Reverse Engineering Algorithmic Mechanism Behind WeChat Red Envelope

YIMIN TANG

ShanghaiTech University

tangym@shanghaitech.edu.cn

2018 年 1 月 20 日

Overview

- Introduction
- Base Data Visualization
- Code Analysis
- Code Data Visualization
- Conclusion

Introduction

We want to find out the distribution of the Red Envelope. The problem is that it is like a black box and we can only reverse it by Red Envelope generated data. We decide to get the answer starting with the program and get the distribution in the end.
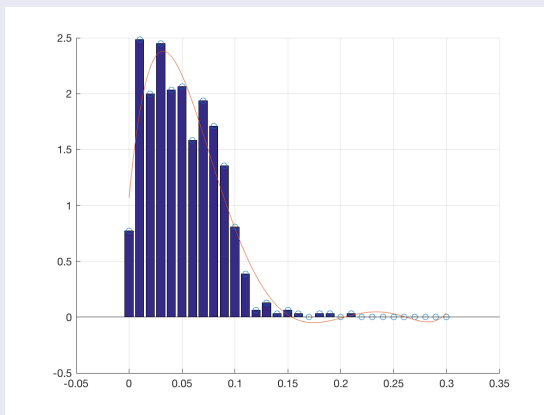
Base Data Visualization

Overview

Every example for Data visualization, x-axis is the parts of a red Envelope order by time (After normalization)
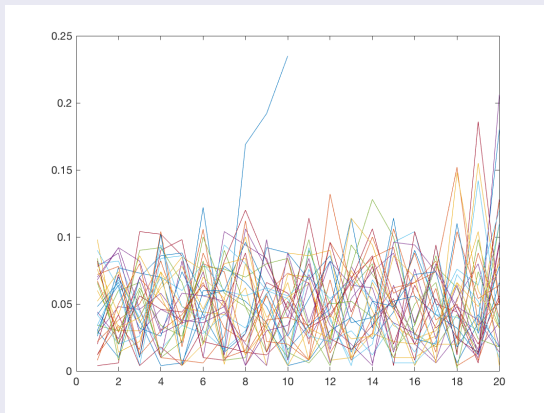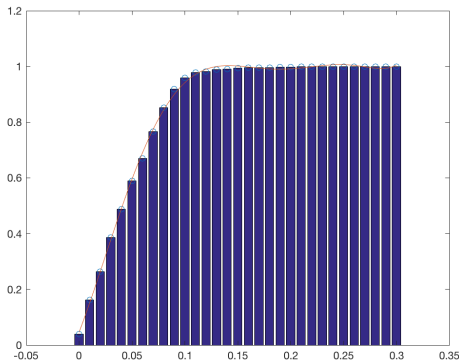
Base Data Visualization

Avg's PDF

PDF of average Data, x-axis is the parts' size, the curve is calculated by linear regression and the max exponent is 7 (After normalization)

Base Data Visualization

Avg's CDF

CDF of average Data, x-axis is the parts' size, the curve is calculated by linear regression and the max exponent is 7(After normalization)
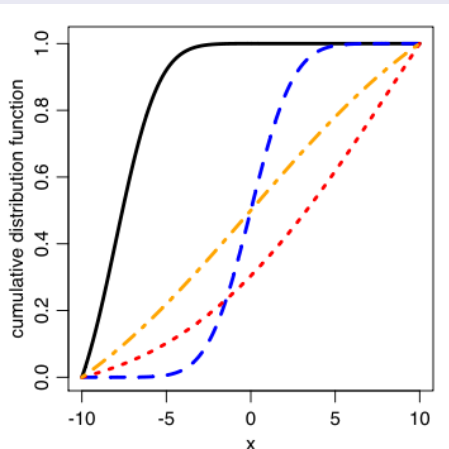
Base Data Visualization

What's the distribution of this?

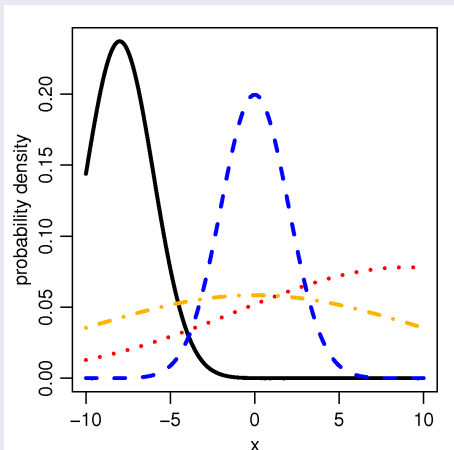Truncated normal distribution!

CDF from wiki

Base Data Visualization

What's the distribution of this?

Truncated normal distribution!

PDF from wiki

Code Analysis

First Group Assumptions

- Simple is good, the part of Red Envelope is not complexity.
- Money involves the part of Red Envelope, so no mistake is the most important.
- The programmer won't show his/her virtuosity in a job that has some relationship with finance.

Conclusion

The program part is easy to reproduce by ourselves.

Code Analysis

Second Group Assumptions

- All money that each people get is not generated one time.
- Generating all money that each people get will cost more store resource.

Conclusion

Each money generation is decided by last time's surplus of money and people.

Code Analysis

Code

https://github.com/TachikakaMin/WechatRedPackets

Code Data Visualization

Average of all Data by Generator 3
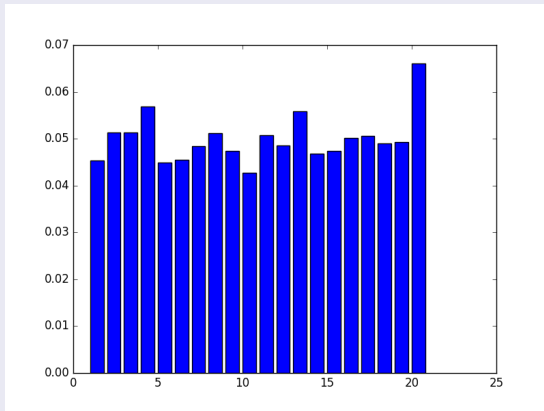

图: Average of all Data by Generator 3

Code Data Visualization
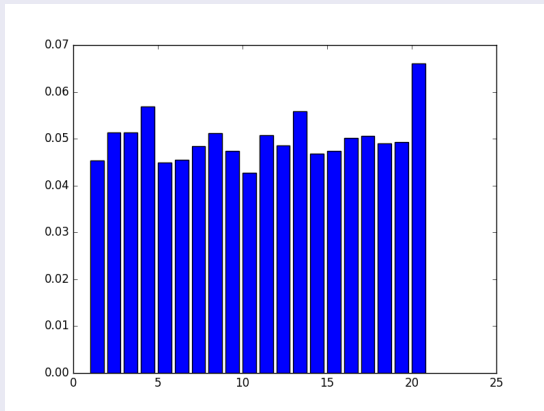
Average of all Data by Generator 3

图: Average of all Data by Generator 3

Code Data Visualization

Average CDF of all Data by Generator 3
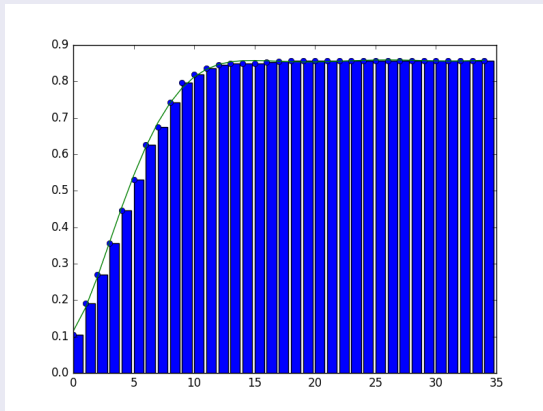

图: CDF by Generator 3

Code Data Visualization
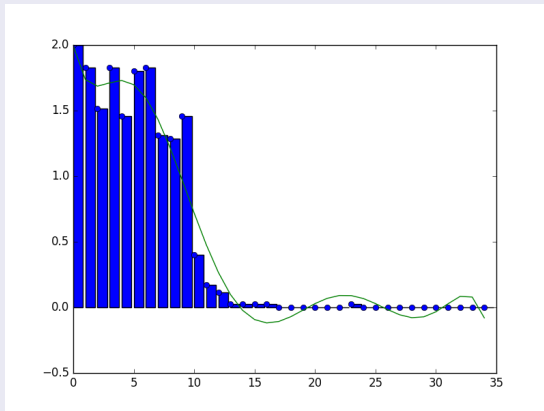
Average PDF of all Data by Generator 3

图: PDF by Generator 3

Code Data Visualization

Give a Red Envelope with two parameters a, n, a is the total money and n is the total number of people. Let $X_1, X_2, \ldots, X_n$ is the money that n people get.

Expectation in theory

$E(X_1) = \frac{a}{n}$

$E(X_2) = E(\frac{a - E(X_1)}{n-1}) = \frac{a}{n}$

Expectation in theory

If $X_1, X_2, \ldots, X_{n-1}$ are all satisfied $E(X_i) = \frac{a}{n}$

$E(X_n) = E(\frac{a - E(X_1 + X_2 + \ldots + X_{n-1})}{n - (n-1)})$

$= a - E(X_1 + X_2 + \ldots + X_{n-1})$

$= \frac{a}{n}$

The code mat be very simple and the generate function is that its expectation is the expectation of last people and money. This simple idea create the result of Truncated normal distribution.