随机模拟方法与应用导论作业一

陈稼霖 *45875852*

2019-09-23

1.5 (Binomial CDF)

Let X be the number of "ones" obtained in 12 rolls of a fair die. Then X has a Binomial$(n = 12, p = 1/6)$ distribution. Compute a table of cumulative binomial probabilities (the CDF) for $x = 0, 1, \ldots, 12$ by two methods: (1) using **cumsum** and the result of Exercise 1.4, and (2) using the **pbinom** function. What is $P(X > 7)$?

(1) 先计算二项式分布的PMF

$$P(X = k) = \left(\begin{array}{c} n \\ k \end{array} \right) p^k (1-p)^{n-k} = \left(\begin{array}{c} 12 \\ k \end{array} \right) \left(\frac{1}{3}\right)^k \left(1 - \frac{1}{3}\right)$$

```
options(digits = 8)
n = 12
p = 1/3
k = c(0:n)
P = choose(n,k) * p^k * (1 - p)^(n - k)
P
```

```
##  [1] 7.7073466e-03 4.6244080e-02 1.2717122e-01 2.1195203e-01 2.3844604e-01
##  [6] 1.9075683e-01 1.1127482e-01 4.7689207e-02 1.4902877e-02 3.3117505e-03
## [11] 4.9676258e-04 4.5160234e-05 1.8816764e-06
```

然后利用函数**cumsum**计算其CDF

```
F = cumsum(P)
F_k = cbind(k,F)
colnames(F_k) = c('k','F(X=k)')
F_k
```

```
##         k       F(X=k)
##  [1,]  0 0.0077073466
##  [2,]  1 0.0539514264
##  [3,]  2 0.1811226458
##  [4,]  3 0.3930746781
##  [5,]  4 0.6315207144
##  [6,]  5 0.8222775435
```

```
##  [7,]  6 0.9335523605
##  [8,]  7 0.9812415677
##  [9,]  8 0.9961444450
## [10,]  9 0.9994561955
## [11,] 10 0.9999529581
## [12,] 11 0.9999981183
## [13,] 12 1.0000000000
```

(2) 直接用函数**pbinom**函数计算二项式分布的CDF

```
F = pbinom(k,n,p)
F_k = cbind(k,F)
colnames(F_k) = c('k','F(X=k)')
F_k
```

```
##        k       F(X=k)
##  [1,]  0 0.0077073466
##  [2,]  1 0.0539514264
##  [3,]  2 0.1811226458
##  [4,]  3 0.3930746781
##  [5,]  4 0.6315207144
##  [6,]  5 0.8222775435
##  [7,]  6 0.9335523605
##  [8,]  7 0.9812415677
##  [9,]  8 0.9961444450
## [10,]  9 0.9994561955
## [11,] 10 0.9999529581
## [12,] 11 0.9999981183
## [13,] 12 1.0000000000
```

计算$P(x > 7)$

```
F = 1-pbinom(7,n,p)
1-pbinom(7,n,p)
```

```
## [1] 0.018758432
```

1.13 (lunatics data)

Obtain a five-number summary for the numeric variables in the lunatics data set (see Example 1.12). From the summary we can get an idea about the skewness of variables by comparing the median and the mean population. Which of the distributions are skewed, and in which direction?

先生成所需数据集

```r
x = matrix(rnorm(20),10,2)
x
```

```
##              [,1]         [,2]
##  [1,] -0.65822854 -2.272548683
##  [2,]  0.83044396 -0.549542906
##  [3,] -1.16299565  1.535298576
##  [4,] -0.37376462 -1.203134547
##  [5,] -1.24063003  0.056282405
##  [6,] -0.18121616 -0.860101256
##  [7,]  2.76138371  0.417178376
##  [8,]  2.01819901 -0.092126212
##  [9,] -0.73820170 -0.606404806
## [10,] -0.91010968 -1.574098261
```

生成数据集的summary

```r
summary(x)
```

```
##        V1                  V2
##  Min.   :-1.240630   Min.   :-2.27255
##  1st Qu.:-0.867133   1st Qu.:-1.11738
##  Median :-0.515997   Median :-0.57797
##  Mean   : 0.034488   Mean   :-0.51492
##  3rd Qu.: 0.577529   3rd Qu.: 0.01918
##  Max.   : 2.761384   Max.   : 1.53530
```

从中可见，V1的平均值大于其中位值，因此V1正偏(positive skewed)，V2的平均值大于其中位值，因此V2也正偏。