计算物理第9题

PB18000039 徐祺云

一　作业题目

自设若干个随机分布（相同或不同分布，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)。

二　算法及主要公式

设若干个随机分布，这里取三种随机分布：$[0,1]$上的均匀分布，$(n,p)=(9,1/2)$的二项分布，$\lambda=1$的指数分布。可以计算出他们的期望与方差，如下表所示：

分布	参数	分布律	数学期望μ	标准差σ
均匀分布	$[a,b]=[0,1]$	$f(x)=\begin{cases}\frac{1}{b-a} & a<x<b\\ 0 & others\end{cases}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
二项分布	$(n,p)=(9,\frac{1}{2})$	$P\{X=k\}=C_n^k p^k q^{n-k}$	np	$\sqrt{np(1-p)}$
指数分布	$\lambda=1$	$f(x)=\begin{cases}\lambda e^{-\lambda x} & x>0\\ 0 & x\le 0\end{cases}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda}$

中心极限定理讲的是，对于一系列独立同分布的随机变量$X_1,...X_n$，有确定的期望与标准差(μ,σ)，那么对任意x，

$$F_n(x)=P\left\{\frac{\sum_{i=1}^n X_i-n\mu}{\sigma\sqrt{n}}\le x\right\}$$

满足

$$\lim_{n\to\infty}F_n(x)=\lim_{n\to\infty}\left\{\frac{\sum_{i=1}^n X_i-n\mu}{\sqrt{n}\sigma}\le x\right\}=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{t^2}{2}}dt=\emptyset(x)$$

即随机变量
$$Y_n = \frac{\langle X\rangle - \mu}{\sigma/\sqrt{n}}$$
近似服从标准正态分布N(0,1).

对于每一种分布f，可以进行N次抽样，求得均值$\langle f\rangle$后代入计算$|\frac{\langle f\rangle-\mu}{\sigma_f/\sqrt{N}}|$的值。重复该过程$num$次，(num取了$10^5$)，可以得到一个长度为$num$的一维数据表，导入ORIGIN中绘制直方图，并绘制标准正态分布曲线与一张图中，比较他们的趋势特性并分析。

抽样方法分析

均匀分布

直接使用16807随机数发生器产生$[0,1]$上的随机数序列，选取N个随机变量计算均值，重复num次即可。

二项分布

对于离散分布，可以采用直接抽样法，可计算概率分布$P(X=k)$与累积分布$Y(X=k)$，通过生成的$[0,1]$间随机数，$Y(X=k)<\xi\leq Y(X=k+1)$确定对应自变量的区间，从而计数。最后得到计算均值，重复num次即可。

指数分布

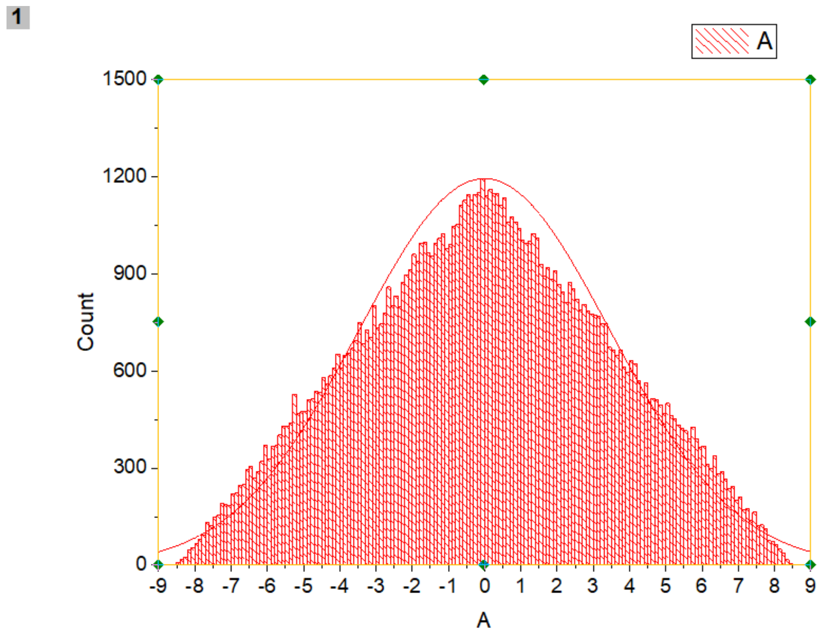
连续性分布可直接抽样，先计算累积分布函数：$\xi=\int_0^x \lambda e^{-\lambda t}dt = 1-e^{-\lambda x}$，反解得到$x=-\lambda^{-1}\ln(1-\xi)=-\lambda^{-1}\ln(\xi)$,通过生的$[0,1]$间随机数$\xi$，计算得到$x$。最后计算均值，重复num次即可。

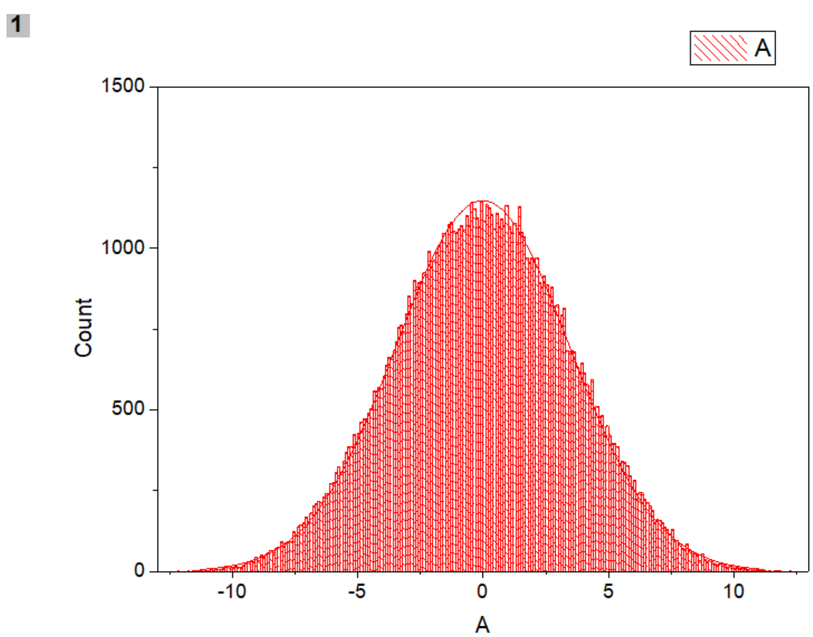
三　计算结果与分析

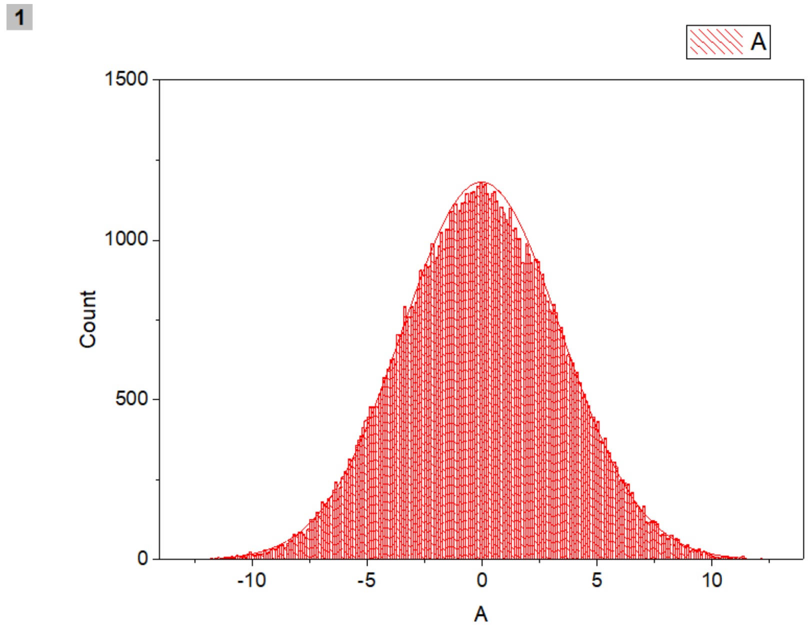
运行程序，得到数据导入软件并绘图。结果如下：

均匀分布

下面三张图依次为$N=2,5,10$:


1
A
1500
1200
900
600
300
0
Count
-9 -8 -7 -6 -5 -4 -3 -2 -1 0 1 2 3 4 5 6 7 8 9
A

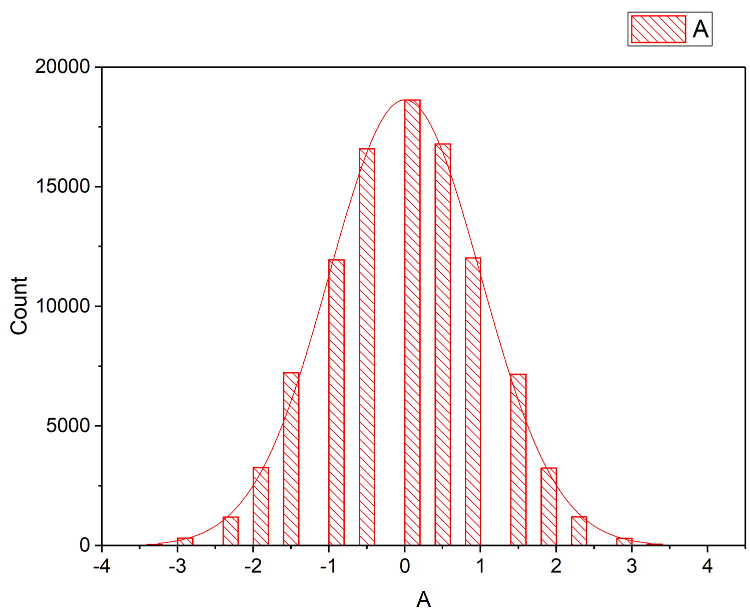

1
A
1500
1000
500
0
Count
-10 -5 0 5 10
A

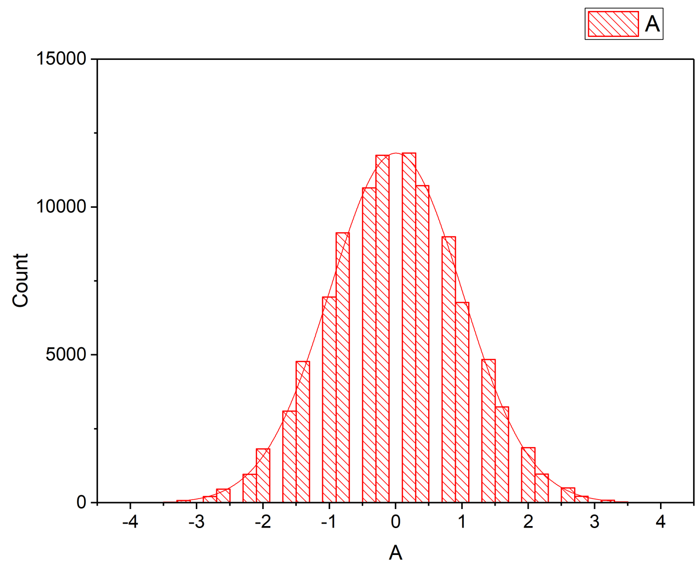


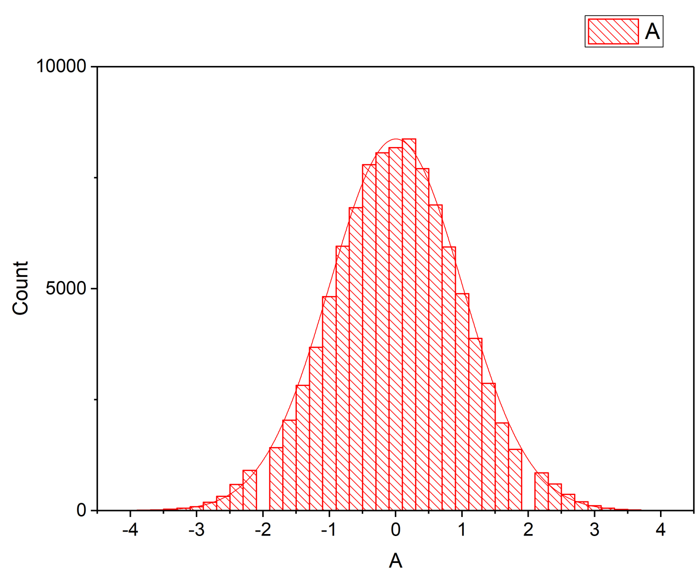
可见，随着N取值的增大，均匀分布逐渐趋于标准正态分布。注意到N并不大，而图像吻合得很好，说明均匀分布收敛于N(0,1)的速度很可观。

二项分布

下面三张图依次为$N = 2, 5, 10$:



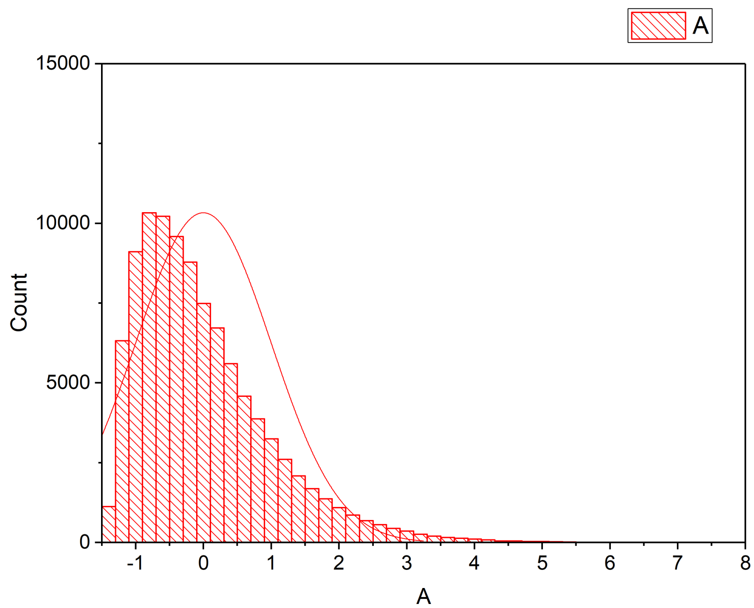

A
15000
10000
5000
0
Count
-4
-3
-2
-1
0
1
2
3
4
A

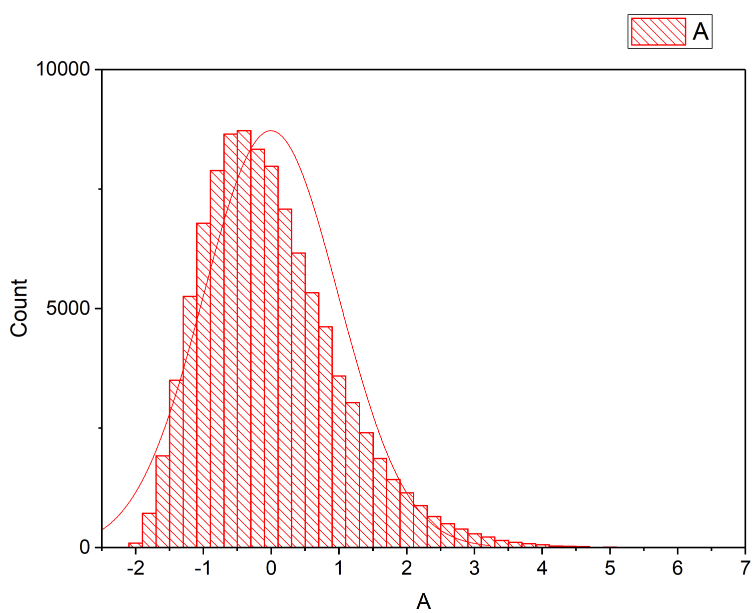

A
10000
5000
0
Count
-4
-3
-2
-1
0
1
2
3
4
A

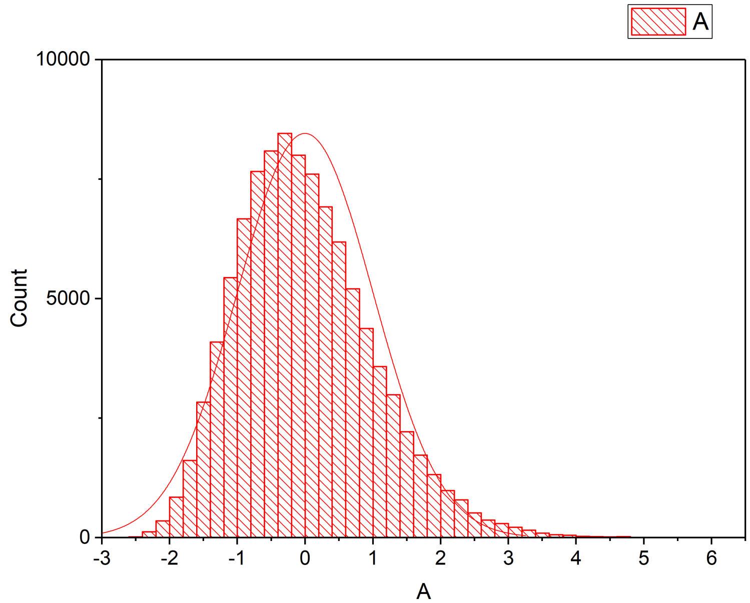
可见，二项分布虽然有缺失的横坐标项(因为是离散分布)，但其直方图分布与标准正态分布吻合得比较好，随着N的增大，图像接近于N(0,1)，收敛速度良好。可以预测随着N的继续增大，单位区间内离散取值数增多，缺失项将会填充，更逼近于N(0,1)。

指数分布

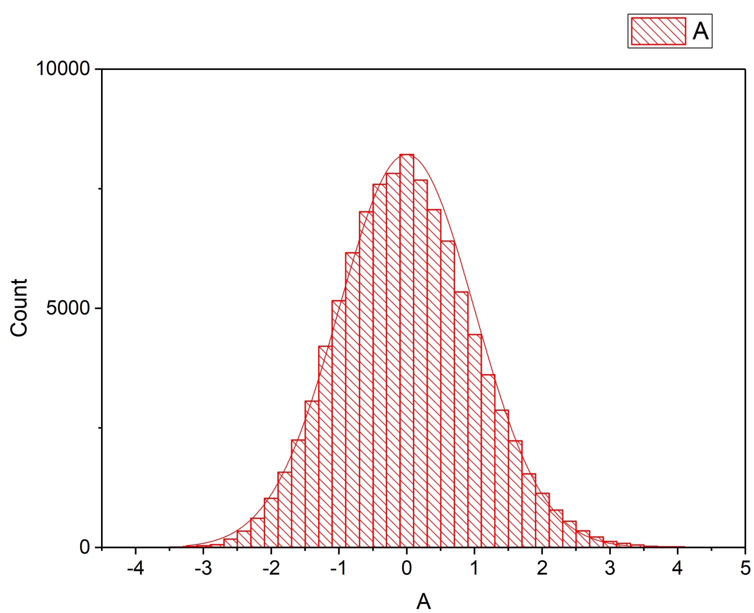
下面三张图依次为$N = 2, 5, 10$:







可见，在N较小时，指数分布直方图并不对称，呈现左陡右宽的形状，随着N的逐步增大，直方图分布向着标准正态分布曲线趋近，但图像效果不如前两个分布。这是因为指数分布本身非对称，其中心极限定理的收敛速度稍逊于前两者。这里令$N = 100$，绘制图像如下：



此时直方图与正态分布吻合得比较良好，这是在我们的预期内的。

四　结论

本题选取了三种分布：$[0,1]$上的均匀分布，$(n,p) = (9,1/2)$的二项分布，$\lambda = 1$的指数分布，计算了他们的抽样方法，绘制了不同N下的直方图，并与标准正态分布比较，发现随着N的增大，二者吻合性提高，验证了中心极限定理的正确性。

值得一提的是，在N相同的情况下，指数分布的图样略差于前两者，即指数分布收敛速度慢于前两者，这是因为分布本身是非对称的，而均匀分布、二项分布的对称性良好，故有此差异。当N继续增大，最后能趋于标准正态分布。