A Scaling Result for Explosive Processes

M. Mitzenmacher* J. Spencer

We consider the following balls and bins model, as described in [2, 4]. Balls are sequentially thrown into bins so that the probability that a bin with x balls obtains the next ball is proportional to x^p for some constant $p > 1$. Specifically, we consider the case of two bins, in which case the state (x, y) denotes that bin 1 has x balls and bin 2 has y balls. In this case, the probability that the next ball lands in bin 1 is $\frac{x^p}{x^p+y^p}$.

This model is motivated by the phenomenon of *positive feedback*. In economics, positive feedback refers to the situation where a small number of companies compete in a market until one obtains a non-negligible advantage in the market share, at which point its share rapidly grows to a monopoly or near-monopoly. One loose explanation for this principle, commonly referred to as Metcalfe's Law, is that the inherent potential value of a system grows super-linearly in the number of existing users. Positive feedback also occurs in chemical and biological processes; for more information, see e.g. [1]. Here we consider positive feedback between two competitors, with the strength of the feedback modeled by the paramter p. Our methods can also be applied to similar problems with more competitors.

It is known that for the model above that when $p > 1$ eventually one bin obtains a monopoly in the following sense: as the total number of balls n goes to infinity, one bin obtains $n - o(n)$ balls with probability 1 [2, 4]. Indeed, with probability 1 there exists a time after which all subsequent balls fall into just one of the bins [4]. Given this limiting behavior, we now ask what is the probability that bin 1 will eventually obtain the monopoly starting from state (x, y). We provide an asymptotic analysis, based an examining the appropriate scaling of the system. This approach is reminiscent of techniques used to study phase transitions in random graphs, as well as other similar phenomena.

Let $a = (x + y)/2$. We show that when $x = a + \lambda\sqrt{a}$ that in the limit as a grows large, the probability that x obtains the monopoly converges to $\Phi(\lambda\sqrt{4p-2})$, where Φ is the cumulative distribution function for the normal distribution with mean 0 and variance 1.

Theorem 1 *For the balls and bins theorem described above, from the state (x, y), the probability that bin 1 obtains the eventual monopoly is $\Phi(\lambda\sqrt{4p-2}) + O(1/\sqrt{a})$.*

*Supported in part by an Alfred P. Sloan Research Fellowship and NSF grants CCR-9983832, CCR-0118701, and CCR-0121154.

Proof: We follow the approach developed in [4]. We adopt a useful but non-intuitive notion of time; a bin with z balls at time t receives its next ball at a time $t+T_z$, where T_z is a random variable exponentially distributed with mean z^{-p}. From the properties of the exponential distribution, we can deduce that this maintains the property that the probability that the next ball lands in a bin with x balls is proportional to x^p. Specifically, the probability that the minimum of two exponentially distributed random variables T_x with mean x^{-p} and T_y with mean y^{-p} is T_x with probability $\frac{x^p}{x^p+y^p}$. Moreover, from the memorylessness of the exponential distribution, when a ball arrives at bin 1 (respectively, bin 2) the time T_x (T_y) until the next ball arrives at bin 2 (bin 1) is still exponentially distributed with the same mean.

The *explosion time* for a bin is the time under this framework when a bin receives an infinite number of balls. If we begin at time 0, the explosion time F_x for bin 1 satisfies

$$F_x = \sum_{j=x}^{\infty} T_j = \sum_{j=a+\lambda\sqrt{a}}^{\infty} T_j$$

and similarly for bin 2. Note that $E[F_x]$ and $E[F_y]$ are finite; indeed, the explosion time for each bin is finite with probability 1. It is therefore evident that the bin with the smaller explosion time at some point obtains all balls thrown past some point, as proven formally in [4].

We first demonstrate that for sufficiently large a, F_x and F_y are approximately normally distributed. This would follow immediately from the Central Limit Theorem (specifically, the variation where random variables are independent but not necessarily identically distributed) if the sum of the variances of the random variables T_j grew to infinity. Unfortunately,

$$\sum_{j=x}^{\infty} \mathrm{Var}[T_j] = \sum_{j=x}^{\infty} j^{-2p} < \infty,$$

and hence standard forms of the Central Limit Theorem do not apply.

Fortunately, we may apply Esseen's inequality, a generalization of the Central Limit Theorem, which can be found in, for example, [3][Theorem 5.4].

Lemma 1 (Esseen's inequality) *Let $X_1, X_2, \ldots, X_n$ be independent random variables with $E[X_j] = 0$, Var$[X_j] = \sigma_j^2$, and $E[|X_j|^3] < \infty$ for $j = 1, \ldots, n$. Let $B_n = \sum_{i=0}^{n} \sigma_j^2$, $F(x) = \mathbf{Pr}(B_n^{-1/2} \sum_{j=1}^{n} X_j < x)$, and $L = B_n^{-3/2} \sum_{j=1}^{n} E[|X_j|^3]$. Then*

$$\sup_x |F(x) - \Phi(x)| \le cL$$

for some universal constant c.

In our setting, let $X_j = T_{x+j-1} - (x+j-1)^{-p}$. We note that there are no problems applying Esseen's theorem to the infinite summations of our problem. Consider

$$F^x(z) = \mathbf{Pr}\left(\frac{\sum_{j=x}^{\infty}(T_j - j^{-p})}{\sqrt{\sum_{j=x}^{\infty} j^{-2p}}} < z\right).$$

That is, $F^x(z)$ is the probability that F_x, appropriately normalized to match a standard normal of mean 0 and variance 1, is less than or equal to z. Then we have

$$\sup_z |F^x(z) - \Phi(z)| \leq O\left(\frac{1}{\sqrt{x}}\right).$$

Hence $F^x(z)$ approaches a normal distribution as x grows large.

We also have

$$E[F_x] = \sum_{j=x}^{\infty} E[T_j] = \sum_{j=x}^{\infty} \frac{1}{j^p} = \int_{j=x}^{\infty} \frac{1}{j^p} dj + O\left(\frac{1}{x^p}\right) = \frac{x^{1-p}}{p-1} + O(x^{-p}),$$

and

$$\mathrm{Var}[F_x] = \sum_{j=x}^{\infty} \mathrm{Var}[T_j] = \sum_{j=x}^{\infty} \frac{1}{j^{2p}} = \int_{j=x}^{\infty} \frac{1}{j^{2p}} dj + O\left(\frac{1}{x^{2p}}\right) = \frac{x^{1-2p}}{2p-1} + O(x^{-2p}).$$

We wish to determine the probability that $F_x - F_y < 0$. Now $F_x - F_y$ is (approximately) normally distributed with mean μ where

$$\begin{aligned}
\mu = E[F_x] - E[F_y] &= \frac{(a+\lambda\sqrt{a})^{1-p}}{p-1} - \frac{(a-\lambda\sqrt{a})^{1-p}}{p-1} + O(x^{-p}) + O(y^{-p}) \\
&= \frac{1}{p-1}\frac{(a-\lambda\sqrt{a})^{p-1} - (a+\lambda\sqrt{a})^{p-1}}{(a^2-\lambda^2 a)^{p-1}} + O(a^{-p}) \\
&= -2\lambda a^{1/2-p} + O(a^{-p})
\end{aligned}$$

and variance σ^2 where

$$\begin{aligned}
\sigma^2 = \mathrm{Var}[F_x] + \mathrm{Var}[F_y] &= \frac{(a+\lambda\sqrt{a})^{1-2p}}{2p-1} + \frac{(a-\lambda\sqrt{a})^{1-2p}}{2p-1} + O(x^{-2p}) + O(y^{-2p}) \\
&= \frac{1}{2p-1}\frac{(a-\lambda\sqrt{a})^{2p-1} + (a+\lambda\sqrt{a})^{2p-1}}{(a^2-\lambda^2 a)^{2p-1}} + O(a^{-2p}) \\
&= \frac{2}{2p-1} a^{1-2p} + O(a^{-2p})
\end{aligned}$$

Hence the probability that $F_x - F_y < 0$ is $\Phi(\lambda\sqrt{4p-2} + O(1/\sqrt{a})) + O(1/\sqrt{a})$, which is just $\Phi(\lambda\sqrt{4p-2}) + O(1/\sqrt{a})$. $\square$

We also present a heuristic argument, which yields the same result. While this argument does not (yet) provide a formal proof, we believe that variations of this heuristic approach may be useful for studying similar questions.

Let $G(\lambda, a)$ be the probability that bin 1 achieves monopoly when $x = a + \lambda\sqrt{a}$ and $y = a - \lambda\sqrt{a}$. From state (x, y), after the next is ball thrown, the new average a' equals $a' + 1/2$ and we have a new values x' and λ' so that $x' = a' + \lambda'\sqrt{a'}$. With probability $\frac{(a+\lambda\sqrt{a})^p}{(a+\lambda\sqrt{a})^p+(a-\lambda\sqrt{a})^p}$, $x' = x + 1$, and hence

$$\begin{aligned} a' + \lambda'\sqrt{a'} &= a + 1 + \lambda\sqrt{a} \\ \lambda' &= \frac{1/2 + \lambda\sqrt{a}}{\sqrt{a + 1/2}}. \end{aligned}$$

Similarly, with probability $\frac{(a-\lambda\sqrt{a})^p}{(a+\lambda\sqrt{a})^p+(a-\lambda\sqrt{a})^p}$, we have

$$\lambda' = \frac{-1/2 + \lambda\sqrt{a}}{\sqrt{a + 1/2}}.$$

Hence,

$$G(\lambda, a) = \left(\frac{1}{2} + \Delta_1\right) G(\lambda + \Delta_2 - \Delta_3, a + 1/2) + \left(\frac{1}{2} - \Delta_1\right) G(\lambda - \Delta_2 - \Delta_3, a + 1/2)$$

for $\Delta_1 = \frac{(a+\lambda\sqrt{a})^p}{(a+\lambda\sqrt{a})^p+(a-\lambda\sqrt{a})^p} - \frac{1}{2}$, $\Delta_2 = 1/2\sqrt{a + 1/2}$, and $\Delta_3 = \lambda\sqrt{a}/\sqrt{a + 1/2} - \lambda$.

Suppose that there exists a continuous and twice differentiable function $F(\lambda)$ such that $F(\lambda) = \lim_{a\to\infty} G(\lambda, a)$. So

$$F(\lambda) = \left(\frac{1}{2} + \Delta_1\right) F(\lambda + \Delta_2 - \Delta_3) + \left(\frac{1}{2} - \Delta_1\right) F(\lambda - \Delta_2 - \Delta_3).$$

Now in the limit as Δ_1, Δ_2, and Δ_3 go to 0 as a grows large, with $\Delta_3 = o(\Delta_2)$

$$\begin{aligned} 0 &= \frac{F(\lambda + \Delta_2 - \Delta_3) - F(\lambda)}{2} - \frac{F(\lambda) - F(\lambda - \Delta_2 - \Delta_3)}{2} \\ &\quad + \Delta_1(F(\lambda + \Delta_2 - \Delta_3) - F(\lambda - \Delta_2 - \Delta_3)) \\ &= \frac{F'(\lambda)(\Delta_2 - \Delta_3)}{2} - \frac{F'(\lambda - \Delta_2 - \Delta_3)(\Delta_2 + \Delta_3)}{2} \\ &\quad + \Delta_1(F(\lambda + \Delta_2 - \Delta_3) - F(\lambda - \Delta_3) + F(\lambda - \Delta_3) - F(\lambda - \Delta_2 - \Delta_3)) \\ &= \frac{(\Delta_2 + \Delta_3)^2}{2} F''(\lambda - \Delta_2 - \Delta_3) - \Delta_3 F'(\lambda) + \Delta_1\Delta_2(F'(\lambda - \Delta_3) + F'(\lambda - \Delta_2 - \Delta_3)). \end{aligned}$$

Hence we have the limiting equation

$$F''(\lambda) + \left(\frac{4\Delta_1}{\Delta_2} - \frac{2\Delta_3}{\Delta_2^2}\right) F'(\lambda) = 0.$$

As a grows large, $4\Delta_1/\Delta_2$ converges to $4p\lambda$ and $2\Delta_3/\Delta_2^2$ converges to 2λ. So we require

$$F''(\lambda) + (4p-2)\lambda F'(\lambda) = 0.$$

Substituting $H(\lambda) = F'(\lambda)$, it is easy to solve to find

$$F(\lambda) = \Phi(\lambda\sqrt{4p-2}),$$

which is indeed the correct limiting result.

We provide an example demonstrating the accuracy of Theorem 1 in Table 1. We consider initial states with 200 balls in the system, with the first bin containing between 101 and 110 balls (so λ ranges from 0.1 to 1). We calculate the exact distribution when there are 160,000 balls in the system for the case $p = 2$, using methods described in [2]. With this data, we make the very accurate approximation bin 1 eventually achieves monopoly if it has 53% of the balls at this point. We also apply symmetry; if at this point bin 1 has k balls with probability p_1 and bin 2 has k balls with probability $p_2 < p_1$, then bin 1 reaches monopoly at least p_2 out of this $p_1 + p_2$ fraction of the time. This approach is sufficent to accurately determine the probability that the first bin eventually reaches monopoly to four decimal places. Comparing these results demonstrates the accuracy of the normal estimate. Table 2 shows similar results for the case of $p = 1.5$; here we calculate the distribution with 640,000 balls in the system and use a 52% cutoff to estimate the probability of monopoly, and the numbers are correct to four decimal places. Again, the normal estimate provides a great deal of accuracy.

References

[1] B. Arthur. **Increasing Returns and Path Dependence in the Economy**. The University of Michigan Press, 1994.

[2] E. Drinea, A. Frieze, and M. Mitzenmacher. Balls and Bins Models with Feedback. In *Proceedings of SODA 2002*.

[3] V. Petrov. **Limit Theorems of Probability Theory**. Oxford University Press, 1995.

[4] J. Spencer and N. Wormald. Explosive processes. Draft manuscript.

x	101	102	103	104	105
Calc.	0.5955	0.6870	0.7682	0.8361	0.8896
$\Phi(\lambda\sqrt{4p-2})$	0.5970	0.6883	0.7693	0.8370	0.8902
x	106	107	108	109	110
Calc.	0.9292	0.9569	0.9751	0.9863	0.9929
$\Phi(\lambda\sqrt{4p-2})$	0.9297	0.9572	0.9753	0.9865	0.9930

Table 1: A calculation vs. the asymptotic estimate of our theorem when $a = 100$ and $p = 2$.

x	101	102	103	104	105
Calc.	0.5794	0.6557	0.7261	0.7886	0.8419
$\Phi(\lambda\sqrt{4p-2})$	0.5793	0.6554	0.7257	0.7881	0.8413
x	106	107	108	109	110
Calc.	0.8854	0.9197	0.9456	0.9644	0.9775
$\Phi(\lambda\sqrt{4p-2})$	0.8849	0.9192	0.9452	0.9641	0.9772

Table 2: A calculation vs. the asymptotic estimate of our theorem when $a = 100$ and $p = 1.5$.