Simulation Of Binomial Distribution By Monte Carlo Method

Xueshuang Fan[1,a,*], Yunjuan[1,b] Li and Fengjun Yan[1,c]

[1]Xi 'an siyuan university Xi 'an, Shanxi, China

[a]fanxueshuang11@163.com, [b]liyunjuan22@163com,[c]yanfj1979@163.com

*corresponding author

Keywords: Monte Carlo Method; Binomial Distribution; Mathematical Expectation

Abstract: In this paper, the distribution law of binomial distribution simulated by Monte Carlo method is presented, and then the point estimation and interval estimation methods and steps of binomial distribution mathematical expectation and variance are given by using the sample digital characteristic method, and the results are compared with the real values.Meanwhile, the corresponding Matlab program is given.

1. The Introduction

Monte Carlo method, also known as statistical simulation method, was first proposed in the 1940s by S.M. Ulam and J. von Neumann.With the development of electronic computer technology, Monte Carlo method has been widely used in financial engineering, computational physics, artificial intelligence and other fields [1-4].

The binomial distribution is an important probability distribution. Binomial distribution is based on Bernoulli experiment, which is instructive for many practical problems. Next, we give the distribution law, expectation and variance of binomial distribution simulated by Monte Carlo method, and realize it by mathematical software Matlab, so as to further reveal the statistical meaning of probability definition and the mean meaning of expectation.

2. Binomial Distribution Distribution Law, Simulation Of Expectation And Variance

In order to quantify the sample space of random trials, the important concept of random variables is introduced in the course of probability theory. Random variables can be classified into discrete, continuous and mixed types. There is an important probability distribution of discrete random variables, namely binomial distribution. So let me give you the concept of a binomial distribution.

Defination1 Let the possible values of the random variable X be 0,1,2..., n, its distribution law is $P\{X=k\}=C_n^k p^k(1-p)^{n-k}, k=0,1,2,\ldots,n\ (0<p<1)$.

Then, the binomial distribution X of the obedient parameters n,p is denoted as $X\sim B(n,p)$ [5].

In the process of solving the distribution law of binomial distribution, when the parameter n is relatively large, it is obviously troublesome to directly calculate $C_n^k p^k(1-p)^{n-k}$. When n is large and p is very small, we generally use $\frac{\lambda^k e^{-\lambda}}{k!}$ ($\lambda=np$) as its approximate calculation formula.Next, we get the distribution law of binomial distribution by doing random experiment and using statistical simulation method to get the approximate value of $C_n^k p^k(1-p)^{n-k}$.

The steps of monte Carlo method to simulate the distribution law of binomial distribution:

Step 1: Given M, the n Bernoulli test is performed.

Step 2: Generate random Numbers on the interval. When the generated random Numbers belong to the interval $[0,p]$, it means that the event occurs; when the generated random Numbers belong to the interval $(p,1]$, it means that the event does not occur.

DOI: 10.38007/Proceedings.0001512

ISBN: 978-1-80052-010-3

Step 3: Given the sequence $\{a_i \mid i = 1,2,\ldots M\}$, a_k represents the number of events that occurred in the n Bernoulli test.

Step 4: Given the sequence $\{f_k \mid k = 0,1,2,\ldots,n\}$, f_k represents the frequency of occurrence of k in the sequence $\{a_i \mid i = 1,2,\ldots M\}$.

Step 5: Determine the distribution law $\{p_k \mid k = 0,1,2,\ldots,n\}$, where $p_k = \frac{f_k}{M},\ k = 0,1,2,\ldots,n$.

Use the following example to illustrate the above steps.

Example 1 The probability of congestion in a certain section in a day is 0.2, which is used to represent the number of days of congestion in this section in a week, and the distribution law simulated by monte Carlo method is used.

Matlab program is as follows:

```
syms M n P x p;
M = 10000; N = 7; P = 0.2; % The number of tests and parameters of binomial distribution are given.
for i = 1: M
      a (i) = 0;
         for j = 1: n
                x= unifrnd (0, 1);          % generates random Numbers.
                if x < p
                      a (i) = a (i) + 1;        % Total number of events in n - weight experiments.
           end
end
  end
for k = 1: (n + 1)
      b (k) = 0.
      for i = 1: M
        If    a (i) = = (k - 1)
         b = b (k) (k) + 1;        % Statistical frequency
    end
        end
  end
P= b /M      % calculated frequency.
```

Operation result: P = 0.2076   0.3634   0.2796   0.1155   0.0293   0.0043   0.0003   0

The following is the solution for the exact value of the binomial distribution.

```
Syms n p q P;
n = 7; p = 0.2; q = 1 - p;
For i= 1: (n + 1)
  P (i) =    (nchoosek (n, i - 1) * (P ^ (i - 1))) * q ^ (n -i + 1);
%      Calculate the probability of the occurrence of the event.
  end
P;    vpa (P)
```

The operation results are 0.2097152, 0.3670016, 0.2752512, 0.114688, 0.028672, 0.0043008, 0.0003584, 0.0000128.

Use the following table to compare the simulated and real values.

Table 1. Comparison Of Simulated Values And Real Values Of Binomial Distribution Distribution Law

X	0	1	2	3	4	5	6	7
P_i Simulation value	0.2076	0.3634	0.2796	0.1155	0.0293	0.0043	0.0003	0

X	0	1	2	3	4	5	6	7
P_i The real value	0.2097152	0.3670016	0.2752512	0.114688	0.028672	0.0043008	0.0003584	0.0000128
Absolute error	0.00211152	0.0036016	0.0043488	0.000812	0.00063	0.0000008	0.0000584	0.0000128
The relative error	0.0101	0.0098	0.0158	0.0054	0.0219	0.0002	0.1629	1

Mathematical expectation and variance are two important characteristics of random variables. Mathematical expectations reflect the average value of a random variable. And the variance reflects the deviation of the value of the random variable.

Definition 2 Let the distribution law of discrete random variable X be $P\{X = x_i\} = p_i, i = 1,2,3\ldots$, $\sum_{i=1}^{+\infty} x_i p_i$ called the mathematical expectation of random variable X, denoted by $E(X)$, and $\sum_{i=1}^{+\infty} (x_i - E(X))^2 p_i$ is denoted by variance, denoted by $D(X)$ [6].

There are two main methods to estimate the mathematical characteristics of random variables: point estimation and interval estimation. To estimate the value of population parameter with samples, namely point estimation, there are mainly sample numerical feature method and moment estimation method.The main principle of the sample numerical characteristic method is to use the sample mean $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ as the estimate of the mathematical expectation of $E(X)$ and the sample variance $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$ as the estimate of the population variance $D(X)$.

Next, the mathematical expectation and variance of the random variable X in example 1 are given by using the sample numerical feature method.

syms M n P x; M = 10000; N = 7; P = 0.2; %The number of tests and parameters of binomial distribution are given.

```
for i= 1: M
        a (i) = 0;
        for j = 1: n
            x = unifrnd (0, 1);
          if x < p
                a(i) = a (i) + 1;      % Total number of events in n - weight experiments.
          end
        end
  end
b = 0;
For i = 1: M
      b = b + a (i);
  end
EX= b/M     % calculate the mathematical expectation of a binomial distribution.
c= 0;
fFor i = 1: M
      c = c + (a (i) -   EX) ^ 2;
end
DX=c/(M-1)         %   calculate the variance of the binomial distribution.
```

EX = 1.3910, DX =1.1224.

The following is the matlab program to solve the binomial distribution mathematical expectation and variance exact value.

```
syms n p;n=7;p=0.2;[EX, DX]=binostat (n, p)
```

Operation solution EX =1.4000, DX =1.1200.

Using Monte Carlo method, the estimated values of $E(X)$ and $D(X)$ are 1.3910 and 1.1224 respectively. Compared with the real values of 1.4 and 1.12, the relative errors are 0.0064 and 0.0021 respectively. So the simulation results are relatively ideal.

The point estimate does not give the degree of confidence of the estimate. In order to compensate for the shortcoming of point estimation, another important parameter estimation method, interval estimation, is given.

Definition3 Let $\hat{\theta}_1(X_1, X_2, \ldots, X_n)$ and $\hat{\theta}_2(X_1, X_2, \ldots, X_n)$ be two statistics. If, for a given probability $1-\alpha$ $(0<\alpha<1)$, there is $P\{\hat{\theta}_1 < \theta < \hat{\theta}_2\} = 1-\alpha$, then the random interval $(\hat{\theta}_1, \hat{\theta}_2)$ is the confidence interval of the parameter θ ,and. $\hat{\theta}_1$ is the lower confidence limit. $\hat{\theta}_2$ is the upper confidence limit and $1-\alpha$ is confidence degree [7].

According to the central limit theorem, when n sufficiently large, approximately has $\sum_{i=1}^{n} X_i \sim N(n\mu, n\sigma^2)$, where $E(X_k) = \mu, D(X_k) = \sigma^2$ .Therefore, the confidence intervals of expectation $E(X)$ and variance confidence level are

$$(\overline{X} - \frac{S}{\sqrt{n}} t_{\alpha/2}(n-1), \overline{X} + \frac{S}{\sqrt{n}} t_{\alpha/2}(n-1)), ((n-1)S^2 / \chi^2_{\alpha/2}(n-1), (n-1)S^2 / \chi^2_{1-\alpha/2}(n-1)) \tag{1}$$

The following is the matlab program for interval estimation of binomial distribution expectation and variance, where the confidence level is set as 0.95.

```
syms M n P x XB S L1 R1 L2 R2;
M= 10000; n=7; p= 0.2;alf = 0.05; %The test number and binomial distribution parameters and confidence level are given.
For i=1: M
     a(i) = 0;
     for j=1: n
          x= unifrnd(0, 1);
          if x < p
               a(i)=a(i)+1; % number of events occurring in n - weight tests.
         end
     end
 end
b = 0;
for i= 1: M
     b=b+a (i); % Total number of times the event occurred
 end
XB=b/M; %sample mean.
c=0;
for i = 1: M
    c=c +(a(i)- XB) ^ 2;
end
S=sqrt(c/(M-1));S2=c/(M-1);    % sample variance.
L1=XB+(S*icdf('t',alf/2,M-1))/sqrt(M);
R1=XB-(S*icdf('t',alf/2,M-1))/sqrt(M);    % Calculate confidence intervals for mathematical expectations.
L1
```

R1
L2=(M-1)*S2/icdf('chi2',alf/2,M-1);
R2=(M-1)*S2/icdf('chi2',1-alf/2,M-1);   % calculates confidence interval for variance.
L2
R2

Operation results: L1 = 1.3702, R1 =1.4118, L2 =1.1639, R2=1.1011.

Interval estimates of mathematical expectation and variance with confidence level of 0.95 are (1.3702, 1.4118) and (1.1011, 1.1639).

Conclusion

As a numerical method based on probability and statistics theory, Monte Carlo method is simple and feasible. With the help of mathematical software MATLAB, it is easy to get the simulated values of binomial distribution law, mathematical expectation and variance. Similar methods and steps can be used for other types of distributions [8].

Since 2016, the Alpha Go artificial intelligence robot developed by Google DeepMind company has attracted wide attention [9-10]. Its main working principle is deep learning, the core technology is monte Carlo method. It can be predicted that monte Carlo method will change people's cognition and life with its wide application prospect.

Acknowledgements

Thanks for the support of the school-level project probability theory and Mathematical Statistics of xi 'an Siyuan University in 2020.

References

[1] Yan Chun, Solution bingxin, Log-normal model with Outliers in Mind and Monte Carlo Method implementation of predictive distribution, Statistics and Information Forum, 09, 2018.

[2] Chun-yan Fan, Shang-Guo Liu, Sh-zhai Yang, analysis of monte Carlo method for evaluating uncertainty of total station measurement, geospatial information, 04, 2020.

[3] Xiong Danfeng, LIU Chunhong, Wu Haibin, application of Monte Carlo method in colorimetric temperature measurement under high temperature background radiation, Journal of Atmospheric and Environmental Optics, 03, 2019.

[4] Sun Xiaoxu, Application of Monte Carlo method in the prediction of gold reserves in the Taiping Area of Inner Mongolia, World Nonferrous Metals, Issue 03, 2018.

[5] Zhang Lishi, probability theory and Mathematical Statistics, Tsinghua University Press, 2015.

[6] Li Xieling, Tan Weiming, Probability theory and Mathematical Statistics, Aviation Industry Press, 2011.

[7] Hu Min, Probability theory and Mathematical Statistics, Shanghai Jiao Tong University Press, 2012.

[8] Si Shoukui, Sun Zhaoliang, Mathematical Modeling Algorithm and Application, National Defense Industry Press, 2017.

[9] Pan Feng, AlphaGo, China Electronic News, November 21, 2017.

[10] Zheng Huanbin, Voice from inside Deep Thinking, Science and Technology Daily, April 1, 2016.