# Optimal Strategies to Maximise Earnings on Virtual Red Packets

Chen, Tianlai; Parikh, Gaurav Rajesh; Schrage, Zackary

February 2021

## 1 Introduction


In Chinese and other East and Southeast Asian societies, a red envelope (紅包) is a monetary gift which is given during holidays or special occasions such as weddings, graduation, or the birth of a baby. Traditionally, the red envelope symbolises good luck and contains best wishes to receivers. In 2014, the virtual red envelope was first coined and implemented by WeChat, a social media platform. The virtual red envelope soon becomes a hit. During the Spring festival of 2019, over 82 billion virtual red envelopes were sent. The purpose of this report is to investigate the distribution algorithm of the WeChat virtual red envelope. How is money distributed in the virtual red envelope? What is the optimal strategy for obtaining more money? This report begins with methodology and experimental data analysis, followed by simulation and mathematical deduction.

## 2 Methodology

Seven students of STATS 210 were recruited for this project. In order to collect sufficient data of red envelopes, each participant was asked to send 20 red envelopes of 1 RMB for seven people. The whole experiment was run using the WeChat platform. Descriptive data of money were generated and collected for people opening envelopes in order. Mean and variance was calculated for each order of money taker. Further manipulation was conducted in MATLAB, and code is provided and discussed in the following section.

# 3 Experimental Results

Based on the methodology explained above, we now illustrate the results for 140 trials. First, we can show the summary of all trials by plotting the amount of money received by each player (drawn as a separate coloured line) for each trial as shown in figure 1.

We identify a suitable range based on our data by observing the maximum and minimum values obtained for money drawn, and then select a bin size by trial and error such that a detailed enough yet not overly fragmented histogram is obtained.

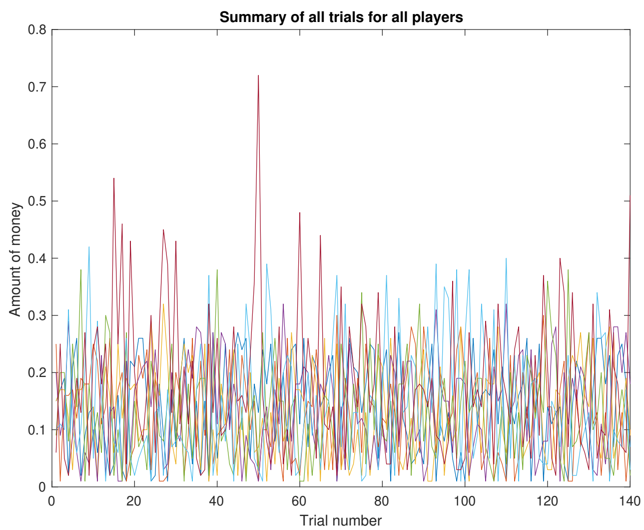



Figure 1: Visualisation of all trials and draws

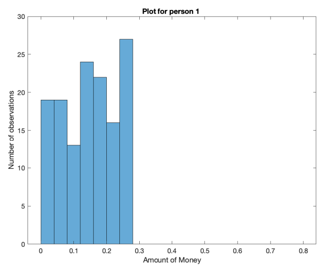
We settle for a range of 0 to 0.8 since the maximum value that we observe for a single money draw is 0.72 RMB while the minimum is 0.01 RMB. We also take the bin size to be 0.04 RMB since this gives us a fairly detailed view of the plots, as it effectively divides the range into a maximum of 20 bars. This is illustrated in Figure 2 on the next page.

Notably it is worth observing that the range of values differs for each person, i.e. the first person is constrained to a lower range of values while for the last person, there is a higher spread of values. While the distributions look random, we can also reason that they may evolve into a uniform distribution if sufficiently many trials are carried out.

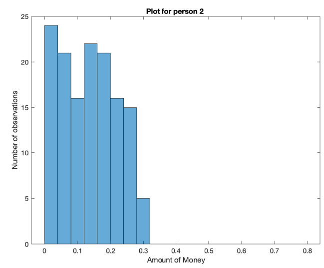
Of course, this is an assumption and we will later explore this in the Discussion part of the report.

We will plot these results, with the help of MATLAB using the code as shown below. Note that we have an input file of raw data which is obtained by manually carrying out the experiment which is a .csv file. We obtain this by recording the data into an Excel spreadsheet and then importing it into the required format. We will submit this along with the code submission.

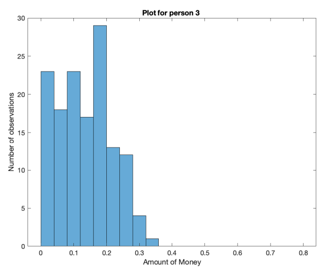
```
function output=results
A = readmatrix('Data.csv');
varmatrix=[];
meanmatrix=[];
for i=2:1:8
   histogram(A(:,i),20,'FaceColor','auto','BinLimits',[0,0.8]);
   xlabel('Amount of Money')
   ylabel('Number of observations')
   title (sprintf('Plot for person %d ',i-1))
   if i<8
   figure()
   end
end
figure()
for i=2:1:8
   varmatrix(i-1)=var(A(:,i));
   meanmatrix(i-1)=mean(A(:,i));
end
plot([1:7],varmatrix);
xlabel('nth Person in order of grabbing')
ylabel('Variance')
title ('Plot of Variance')
figure()
plot([1:7],meanmatrix);
xlabel('nth Person in order of grabbing')
ylabel('mean value for all trials')
title ('Plot of Expected Value')
figure()
for i=2:1:8
   plot(A(:,1),A(:,i))
   xlabel('Trial number')
   ylabel('Amount of money')
   title ('Summary of all trials for all players')
   hold on
end
```



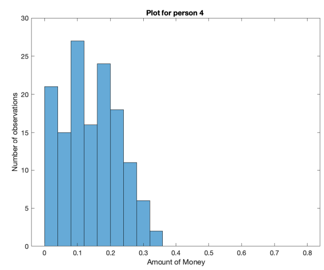

(a) Person 1



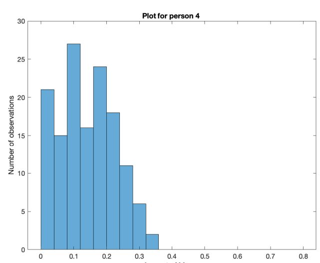

(b) Person 2



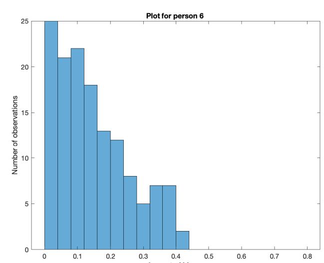

(c) Person 3



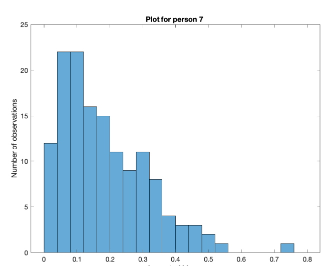

(d) Person 4




(e) Person 5




(f) Person 6




(g) Person 7

Figure 2: Histogram of distributions for money received

# 4 Simulation

## 4.1 Description

Based on the experimental results we now devise an algorithm to distribute money of our own. For this algorithm, we will fulfil two conditions, that at any point we only care about

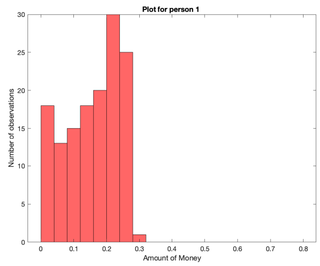
- amount of money that is remaining
- the number of people who are remaining

We do not care about how much money has been given before, and we are not predetermining values for the distribution of money. To do this, we presume (perhaps also for some inherent belief in fairness that all people should be treated equally) that the expected value of all the participants given that they participate infinitely many times in this process should be the same. Based on this notion we define the process of money distribution as a stochastic process with a Mean-preserving spread. Now let there be N random variable $X_i$ such that each $X$ is is uniformly distributed between 0 and the twice the amount of money left divided by the number of people left then for the $n^{th}$ person we will change from the probability distribution $X_i$ to another probability distribution $X_{(i+1)}$, where $X_{(i+1)}$ is formed by spreading out one or more portions of $X_{(i)}$'s probability density function or probability mass function while leaving the mean (the expected value) unchanged. Further, we assume that the stochastic process is modelled by n uniform random variables such that the mean is preserved. Such an approach, as we will show in the Discussion, guarantees that for infinite trials the expected value for all the participants is indeed equal. It also shows us some interesting strategies for whether to be the first person to pick or to pick later. Finally, based on this algorithm we plot our results in Figure 3, the histograms for the distribution of money for all 7 people. We use a range of 0 to 0.8 to keep our plots consistent with the experimental plots. We also take the bin size to be 0.04 RMB for consistency.

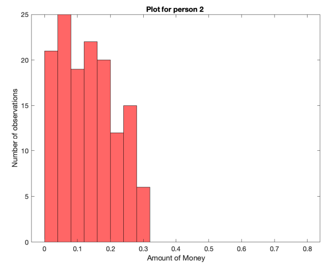
## 4.2 Algorithm and Plots

We now illustrate our code for this section. Our algorithm takes in input parameters for amount of money, number of people and number of trials. For the simulation, we will run this with exactly 1 RMB to be distributed for a total of 140 trials and for 7 people so as to replicate our experimental conditions.

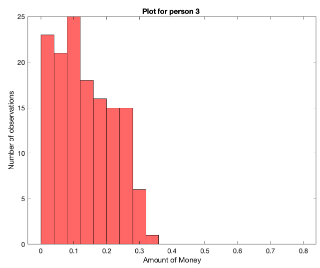
```
function Main(money,people,trials)
r=[];
varmatrix=[];
meanmatrix=[];
moneyleft=money;
for i=1:1:trials
   for j=1:1:people
      if j<people
      r(i,j)=(2*moneyleft/(people+1-j))*rand();
      moneyleft=moneyleft-r(i,j);
      else
         r(i,j)=moneyleft;
      end
   end
   moneyleft=money;
end
for i=1:1:people
   varmatrix(i)=var(r(:,i));
   meanmatrix(i)=mean(r(:,i));
end
plot([1:people],varmatrix);
xlabel('nth Person in order of grabbing')
ylabel('Variance')
title ('Plot of Variance')
figure()
plot([1:people],meanmatrix);
xlabel('nth Person in order of grabbing')
ylabel('mean value for all trials')
title ('Plot of Expected Value')
figure()
for i=1:1:7
   histogram(r(:,i),20,'FaceColor','r','BinLimits',[0,0.8]);
   xlabel('Amount of Money')
   ylabel('Number of observations')
   title (sprintf('Plot for person %d ',i))
   if i<7
   figure()
   end
end
```



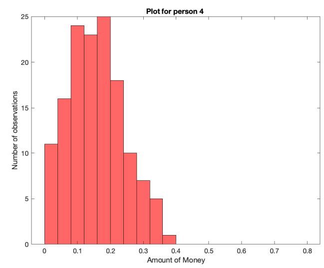

(a) Person 1



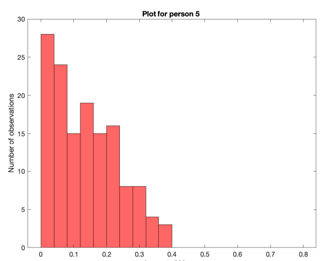

(b) Person 2



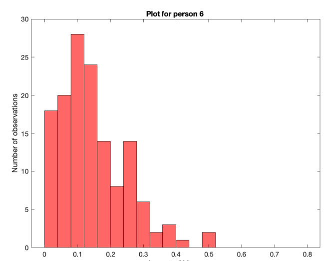

(c) Person 3



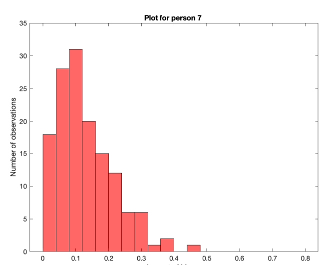

(d) Person 4




(e) Person 5




(f) Person 6




(g) Person 7

Figure 3: Histogram of distributions for money based on simulation algorithm

# 5 Discussion

## 5.1 Expected Value

The expected value, in this case the expected amount of money to be received by an individual opening an red packet with M initial money and N people to open it, is $M/N$. This result is intuitive since this equally distributes money fairly. The proof for this comes from the formula for the expected value.

$$E[X] = \int_{-\infty}^{\infty} xp(x)dx$$

Since we have a uniform distribution from 0 to $\frac{2m_i}{n_i}$ where $m_i$ is the current amount of money remaining at the ith iteration and $n_i$ is the number of people remaining to open the red packet at the ith iteration ($n_i = N - i + 1$), from this we have

$$p(x) = \frac{1}{b-a} = \frac{1}{\frac{2m}{n} - 0} = \frac{n}{2m}$$

$$\int_a^b xp(x)dx = \frac{n}{2m}\int_0^{\frac{2m}{n}} xdx = \frac{n}{2m}\left[\frac{1}{2}x^2\right]_0^{\frac{2m}{n}} = \frac{m}{n}$$

$\frac{m}{n}$ is the expected amount of money to be received at some given iteration, however we also can write the probability distribution iteratively as follows. Our first person has the uniform distribution

$$\begin{cases} \frac{N}{2M} & \text{for } 0 \le x \le \frac{2M}{N} \\ 0 & otherwise \end{cases}$$

where $M$ and $N$ are the fixed total amount of money and total number of people to collect the red packet respectively. Since this is the first iteration, we have that $m = M$ and $n = N$ since none of the money has been collected and the first person is the first person to collect money, and thus we know that the expected value for this person is simply $\frac{M}{N}$ . We can then use this to find how much money the next person is expected to get. The second person now has the uniform distribution

$$\begin{cases} \frac{N}{2M} & \text{for } 0 \le x \le \frac{2(M-\frac{M}{N})}{N-1} = \frac{2(MN-M1)}{N(N-1)} = \frac{2M(N-1)}{N(N-1)} = \frac{2M}{N} \\ 0 & otherwise \end{cases}$$

From this we see that the second person also has an expected value of $\frac{M}{N}$ and by induction we know that the ith person now has the uniform distribution

$$\begin{cases} \frac{N}{2M} & \text{for } 0 \le x \le \frac{2(M-(\frac{M}{N})i)}{N-i} = \frac{2(MN-Mi)}{N(N-i)} = \frac{2M(N-i)}{N(N-i)} = \frac{2M}{N} \\ 0 & otherwise \end{cases}$$

Thus we conclude that the expected value for any given individual, regardless of the ordering that they collect money in, will result in them having the same expected value determined initial by the amount of money in the packet and number of collectors $\frac{M}{N}$.

## 5.2 Trends in Variance

Similar to the calculation for the expected value, we can calculate the variance by

$$\begin{aligned} Var(X) =& E[x^2] - E[x]^2 \\ Var(X) =& \int_a^b x^2 p(x)dx - \left(\int_a^b xp(x)dx\right)^2 \\ Var(X) =& \frac{n}{2m}\int_0^{\frac{2m}{n}} x^2 dx - \left(\frac{m}{n}\right)^2 \\ Var(X) =& \frac{n}{2m}\left[\frac{1}{3}x^3\right]_0^{\frac{2m}{n}} - \frac{m^2}{n^2} \\ Var(X) =& \frac{4m^2}{3n^2} - \frac{m^2}{n^2} \\ Var(X) =& \frac{m^2}{3n^2} \end{aligned}$$

### 5.2.1 Approach 1

We see that this expression for variance is dependent on the current amount of money and people to open the packet remaining. Just as with the expected value though, we can do an inductive process to determine an expression for variance at the ith iteration that depends on $M$ and $N$ instead.

$$\begin{aligned} \frac{(M - \frac{M}{N}i)^2}{3(N-i)^2} &= \frac{M^2 - \frac{2M^2 i}{N} + \frac{M^2 i^2}{N^2}}{3(N-i)^2} = \frac{M^2(1 - \frac{2i}{N} + \frac{i^2}{N^2})}{3(N-i)^2} \\ &= \frac{M^2(N^2 - 2Ni + i^2)}{3N^2(N-i)^2} = \frac{M^2(N-i)^2}{3N^2(N-i)^2} = \frac{M^2}{3N^2} \end{aligned}$$

This implies that the variance is constant, however from the experimental results and the subsequent simulations, we see that this is not true, and that is because while the expected value is indeed constant, the random variable isn't guaranteed to draw the expected value each time, thus we can adjust for this by adding in a perturbance term $\xi$.

$$\begin{aligned}\frac{(M-(\frac{M}{N}+\xi)i)^2}{3(N-i)^2} =& \frac{M^2(N-i)^2+2MN^2\xi i-2MN\xi i^2+N^2\xi^2 i}{3N^2(N-i)^2} \\ =& \frac{M^2(N-i)^2+2MN\xi i(N-i)+N^2\xi^2 i}{3N^2(N-i)^2} \\ =& \frac{M^2}{3N^2}+\frac{2M\xi i}{3N(N-i)}+\frac{\xi^2 i}{3(N-i)^2} \\ =& \frac{M^2}{3N^2}+\frac{2M\xi}{3N}\frac{i}{N-i}+\frac{\xi^2}{3}\frac{i}{(N-i)^2}\end{aligned}$$

This result shows us that the variance is the same expression $\frac{M^2}{3N^2}$ as before, but with additional terms that have a dependency on $i$ multiplied by some constant.

$$\frac{M^2}{3N^2}+C_1\frac{i}{N-i}+C_2\frac{i}{(N-i)^2}$$

Given this equation for variance, we can explain the phenomenon of the variance increasing for the last few individuals that collect the packet. Since $i \in [1, N]$ we have that

$$\lim_{i\to N}\frac{i}{N-i}=\infty$$

$$\lim_{i\to N}\frac{i}{(N-i)^2}=\infty$$

So thus, we see that the variance is approximately $\frac{M^2}{3N^2}$ initially and as the last few people collect their packets the variance increases hyperbolicaly.

To demonstrate this phenomenon, we conducted a simulation of 10 yuan and 100 people over 100 trials at different values of $\xi$ where the range of the probability function went from $\mu-\xi<x<\mu+\xi$ for values of $\xi$ ranging from 0 to $\mu$.

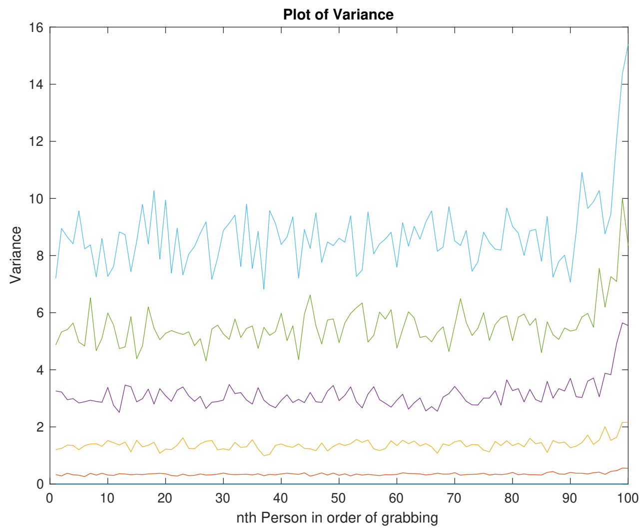



Figure 4: Variance over range of perturbations

What we notice is that in the dark blue case (straddling the x axis) when $\xi = 0$ and every draw of money gives the expected value each time, we get a constant variance that doesn't change. Then as $\xi$ is increased, the variance increases because of the second and third terms in the equation no longer evaluate to 0, and then as we approach the last few people collecting the packet, the hyperbolic blowing up of those terms becomes more and more apparent, such as in the blue case where $\xi = \mu$.

This expression for variance has one glaring problem though, while it can help to predict the end behaviour of the variance function, this model also regards $\xi$ as a constant, which is not accurate.

#### 5.2.2 Approach 2

As aforementioned, the problem with the first approach is that $\xi$ is fixed as a constant, however in actuality it can take many values at different iterations i since the picking of a red packet is a random variable. So instead, we should replace this with a random variable to represent the picking of the sum of the previous i picks. Therefore, we can represent the variance as the following

$$Var(X) = \frac{(M - \Psi_i(x)^*)^2}{3(N-i)^2} \qquad \Psi_i = N\big(\frac{iM}{N}, i\delta\big)$$

This means that our variance $\frac{m^2}{3n^2}$ can be expressed in terms of the initial quantities of money and people. Theoretically, what we expect to see is that everyone draws the mean $\frac{M}{N}$ and so after i people, we should see the sum be $\frac{iM}{N}$. The random variable

$\Psi$ for this, which is a sum of uniform distributions, is a normal distribution by the central limit theorem, with a mean at this value $\frac{iM}{N}$, and as i increases the spread away from the mean increases so its standard deviation is dependent on $i$ and scaled by some constant $\delta$. Since we are interested in the actual x values of this distribution and not plugging the probability into the equation, we now have that $\Psi_i(x)^*$ is the value of some x chosen within a 95% (more or less depending on $\delta$) confidence interval on the normal distribution $N(\frac{iM}{N}, i\delta)$. Lastly, the denominator is unchanged from before.

To demonstrate this fact, we ran a simulation with 1 yuan and 7 people and then plotted $\Psi_i(\mu)$ given the variance of the simulation using a rearranged form of the equation above

$$\Psi_i(\mu) = M - \sqrt{3Var(x)(N-i)^2}$$

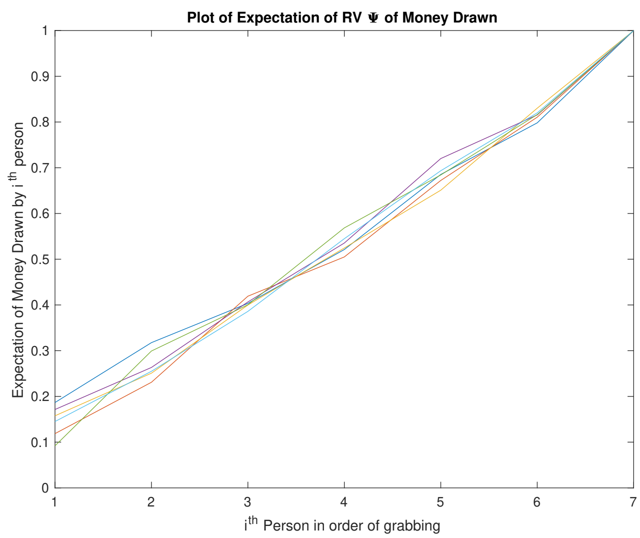



Figure 5: Simulation Plot of Expectation of RV $\Psi$ of Money Drawn

The results for this are correct and corroborate what we would expect to get. The first person leaves with $\frac{M}{N}$ removed from the pool, then the next person takes his cut how having $\frac{2M}{N}$ removed from the pool and so on.

So now that we see that the expression for variance is accurate, how can we use this to find any variance given the parameters $M, N, i$. The way we go about doing this is that we take the bounding values $a$ and $b$ of the normal distribution $\Psi_i$ for some confidence interval that we are satisfied with, and plug those into the equation for variance as the $\Psi_i(x)^*$ term. This will give us an upper and lower bound on our variance.

### 5.3 Additional Findings

One of the things that we observe in our initial data is that while we had assumed each of our distributions to be uniform, in fact the histograms we obtain seem to have a decreasing trend. I.e. for the last person, there appears to be a greater range of values that are obtainable but the highest values have a lower likelihood. This makes sense as if we want to preserve mean and given a higher variance, must see that some high values will come in and in order to average out their effect there must be more lower values.

One interesting observation based on this is that we see a truncated normal distribution for the data for the plot. We suspect this is related to some intrinsic properties of the Mean Preserving Spread stochastic process but a deeper explanation is beyond the scope of this report (and beyond our current abilities). We illustrate the truncated normal plot in Figure 6.

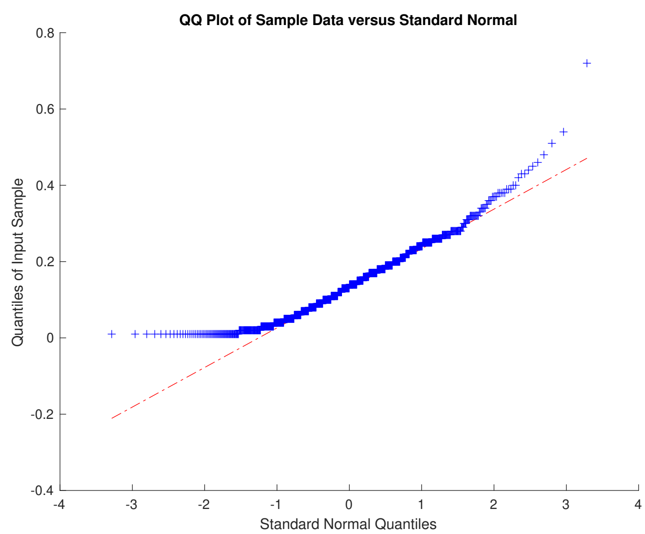



Figure 6: Truncated Normal Distribution

### 5.4 Optimal Strategy

What we have learned about expected value and variance is useful to us in the context of finding an optimal strategy for a participant in this activity. Let us assume that the objective for a person participating in this experiment is to maximise their returns and they know the number of times in total the red packet will be sent ( let us call each time people draw money from a red packet a game ). While so far in this game we have defined it to be such that each participant is sending out some red packets such that the cost to participate in any one game is 0 for all participants and the entire sum of money for one of the participants, an alternate formulation could be that each participant sends out a lump sum amount of money to an external non player who then sends out the red packet( but doesn't participate in the game). Such a formulation would mean that we can now define the cost to each participant to participate on a per game basis, which would make it useful to consider strategic approaches.

#### 5.4.1 Large Number of Games

Now as we observe, if the game is played infinite times, in theory it should not matter whether the person opens the red packet first or last. However, we know that in reality we do not actually play infinite games nor that we prefer to take up higher risk for any given return rather than is necessary. In the language of economics "based on the theory of expected utility maximisation people are risk averse and they have a concave utility function such that if X and Y have the same mean every risk averse individual prefers X to Y, $EU(X) > EU(Y)$ $\forall$ concave U where U is the Utility function".

What we learn from this is that the the optimal strategy for a risk averse player with concave utility is to always be the first player to open the red packet, rather than be the last player.

#### 5.4.2 Few Games

However, if we are playing a finite small number of games, then the strategy to play might be different depending on how risk averse the player is.

For example, if the player is very risk averse they may simply choose to pick the first. However, a more risk tolerant player would choose to play the last since this is when the largest upside and big gains are possible. Now, with this definition, one strategy can be that the last person should participate in the game until gaining one substantial reward that puts their profits to a high level and then stop playing the game entirely.

We can now empirically test this as follows: we will graph sample means. The cost per game for 1 to n trials for all the players is determined simply as being the total amount of money to be distributed divided by the number of people drawing cards. (there are no transaction costs and everyone pays up uniformly then draws the red packet)

Based on our test data if such a strategy would have been employed by the last player this might have almost always been successful if about 28 games were played and then the player might choose to exit.

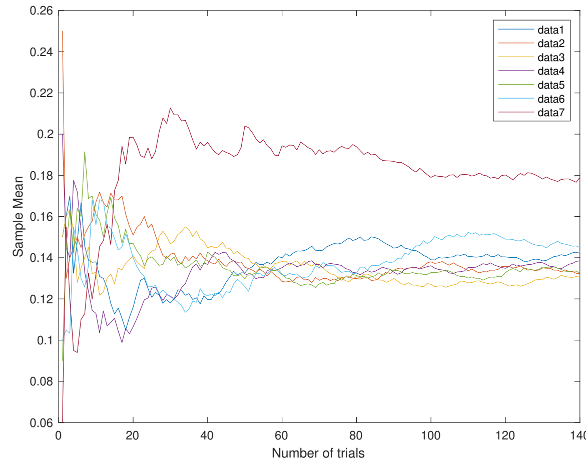



Figure 7: Plot of Average Gains on experimental data over n trials for all players without factoring in cost of play

So, the best way to play is to exit as soon as at the point of having played 20 games the person has made a return greater than their costs, and the most likely approach to achieve this is to be the last player as shown in Figures 7 and 8.

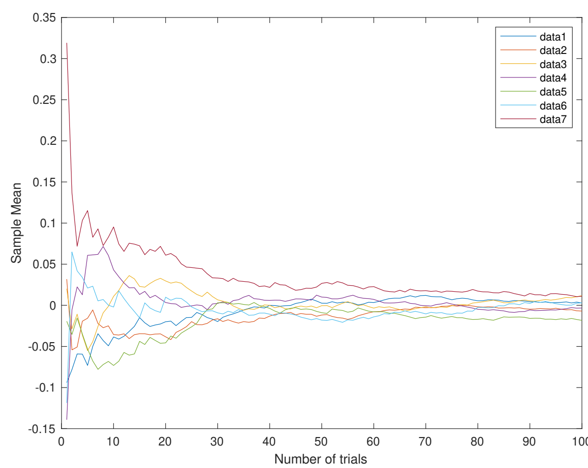



Figure 8: Plot of Average Gains on simulation over n trials for all players with factoring in cost of play

# 6 Conclusion

Through the project, we learned about several ideas in modelling. We have first explored the model for building an algorithm to distribute money based on some intuition and then sought to validate it against some empirical data. Following this we explored a mathematical proof of the mean and the variance through multiple different approaches. We supplemented our findings with statistical mini simulations to validate these results. Finally, we applied this knowledge to model lotteries and games where people are aiming to maximise utility for some given risk tolerance. Overall, our conclusion is that the WeChat algorithm is a rich and interesting statistical simulation that distributes money to people. One unrelated idea that we are all interested in is how does the algorithm works securely without being hacked.