蒙特卡洛算法：基础

计算物理b

高阳

数学基础：随机变量

- 在前面的例子中都需用到在(0,1)的均匀分布随机数
- 随机变量ξ: 有确定的取值区域，但取哪个值不确定，按照某种统计规律获得。
- 例子：掷硬币，头为0，背为1，计ξ标记某次掷的结果，即为随机变量。

数学基础：期望

- 离散型：$\xi \in \{x_i, i = 1,2,\ldots,n\}, p_i = p(\xi = x_i)$.

  需要归一：$\sum_i p_i = 1$
- 连续型：$\xi \in (a,b)$，概率密度$p(\xi)$. 归一：$\int_a^b p(\xi)d\xi = 1$

  $$\mathrm{p}(x)dx = p(x < \xi < x + dx)$$

  分布函数 $p(x) = \frac{d}{dx}P(x)$
- 期望：$\langle\xi\rangle = \sum_i x_i p_i$，$\langle\xi\rangle = \int_a^b d\xi\, \xi p(\xi)$
- 方差 $V(\xi) = \langle\xi^2\rangle - \langle\xi\rangle^2$ 标准差 $\sigma = \sqrt{V}$

数学基础：基本性质

- x,y为随机变量，a,b为常数

$$\langle ax+by\rangle = a\langle x\rangle + b\langle y\rangle$$

$$V(ax+by) = a^2V(x) + b^2V(y) + 2ab\,V(x,y)$$

- 协方差：$V(x,y) = \langle (x-\langle x\rangle)(y-\langle y\rangle)\rangle$
- 协方差说明两个随机变量的关联。

如果两变量独立，则$V(x,y)=0.$

此时有$V(ax+by) = a^2V(x) + b^2V(y)$

对于一系列独立同分布随机变量$\xi_i, i = 1,2,\ldots,N$

$$V\left(\sum \xi_i\right) = NV(\xi_i),\ V\left(\frac{\sum \xi_i}{N}\right) = \frac{1}{N}V(\xi_i), \sigma\left(\frac{\sum \xi_i}{N}\right) = \frac{1}{\sqrt{N}}\sigma(\xi_i)$$

数学基础：独立性

- x, y为随机变量，我们可定义联合概率密度$p(x, y)$
  $p(x, y)dxdy = p(x < \xi_1 < x + dx \ \&\& \ y < \xi_2 < y + dy)$
- 例子：高维空间的概率密度
- 若$p(x, y) = f(x)g(y)$则我们可说x, y相互独立。
- 例子：直接抽样计算π的算法，产生正方形里的点的时候
- 若有多个随机变量，其独立性不能化为成对的独立性
  例子：　$x = r, \mathrm{y} = \mathrm{s}, \mathrm{z} = (\mathrm{r} + \mathrm{s}) \bmod 1.$
  其中，r, s为(0,1)区间的随机数。

数学基础：条件概率

- 事情A的发生可能需要前提B或者C，则

$$p(A) = p(B)p(A|B) + p(C)p(A|C)$$

- 性质：

$$p(A|B) = \frac{p(AB)}{p(B)}$$

- 对于独立事件

$$p(A|B) = p(A)$$

数学基础：大数定律

- ξ_i为独立同分布的随机变量序列，而且ξ_i存在数学期望，记为μ，则对于任意$\epsilon > 0$,我们有

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n}\xi_i - \mu\right| < \epsilon\right) = 1$$

- 例子：掷n次硬币，每次的值ξ_i为随机变量，其期望为1/2 则当n很大时

$$\frac{1}{n}\sum_{i=1}^{n}\xi_i = \frac{1}{2} \quad （概率意义上的相等）$$

- 蒙特卡洛算法应用的基础。

数学基础：中心极限定理

- ξ_i为一列独立同分布随机变量（n个），并且有有限的期望μ和标准差σ,则如下变量近似服从标准正态分布：

$$\xi = \frac{\frac{1}{n}\sum_{i=1}^{n}\xi_i - \mu}{\sigma/\sqrt{n}}$$

也即

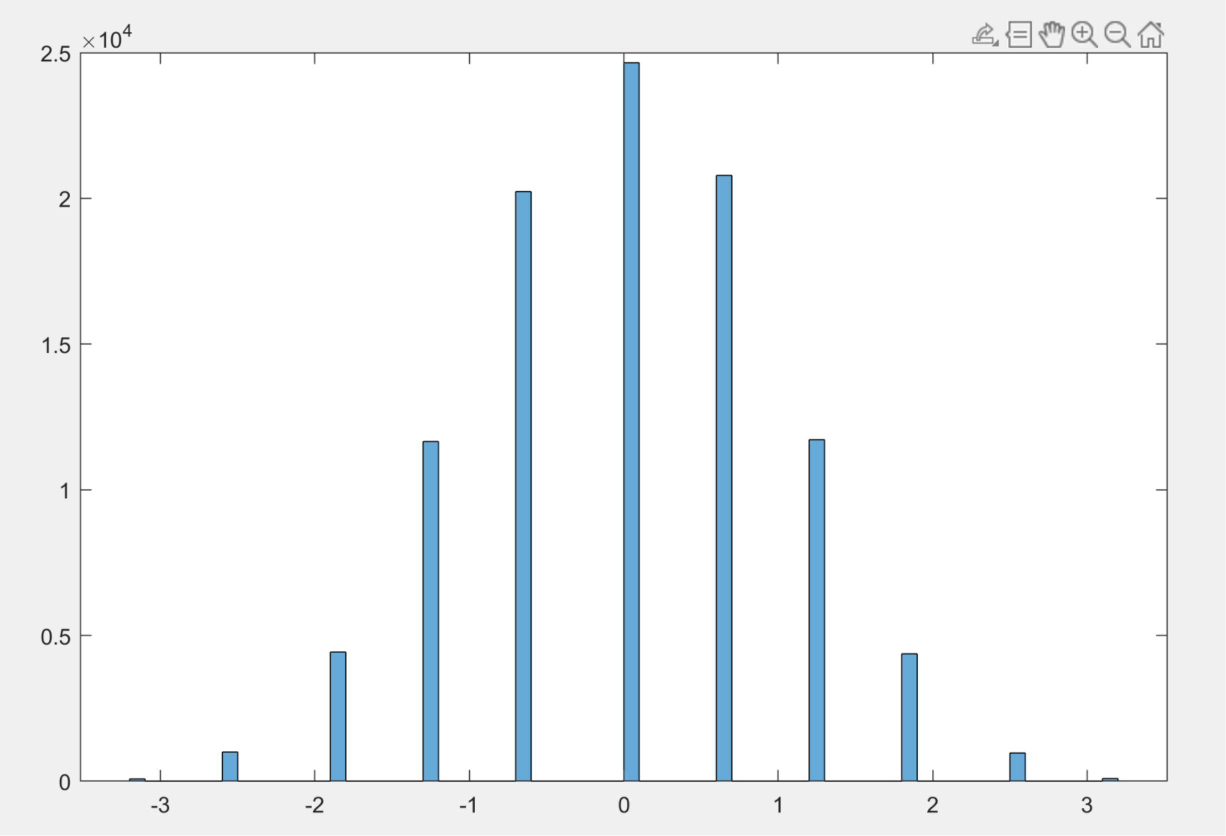
$$\lim_{n\to\infty} P(\xi \le x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{x^2}{2}} dx$$

- 用于误差分析

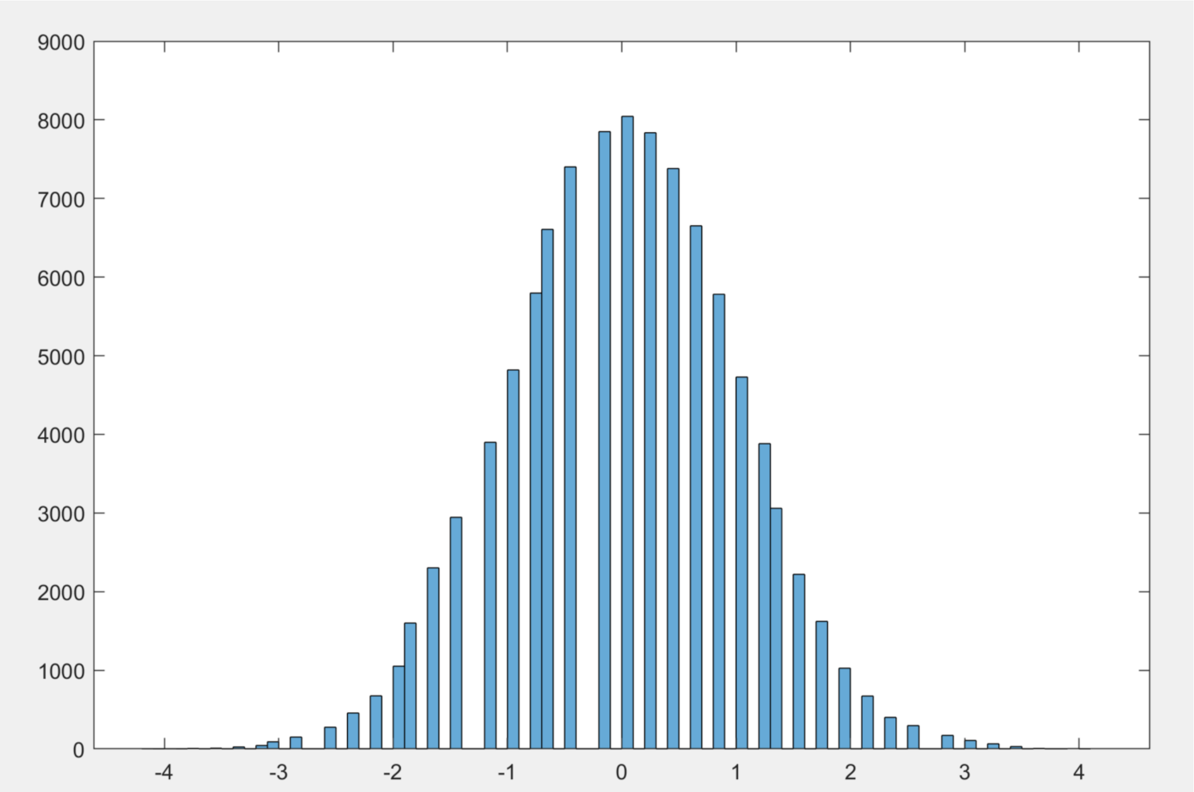
数学基础：检验

- 掷硬币：进行多次重复实验，每次实验掷m次硬币，观察$\xi = \frac{\frac{1}{n}\sum_{i=1}^{n}\xi_i-\mu}{\sigma/\sqrt{n}}$的分布情况。注意，单次实验期望值为1/2, 标准差为1/2

m=10



m=100



数学基础：误差分析

- 期望值的偏离

$$P\left(-\lambda\frac{\sigma}{\sqrt{n}}\le\frac{1}{n}\sum_i\xi_i<\lambda\frac{\sigma}{\sqrt{n}}\right)=\frac{1}{\sqrt{2\pi}}\int_{-\lambda}^{\lambda}e^{-\frac{x^2}{2}}dx$$
$$=\mathrm{erf}\left(\frac{\lambda}{\sqrt{2}}\right)=1-\alpha$$

- 误差在$\lambda\frac{\sigma}{\sqrt{n}}$之内的概率为$1-\alpha$, 此为置信水平。
  α为显著水平。
- 是概率收敛，不是精确收敛。
- 若σ固定，可以通过增大n来减小误差。

数学基础：误差分析示例

每轮900次

轮数	1	2	3	4	5	6	7	8	9
期望值	0.527	0.504	0.489	0.482	0.502	0.520	0.492	0.514	0.489

误差函数，$\lambda = 1$时值为0.68

有68%的概率计算值在$0.5 \pm \frac{0.5}{30} = 0.5 \pm 0.017$的区间内

实际发现占比为$\frac{6}{9} = 67\%$

随机数生成

- 在前面的例子中都需用到在(0,1)的均匀分布随机数
- 真随机数：比如物理过程产生的，涨落之类的。
- 准随机数：不具有随机性，无法通过统计检验，但可能很均匀，仍然可以用来计算。
- 举例：Halton序列(区间均分型，二进制翻转获得）：1/2, ¼, ¾, 1/8,5/8,3/8,7/8,...

伪随机数算法1

- 平方取中法（冯诺依曼提出）

  对于一个2r位的数字，其平方为4r位，取中间的2r位作为随机数，并且构成获得下一个数字的种子。

$$\xi_n = \frac{x_n}{10^{2r}}$$

$$x_{n+1} = [10^{-r}\, x_n^2\,] \left(mod\ 10^{2r}\right)$$

  此方法较差，容易重复，如感兴趣可自己尝试。（课本有误）

伪随机数算法2

- 线性同余法
  通过如下方法获得数列。
  $$x_{n+1} = (a\,x_n + c)\ (mod\ m)$$
  a 为乘子， c为增量， m为模数， x_0 为种子。
  若c=0，为乘同余法
- 与参量值的选择非常相关。
  例子：a=3, m=10, c为偶数，

$$x_0, 3x_0 + c, 9x_0 + 4c, 27x_0 + 13c, 81x_0 + 40c = x_0\ (mod\ 10)$$

一般原则

- (1) c和m互素； （2）若p是m的素因子，$a = 1(mod\ p)$; （3）若 $m = 0(mod\ 4)$, $a = 1(mod\ 4)$; (此为一种方法）此时，周期为m
- 证明请见：https://chagall.med.cornell.edu/BioinfoCourse/PDFs/Lecture4/random_number_generator.pdf
- $m = 2^n$是一种比较便利的选择，可以用到计算机以二进制编码的事实。此时，若$n > 1$（一般肯定满足），则可选$c = 2p + 1, a = 4q + 1,$也即教材中给出的形式。

统计检验：均匀性

χ^2检验：检验实验值是否满足某种分布的方法

对于均匀性检验：将$(0,1)$分成N份，若满足均匀性则在每份的概率应为1/N

做法：产生$(0,1)$内的长度为M的随机数，将（0,1）均匀分成N份，统计每份中点的数目，计算χ^2值，查表判断均匀性。

统计检验：实例

χ^2 检验：产生 (0,1) 内的长度为2000的随机数列，记为A，将（0,1）均匀分成10份。进行假设检验：H0，A在（0,1）内均匀分布
H1，A在（0,1）内不满足均匀分布

区间号	1	2	3	4	5	6	7	8	9	10
频数	188	221	181	197	207	213	209	203	190	191

$\chi^2 = \sum_{i=1}^{N} \frac{(m_i - M/N)^2}{M/N} = 7.22$ 服从$\chi^2(N-1)$，$\chi^2(9,7.22) = 0.39$ 若H0成立，服从χ^2分布，则不小于7.22的概率为0.61，我们确实在此区间，故61%的概率接受H0.

注：也可均分为两个区间，$\chi^2 = 0.072$，$\chi^2(1,0.072) = 0.21$，有79%的概率可认为在二分区间均匀。实际上，撒点变多可提高均匀性。

统计检验：独立性

χ^2检验：检验实验值是否满足某种分布的方法

两事件独立性检验：若独立性成立，我们有如下理论预言，也即联合概率等于分立概率之积。基于此，可适用χ^2检验。

做法：产生M个随机数，两两成对，每队第一个作为变量1，第二个作为变量2，将变量1和2的取值区间分划，利用独立时间条件概率检验独立性。

统计检验：实验值

χ^2检验：产生10000个随机数，形成5000对，将（0,1）区间分成3份，从而对于两个变量的联合分布，有9个区间。列表计算频数。

	$x1 \in \left(0,\frac{1}{3}\right)$	$x1 \in \left(\frac{1}{3},\frac{2}{3}\right)$	$x1 \in \left(\frac{2}{3},1\right)$	计数
$x2 \in \left(0,\frac{1}{3}\right)$	546	548	529	1623
$x2 \in \left(\frac{1}{3},\frac{2}{3}\right)$	548	542	560	1650
$x2 \in \left(\frac{2}{3},1\right)$	584	560	583	1727
计数	1678	1650	1672	

统计检验：理论频数与检验

	$x1 \in \left(0,\frac{1}{3}\right)$	$x1 \in \left(\frac{1}{3},\frac{2}{3}\right)$	$x1 \in \left(\frac{2}{3},1\right)$	计数
$x2 \in \left(0,\frac{1}{3}\right)$	544.68	535.59	542.73	1623
$x2 \in \left(\frac{1}{3},\frac{2}{3}\right)$	553.74	544.50	551.76	1650
$x2 \in \left(\frac{2}{3},1\right)$	579.58	569.91	577.51	1727
计数	1678	1650	1672	

若为均匀分布，则每个矩阵元可以计算出来。比如第一个元的理论值应为：1697/5000*1633=554.24 这里不用5000/9是因为不预先假设变量为均匀分布，做与均匀性无关的独立性检验。

$$\chi^2 = \sum_{i=1}^{N} \frac{(m_i - M/N)^2}{M/N} = 1.0904 \text{ 服从} \chi^2((x-1)*(y-1)),$$

$\chi^2(4,1.0904) = 0.104$,故有89.6%的概率接收H0，也即独立性。

（注：若按照理论均匀频数5000/9来计算，则$\chi^2 = 4.8532, \chi^2(4,1.0904) =$ 0.697，只有30%的概率接收独立性。）

作业

1. 根据上次作业中计算π的值的算法，设置投点步数N=10000,分别写出置信水平为68%和95%的误差区间(对于误差函数，$\lambda = 1$时值为0.68，$\lambda = 2$时值为0.95）。

2. 教材第二章习题1.