Study of Gambling and Commercial Gaming, pp. 439–469. University of Nevada, Reno, (2007)
- Courtois, C, Denuit, M, van Bellegem, S: Discrete s -convex extremal distributions: theory and applications. *Appl. Math. Lett.* **19**, 1367–1377 (2006)
- Daley, DJ, Narayan, P: Series expansions of probability generating functions and bounds for the extinction probability of a branching process. *J. Appl. Probab.* **17**, 939–947 (1980)
- Denuit, M, Lefevre, C: Some new classes of stochastic order relations among arithmetic random variables, with applications in actuarial sciences. *Insur. Math. Econ.* **20**, 197–213 (1997)
- Denuit, M, de Vylder, E, Lefevre, C: Extremal generators and extremal distributions for the continuous s -convex stochastic orderings. *Insur. Math. Econ.* **24**, 201–217 (1999)
- Denuit, M, Lefevre, C, Mesfioui, M: On s -convex stochastic extrema for arithmetic risks. *Insur. Math. Econ.* **25**, 143–155 (1999)
- Ethier, SN, Khoshnevisan, D: Bounds on gambler’s ruin probabilities in terms of moments. *Methodol. Comput. Appl. Probab.* **4**, 55–68 (2002). 10.1023/A:1015705430513
- Hürlimann, W: Improved analytical bounds for gambler’s ruin probabilities. *Methodol. Comput. Appl. Probab.* **7**, 79–95 (2005). 10.1007/s11009-005-6656-4
- Karlin, S, McGregor, JL: The differential equations of birth-and-death-processes, and the stieljtjes moment problem. *T. Am. Math. Soc.* **85**, 489–546 (1957)
- Kelly, J: A new interpretation of information rate. *Bell Sys. Tech. J.* **35**, 917–926 (1956)
- Kimmel, M, Axelrod, DE: *Branching Processes in Biology*. Springer, New York (2002)
- Lewontin, RC, Cohen, D: On population growth in a randomly varying environment. *P. Natl. Acad. Sci. USA.* **62**, 1056–1060 (1969)
- MacLean, LC, Thorp, EO, Ziemba, WT: Good and bad properties of the Kelly criterion. In: *The Kelly Capital Growth Investment Criterion: Theory and Practice*, pp. 563–574. World Scientific Publishing, Singapore, (2010)
- Prékopa, A: The discrete moment problem and linear programming. *Discrete Appl. Math.* **27**, 235–254 (1990)
- R Development Core Team: R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria (2011). <http://www.R-project.org>

Rice, SH: A stochastic version of the Price equation reveals the interplay of deterministic and stochastic processes in evolution. *BMC Evol. Biol.* **8**, 262 (2008)

Shaked, M, Shanthikumar, JG: *Stochastic Orders*. Springer, New York (2007)

Soetaert, K, Herman, PMJ: *A practical guide to ecological modelling. Using r as a simulation platform*. Springer, New York (2009)

Tuljapurkar, S, Orzack, SH: Population dynamics in variable environments I. Long-run growth rates and extinction. *Theor. Popul. Biol.* **18**, 314–342 (1980)

doi:10.1186/2195-5832-1-10

Cite this article as: Sawaya and Klaere: **Extinction in a branching process: why some of the fittest strategies cannot guarantee survival.** *Journal of Statistical Distributions and Applications* 2014 **1**:10.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com
