

二项分布的参数模拟

庄亮亮

Generating random Numbers

```
set.seed(0)
binom = function(n,p,a=1.96){
  #n为随机产生的个数，p为预先生成随机数的参数，
  result = list()
  x = rbinom(n,1,p)
  mean = mean(x)
  result$mean = mean
  var = var(x)*n/(n-1)
  result$var = var
  up = mean + a*sqrt(var/n)
  low = mean - a*sqrt(var/n)
  result$conf.int =c(low,up)
  return(result)
}
```

Simulation

- 当 $p=0.5$ ，计算根据不同试验次数所对应的 p 的估计值

试验次数	p 估计值	偏差	区间估计
10	0.7	0.2	0.38441, 1.01559
50	0.48	0.02	0.3386918, 0.6213082
100	0.48	0.02	0.3810893, 0.5789107

试验次数	p 估计值	偏差	区间估计
500	0.478	0.022	0.4341278, 0.5218722
1000	0.477	0.023	0.4460115, 0.5079885
5000	0.501	0.001	0.487138, 0.514862
10000	0.4999	10×10^{-5}	0.490099, 0.509701

- 当 **p=0.25**，计算根据不同试验次数所对应的 p 的估计值

试验次数	p 估计值	偏差	区间估计
10	0.1	0.15	-0.1066021, 0.3066021
50	0.16	0.09	0.0563081, 0.2636919
100	0.24	0.01	0.1554462, 0.3245538
500	0.218	0.032	0.1817363, 0.2542637
1000	0.242	0.008	0.2154275, 0.2685725
5000	0.2542	0.0042	0.2421286, 0.2662714
10000	0.2507	7×10^{-4}	0.2422042, 0.2591958

- 当 **p=0.75**，计算根据不同试验次数所对应的 p 的估计值

试验次数	p 估计值	偏差	区间估计
10	0.8	0.05	0.5245305, 1.0754695
50	0.68	0.07	0.5480606, 0.8119394
100	0.72	0.03	0.6311073, 0.8088927
500	0.732	0.018	0.6930988, 0.7709012
1000	0.745	0.005	0.717958, 0.772042
5000	0.7616	0.0116	0.7497866, 0.7734134
10000	0.7486	0.0014	0.7400963, 0.7571037

Conclusion

可以看到不管 p 为何值，随着 n 不断变大， p 估计值与实际参数之间的偏差在不断减小，区间估计效果越来越好。