计算物理 作业报告12

PB14203209　张静宁　　2017.11.1 4

第十二题

自设若干个随机分布（相同或不同发布，它们有相同或不同的 μ 和 σ^2 ），通过Monte Carlo模拟，验证中心极限定理成立（$N=2$、5、10）.

算法公式

中心极限定理

设 X_1，$X_2,\ldots X_n,\ldots$ 为独立同分布的随机变量，$E(X_i)=\mu$，$Var(X_i)=\sigma^2$ $(0<\sigma^2<\infty)$.

则对任何实数 x，有

$\lim\limits_{n\to\infty} P(\frac{1}{\sqrt{n}\sigma}(X_1+\ldots+X_n-n\mu)\leq x)=\Phi(x)$

这里，$\Phi(x)$ 是标准正态分布 $N(0,1)$ 的分布函数，即 $\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-t^2/2}dt$.

注意到 $X_1+\ldots+X_n$ 有均值 $n\mu$，方差 $n\sigma^2$，故

$y=\frac{1}{\sqrt{n}\sigma}(X_1+\ldots+X_n-n\mu)=\frac{<X>-\mu}{\sigma/\sqrt{N}}$，其中$<X>=(X_1+\ldots+X_n)/n$.

就是 $X_1+\ldots+X_n$ 的标准化，以下称 $\frac{<X>-\mu}{\sigma/\sqrt{N}}$ 标准化和值 (normalized sum) .

Monte Carlo 模拟验证中心极限定理

其思想是：生成独立同分布的随机变量 $X_1,X_2,\ldots,X_N$，即 X_i 是满足特定分布的随机抽样. 算出每次抽样的标准化和值，进行 M 次抽样（如$M=1000,100000$ ），即可得到标准化和值的频数分布. 随着 N 增大，归一化以后的频率分布越来越接近标准正态分布，就简要的验证了中心极限定理.

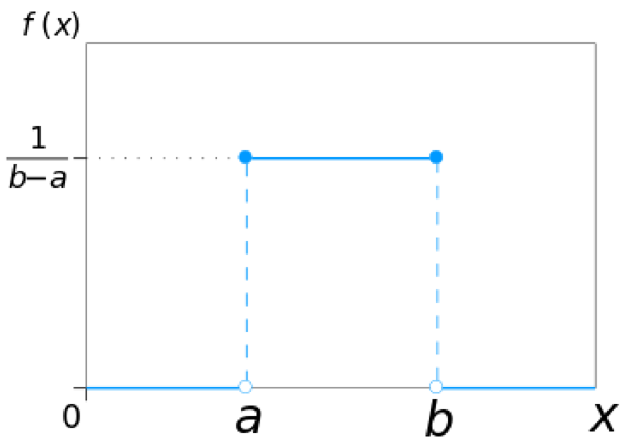
程序说明

- uniform.c　　连续均匀分布，输出标准化和值
- bernoulli.c　　伯努利分布（0-1分布），输出标准化和值

- arbitrary.c 给定概率密度函数，输出标准化和值

计算结果与分析

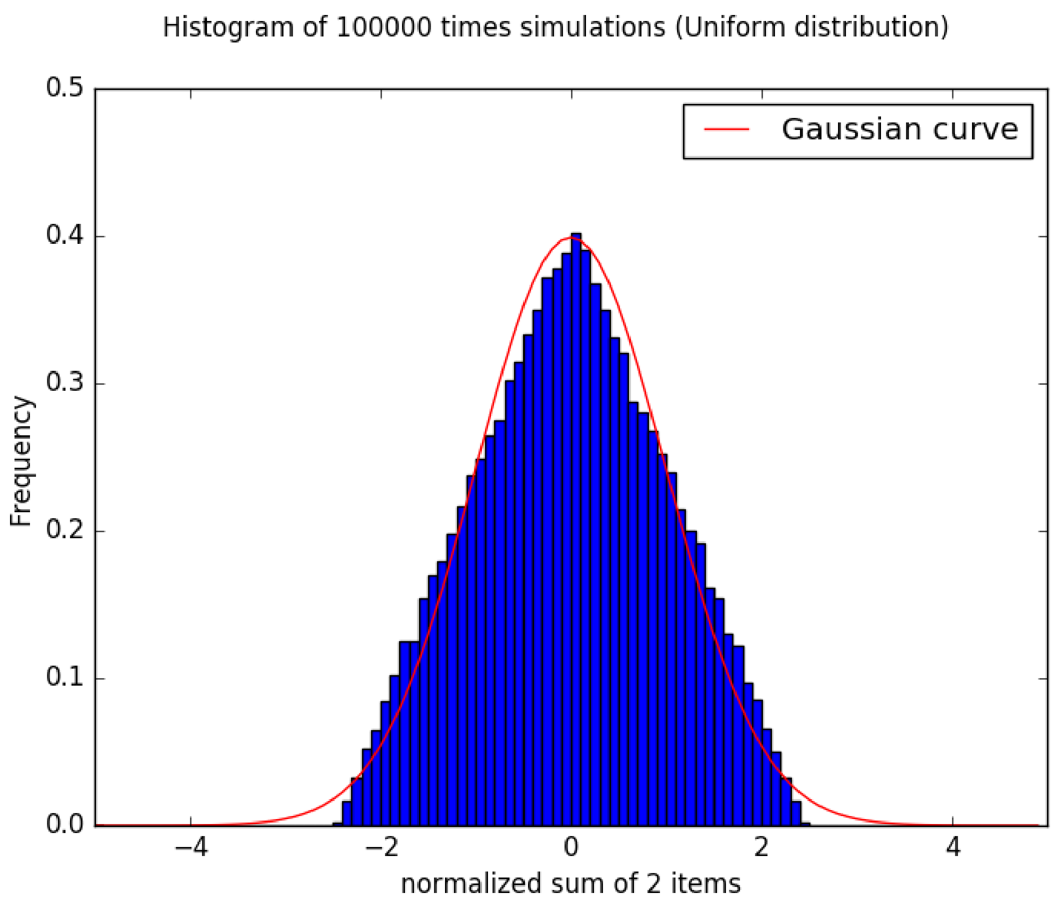
连续均匀分布



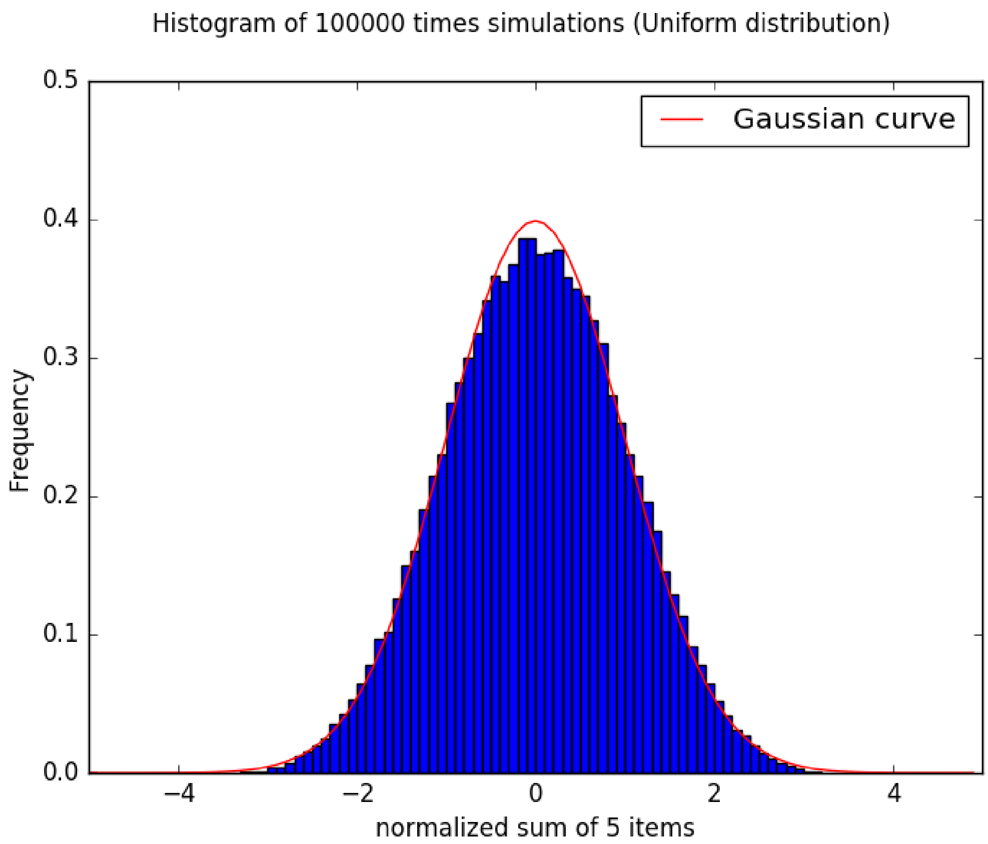
考虑 $[0, 1]$ 的连续均匀分布

均值 $\mu = \frac{1}{2}(a+b) = 0.5$，方差 $\sigma^2 = \frac{1}{12}(b-a)^2 = 1/12$，标准差 $\sigma = 1/\sqrt{12}$

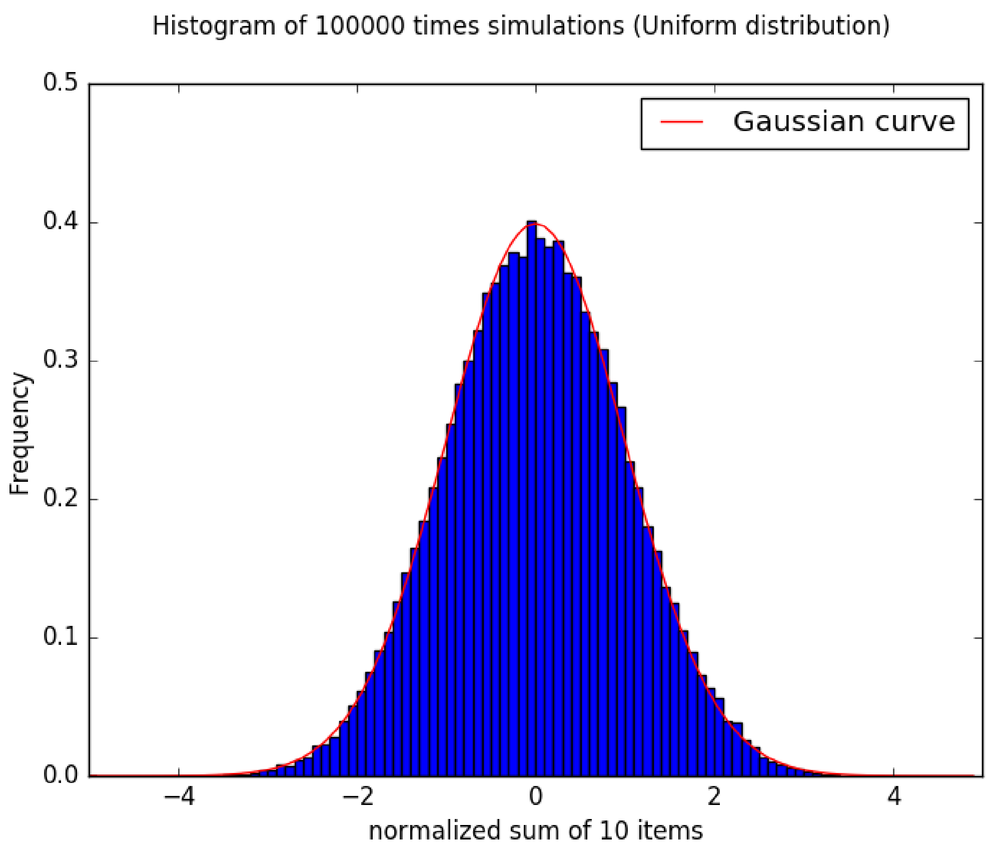
1. $N = 2$



2. $N = 5$


Histogram of 100000 times simulations (Uniform distribution)
Gaussian curve
0.5
0.4
0.3
0.2
0.1
0.0
Frequency
−4
−2
0
2
4
normalized sum of 5 items

3. $N = 10$


Histogram of 100000 times simulations (Uniform distribution)
Gaussian curve
0.5
0.4
0.3
0.2
0.1
0.0
Frequency
−4
−2
0
2
4
normalized sum of 10 items

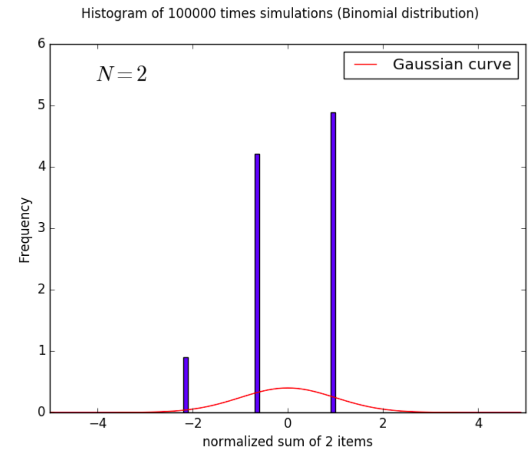
伯努利分布 （0-1分布）

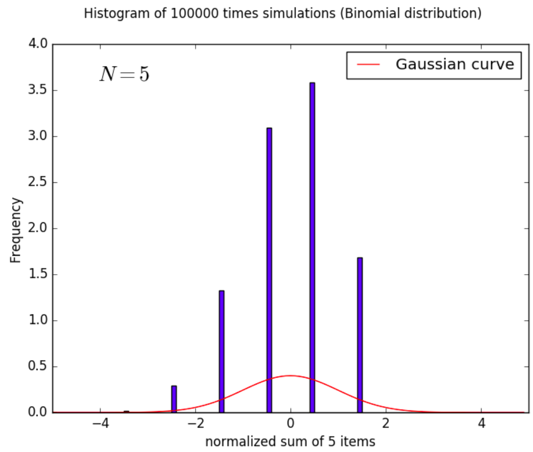
设 $X_1, X_2, \ldots, X_n, \ldots$ 独立同分布，X_i 的分布是伯努利分布，即

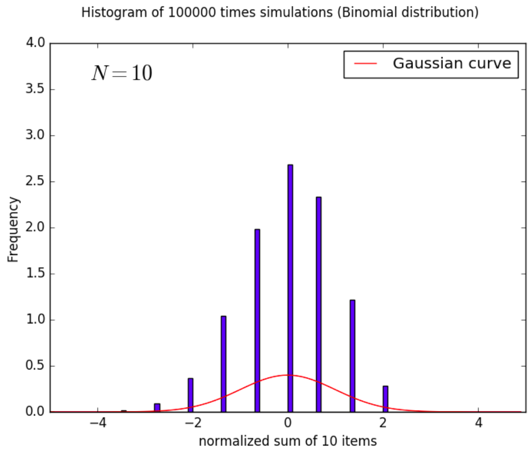
$P(X_i = 1) = p = 0.7$，$P(X_i = 0) = 1 - p = 0.3$，$(0 < p < 1)$

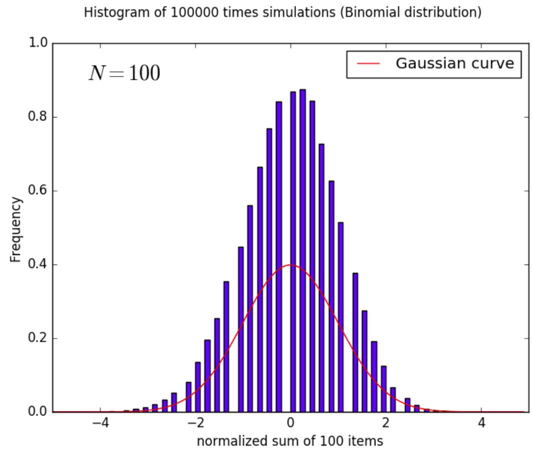
该分布均值 $\mu = p = 0.7$，方差 $\sigma^2 = p(1-p) = 0.21$

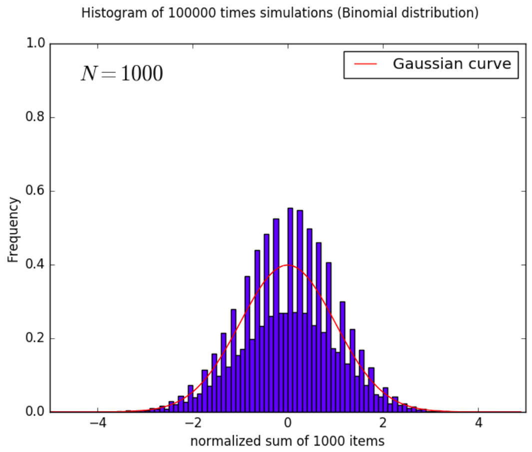
- 以下是 $N = 2$，5，10，10^2，10^3，10^5 的情形

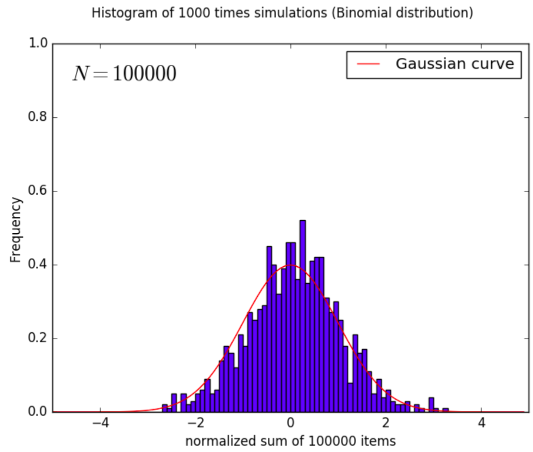










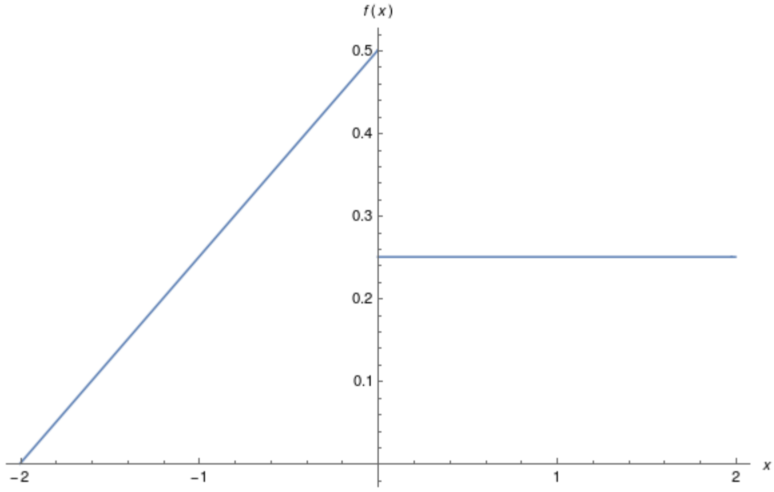


可见，只有当 N 足够大时，伯努利分布 $X_1 + X_2 + \ldots + X_n$ 的标准化才会趋向于正态分布. 也就是 n 很大时，可以用标准正态分布去逼近二项分布$B(n, p)$.

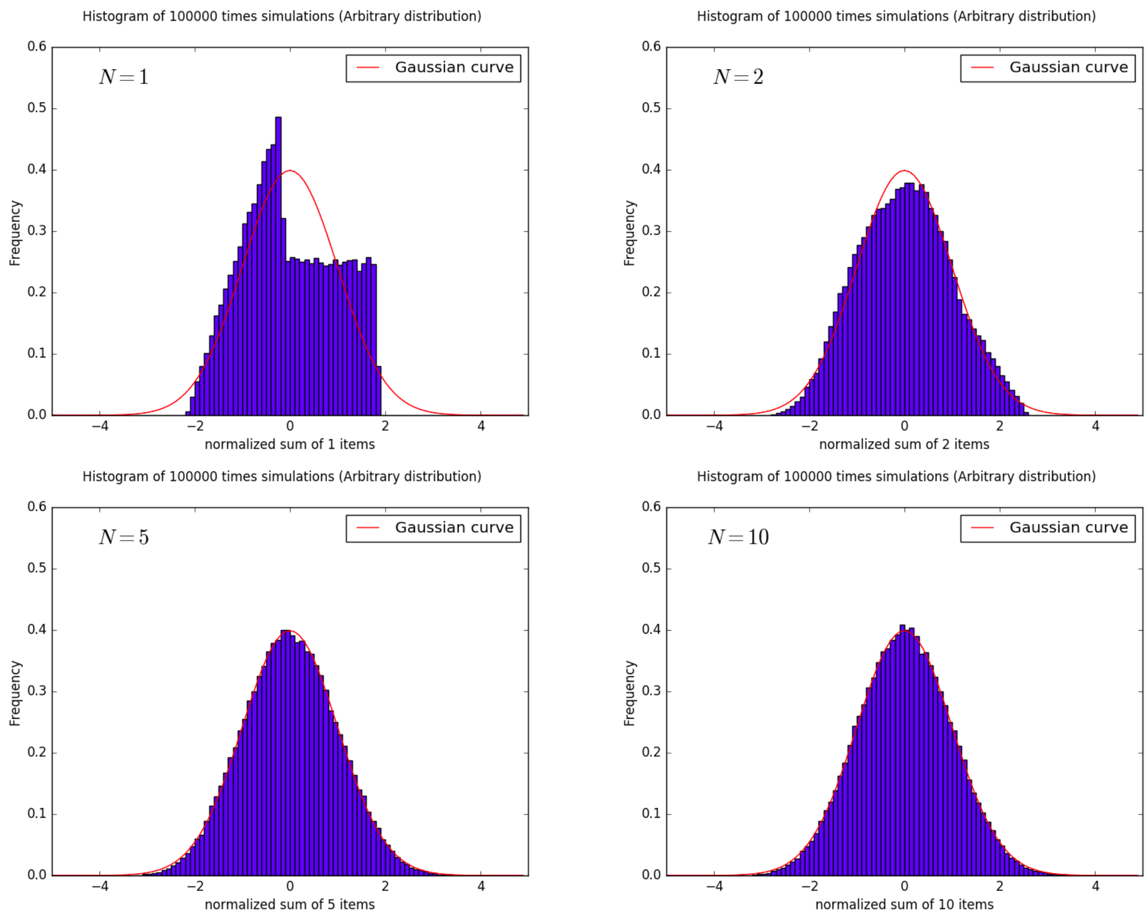
给定分布

给定分布，其概率密度函数如下，均值 $\mu=\frac{1}{6}$，方差 $\sigma^2=1$

$$f(x)=\begin{cases}0.25x+0.5, & -2<x\leq 0\\ 0.25, & 0<x<2\end{cases}$$



- 分别取$N=1$，2，5，10 结果如下



总结

1. 随着N增大随机变量的增多，以上三种抽样的标准化和值都接近于标准正态分布.
2. 连续均匀分布和第三个给定分布，在$N=10$时，都很接近标准正态分布了.
3. 离散的伯努利分布，在$N=100000$ 时，大体满足正态分布相同，但仍然有较大差距. 因此说明，只有 n 很大时，才可以用标准正态分布去逼近二项分布 $B(n, p)$.