Recruitment in a Pyramid Scheme

1 Abstract

Pyramid schemes are businesses which encourage recruitment. As there is a finite limit to the population, a large percentage of the participants in the scheme lose out as they cannot recruit. This paper studies this phenomenon by proving the following:

Theorem. Consider a pyramid scheme where each participant has an equal likelihood of hiring, and where there is a finite limit to the scheme. The expected number of participants that can anticipate recruiting at least one participant is approximately the first 37% of the pyramid scheme population.

This paper is based on Gastwirth (1977).

2 Definitions and Assumptions

According to the US Federal Trade Commission's Koscot test, a business is a pyramid scheme if a participant pays to receive the right to sell a product, but receives more significant rewards for recruitment than for end-user sales (Keep, 2002).

Three assumptions are used to prove the theorem:

A1. All participants have the same probability of hiring the next participant. Suppose 10 participants were in a scheme, the probability of each of the participants hiring the 11th person would be 0.1. This assumes all participants have equal resources when it comes to recruiting. Part of this assumption is that each participant is always actively trying to recruit, and that a participant cannot exit the scheme:

$$P(Hiring\ k+1) = \frac{1}{k}$$

Note that k is the number of participants in the pyramid scheme; the k^{th} participant represents any participant through to N; and the k_i participant represents a particular value with index, i, which runs through all integers from a lower bound, 1, to an upper bound, N. Thus:

A2. Let N be finite; this is size of the pool of potential participants.

A3. Recruitment is a result of 'n' number of Bernoulli trials that meets the following conditions: each trial is independent, and each trial can either 'succeed' or 'fail'.

3 Main Result

Proof.

"A binomial random variable is the number of successes in n consecutive independent Bernoulli trials" (Monroe, 2017). R is a discrete binomial random variable, which represents the number of recruitment 'successes'. R_i is a Bernoulli random variable, which represents success or failure of recruitment for a particular trial. Thus, R is the number of successful Bernoulli trials (where $R_i = 1$).

Therefore, $R \sim Bin(n, p)$, with n being the number of Bernoulli trials:

- n is the number of observations (the amount of times k can hire). k can hire everyone after themselves up until the last person. The last person to be recruited is N, so once N is involved in the scheme, there is no one left to recruit. Therefore, there is no probability of recruitment at N (n ends at $N-1$). Therefore $n = (N-1) - k$.

- p is the probability of hiring such that $p = \frac{1}{i}$.
- R_i is a Bernoulli random variable, that can take two values: 1, when successful (hiring) and 0 when failure (not hiring). Therefore, R is the sum of successes for the n Bernoulli trials:

$$R = \sum_{i=k}^{N-1} R_i : where\ P(R_i = 1) = p = \frac{1}{i}\ and\ P(R_i = 0) = q = 1 - \frac{1}{i}$$

The expectation of a binomial variable is the sum of probabilities: $E(R) = np = p_1 + p_2 + \cdots + p_n$. $E(R)$ is the expected number of people a participant can anticipate to recruit. Therefore, $E(R)$ for the Bernoulli trials between k and $N-1$ is the sum of successes:

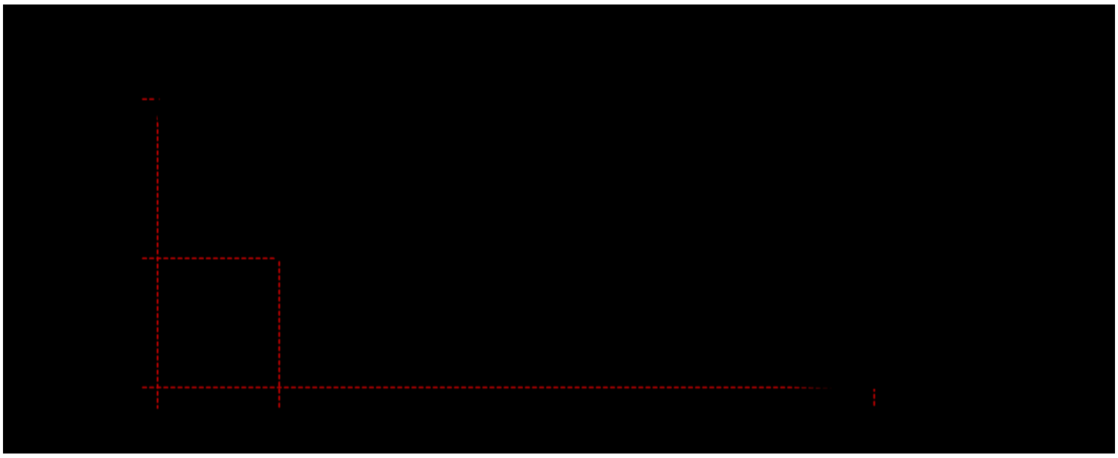
$$E(R) = \sum_{i=k}^{N-1} \frac{1}{i}$$

Therefore, $E(R)$ is the sum of all probabilities of success (where $R_i = 1$), between k, the position of entry (so the first opportunity to hire), and $N-1$, the last opportunity to hire:

$$E(R) = \frac{1}{k} + \frac{1}{k+1} + \cdots + \frac{1}{N-1} \quad (1)$$

The simplest way to demonstrate the relationship between the number of participants in the scheme, k, and the probability of hiring, $P(k+1)$, is to draw a diagram.

The domain is restricted to $[1, N-1]$: the lower bound is the scheme's founder; $N-1$ is the last agent who can expect to recruit another agent. In other words, once N is involved in the scheme, there is no one left to recruit, so there is no probability of recruitment.



Adding all the individual probabilities of success can be approximated to adding all of the 'strips' beneath the curve in Figure 1. Therefore, integrating the curve would give an estimate for the expected number of recruits. Note that from now on, $E(R)$ is being approximated, this is explained later on and in the discussion. Therefore, $E(R)$ can be approximated as:

$$E(R) \approx \int_{k_i}^{N-1} \frac{1}{k}\, dk \quad (2)$$

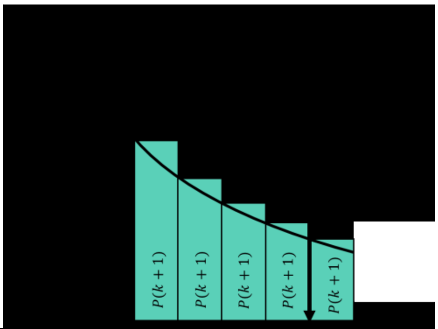
The k^{th} participant can only recruit people who join after themselves. This is the area beneath the curve, constrained by k_i and the final hiring opportunity, when $k = N-1$.
Integrating (2), and inputting these bounds gives us the following:

$$E(R) \approx [\ln(k)]_{k_i}^{N-1}$$
$$E(R) \approx \ln(N-1) - \ln(k_i)$$

We re-write this expression for the expected number of people that participant k_i can anticipate to recruit as:

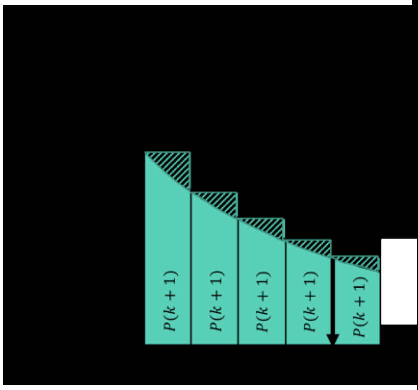
$$E(R) \approx \ln\left(\frac{N-1}{k_i}\right) \qquad (3)$$

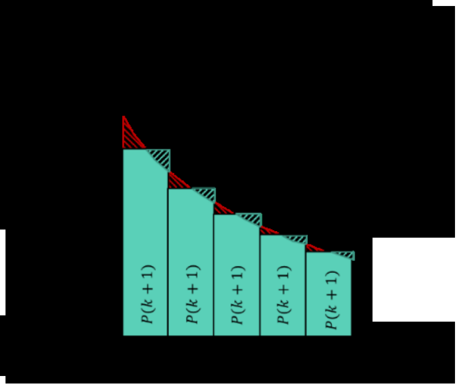
However, there are two adjustments that need to be made to this formula:



Firstly, the model excludes the probability of $k = N - 1$ recruiting. Figure 2 shows that when $N = 6$, the current model would ignore the probability of hiring the 6th participant. This is because the probability should have a width of 1, as k is a discrete variable, on a continuous axis. This ensures that $width \times height = probability$. The solution to this is to adjust the boundaries to include N, so that the probability of hiring the 6th participant is included:

$$E(R) \approx \ln\left(\frac{N}{k_i}\right) \qquad (4)$$





Secondly, Figure 3 shows that the model underestimates the probability of hiring, as the curve 'truncates' the probability. Again, $width \times height = probability$. So, if the height is not consistent, and gets truncated, the probability is underestimated.

This chronic underestimation converges to the Euler-Mascheroni constant, γ, which roughly equals 0.5772 as k approaches ∞. This is the sum of all the underestimations. This constant is the difference between the harmonic series and the natural logarithm function. [$\gamma = H_m - \ln n$]. Given a population of 500, the true R for the first participant (H_m, between 1 and $\frac{1}{N-1}$) is different from the model ($\ln\left(\frac{N}{1}\right)$) by approximately γ:

$$\gamma = H_m - \ln N = \frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{500} - \ln 501 = 6.792823 \ldots - 6.216606 \ldots = 0.576217$$

To reduce this error, the probability columns should be re-centred, with k as their midpoints. This is shown in Figure 4. There is still an underestimation occurring, but it is less than γ. Equation (5) shows this adjustment in the model:

$$E(R) \approx \ln\left(\frac{N-0.5}{k_i-0.5}\right) \qquad (5)$$

To check these results, we can use an example. If $N = 11$, what is the possibility of the 10th person recruiting? The 10th participant would only have one opportunity to hire and should have a 0.1 probability of hiring. Thus, the expected number of recruits should be 0.1 for this participant:

$$E(R) \approx \ln\left(\frac{11-0.5}{10-0.5}\right) \approx \ln\left(\frac{10.5}{9.5}\right) \approx 0.10008347 \ldots \approx 0.10$$

This model is close to the true result; the small values of N and k_i have caused a moderate degree of inaccuracy, but this example used small values for simplicity.

This paper is aiming to find the number of participants that are expected to recruit at least one people. If we assume N and k_i are large, we can use Equation (4), above. The validity of this is discussed later.

To find the value of k when $1 \leq E(R)$, $1 \leq E(R)$, is substituted into the equation:

$$1 \lessapprox \ln\left(\frac{N}{k_i}\right)$$

The next step is to raise e to the power of both sides:

$$e^1 \lessapprox \frac{N}{k_i}$$

Which can be rearranged to:

$$k_i \lessapprox \frac{N}{e^1} \lessapprox 0.367879 \ldots N \lessapprox 0.37\text{N}$$

Therefore, the value of k_i, for which it can be expected to recruit one or more new participants is the first 37% of the population.

This ends the proof.

This methodology can be extended for those expecting to recruit 2 or more, 3 or more, and beyond:

$$E(R) \gtrapprox 2: \; k_i \lessapprox \frac{N}{e^2} \lessapprox 0.145N$$

$$E(R) \gtrapprox 3: \; k_i \lessapprox \frac{N}{e^3} \lessapprox 0.049N$$

4 Applying the model

Further results can be derived using Equation (5). To do this, another assumption is added:

A4. The only source of revenue is from recruiting. This follows the definition of a pyramid scheme, as many of these schemes contain worthless or overvalued products that are difficult to sell:

$$Total\ Revenue = Recruitment\ Bonus \times Number\ of\ Recruits$$

We use Excel to generate random values for N and k, where $k \leq N$, in order to estimate the expected number of recruits, $E(R)$, and their expected profit, $E(\Pi)$. For example, inputting randomly chosen $N = 879{,}293$ *and* $k = 190{,}069$ into Equation (5), the randomly chosen participant is expected to recruit 1.53 people:

$$E(R) \approx \ln\left(\frac{879{,}292.5}{190{,}068.5}\right) \approx 1.53173 \ldots \approx 1.53$$

To join the Bull Investment Group, which was found to be an illegal pyramid scheme, the initial investment cost was \$2,500 and the recruitment bonus was \$900 (Gastwirth, 1977). Using these figures, and assuming A3, this participant's expected profit can be estimated:

$$E(\Pi) \approx \$900 \times E(R) - \$2{,}500 \approx \$900 \times 1.53 - \$2{,}500 \approx -\$1{,}123$$

So, for this value of k, the participant is expected to make a loss. This application can be used for other values of N and k. Figures 5 and 6 display the result of picking 10,000 random values.

Fig.5

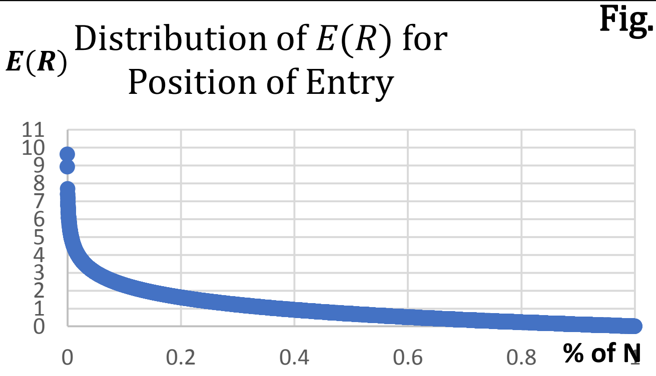
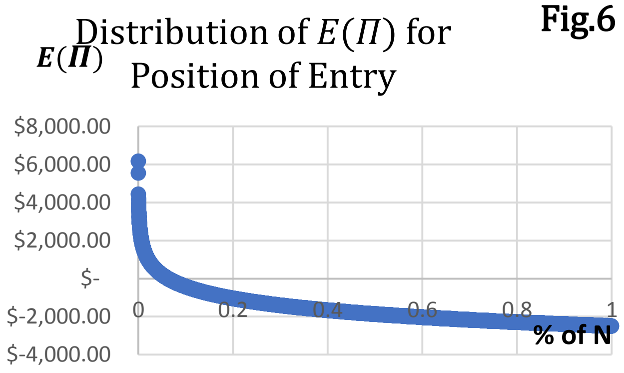


Fig.6



It can be seen that there is exponential decay for both $E(R)$ and $E(\Pi)$. For 36.56% of the simulated partcipants $E(R) \geq 0$, consistent with the theorem's result of c. 37%. Only 612 of the observations had an $E(\Pi) \geq 0$, using the Bull Investment Group's initial cost and recruitment benefits. To break even, a participant would have to recruit 2.778 people. Note that: $E(R) \gtrapprox 2.8$: $k_i \lessapprox \frac{N}{e^{2.8}} \lessapprox 0.0621N$. This shows that 6% of the population is expected to break even. Here, the simulation backs the theorem.

5 Discussion

This theorem proves that the majority of participants are unable to recruit, let alone recruit enough new participants to cover the costs of joining the scheme. Many of these schemes advertise income levels based on recruitment goals, proven to be unrealistic.

However, being in the bottom 63% of the scheme doesn't mean you can't recruit. It could be that everyone hires 1 person, apart from N. Likewise, being in the first 37% doesn't guarantee recruitment. An expected value is not a defining amount, there will be natural variation from this. This is obvious when $E(R)$ is not an integer; it is impossible to recruit a fraction of a person. Some participants will recruit more than their $E(R)$, and some less.

Throughout this paper, $E(R)$ has only been approximated, as the natural logarithm underestimates $E(R)$. This is because the harmonic series, $\sum_{n=1}^{\infty} \frac{1}{n}$, is being approximated by $\ln n$, but they are not identical. The error between the two is the Euler-Mascheroni constant, γ. The model in this paper will have a smaller error as n is finite $[(N-1)-k]$. The main theorem ignored γ. This is due to large values of N and k; with large values, the underestimation is minimised. A significant proportion of γ comes from early values of k (As seen in Figure 3). Therefore, the underestimation after 37% of N is small. Therefore, the natural logarithm is almost identical to the harmonic form with these large values. This means that the adjusted Equation (5) is not required: Equation (4) is sufficient.

Although this model does illustrate the difficulty of recruiting, it may do so too simplistically. Some of the assumptions may not be valid. Realistically, some participants will have more resources, such as money and time, which is likely to increase their probability of hiring. Therefore, assumption A1 is too simple. If some participants probability of hiring was greater than $\frac{1}{k}$, the percentage of N that can expect to recruit at least one person would be lower. It was also assumed that N and k are large, so that Equation (4) could be used. This is usually true, such as in the UK's 'Give and Take' pyramid scheme which involved at least 10,000 victims (BBC News). However, sometimes pyramid schemes can be considerably smaller, so using Equation (4) would underestimate the percentage that can expect to recruit at least one. Also, A4 may be untrue. Most pyramid schemes do have a product to sell, and despite it being difficult to sell, it is possible to have a source of revenue from products. Thus, the profit earned by a participant is likely to be higher than the model estimates.

6 References

1. BBC News (2015) *Promoter of £21m pyramid scam ordered to pay back £1.* Available at: http://www.bbc.co.uk/news/uk-england-bristol-33536824 (Accessed 23/03/18)
2. Gastwirth, J.L. (1977) *A Probability Model of a Pyramid Scheme*, American Statistician, Vol. 31, No.2, p79-82.
3. Keep, W.W. and Vander Nat, P. J. (2002) *Marketing Fraud; An approach for differentiating multilevel marketing from pyramid schemes*. Journal of Public Policy and Marketing, pp. 139 – 151.
4. Lagarias, J.C. (2013) *Euler's Consant: Euler's Work and Modern Developments*, American Mathematical Society, Vol. 50, No. 4, pp. 527-628
5. Monroe, W. (2017) *Bernoulli and Binomial Random Variables,* Stanford University. Available at https://web.stanford.edu/class/archive/cs/cs109/cs109.1178/lectureHandouts/070-bernoulli-binomial.pdf (Accessed: 08/03/2018)