18　乱数とモンテカルロシミュレーション

18.1　疑似乱数

自然現象（そして社会現象）の多くは確率的な要素を含む。仮に系の時間発展が原理的に決定論的なものであったとしても、実際には我々は全ての要素を知ることができず、確率的な議論をしなければならないことは多い。例えば、サイコロを振って次にどの目が出るか、は原理的には初期条件がわかれば運動方程式を解けば求まるのであろうが、実際にサイコロを振る時の初期条件を正確に知ることは困難である。そこで確率的な議論をするわけである。系の時間発展が単純な規則で記述される場合も、初期条件の微小な差が時間とともに拡大される場合がある (いわゆるカオス)。

逆に、確率的な議論を行なうことで、決定論的に扱うのは非常に困難な系を簡単に扱うことができる場合がある。その典型は統計力学である。

いろいろな現象をコンピュータで数値的に解析する場合も、確率的な要素を導入すると便利なことが多い。しかし、コンピュータの計算は決定論的であるため、普通にやると確率的な要素は入ってこない。

コンピュータで確率的な計算をするには、「乱数」つまりランダムな値を取る確率変数が必要である。実際には何らかの決定論的な計算で乱数を作るため、真の乱数は得られない。そこで、正確にはコンピュータで用いる、決定論的な計算によって生成する乱数に代わる数を「疑似乱数」と呼ぶ。「疑似乱数」がどの程度「真の乱数」に近いか、というのは難しい問題である。

当然ながら、計算方法によって疑似乱数の「質」には大きな違いがある。質の悪い疑似乱数は、例えばそれぞれの疑似乱数の相関が強く、独立性が悪いなどの問題がある。ゲーム等で利用する場合には、疑似乱数の質はそれほど問題にならないだろうが、特に大規模なモンテカルロ計算を行なう際には、乱数の質の悪さによって結果が影響され、せっかく計算しても誤った結論を得ることもある。

ほとんどのシステムで用意されている疑似乱数は簡単な方法を用いていて、質が良くない。従って、シミュレーションに用いる場合は、特に保証のある場合を除いて基本的にシステムで用意された疑似乱数は使わない方が良い。

性質の良い乱数として最近注目されているものに Mersenne Twister と呼ばれているものがある。

`http://www.math.sci.hiroshima-u.ac.jp/~m-mat/MT/mt.html`

を参照のこと。ソースコードが配布されているので、ダウンロードして動かしてみよう。最新版のプログラム`mt19937ar.c`は、`~oshikawa/samples/`にも置いておくのでコピーしても良い。`init_genrand`が初期化を行う関数、`genrand_rand1`等が実際に 0 から 1 までの一様乱数を返す関数である。

18.2　簡単な乱数の利用方法

例えばメインルーチンを次のようにして、Mersenne Twister の関数を呼び出すと、0 から 1 までの一様乱数を表示することができる。

```
int main(void)
{
  int i,seed;
  double x;

  seed = 100;
  init_genrand(seed);


  for(i=0;i<10;i++)
    {
      x = genrand_real1();
      printf ("%f \n",x);
    }
}
```

実際に実行してみると、genrand_real1() を用いる毎に乱数が与えられていることがわかるだろう。同じプログラムをもう一度実行してみると、先ほどと同じ結果を与える。これはgenrand_real1() が真の乱数ではなく、ある決まった規則に従って生成されている疑似乱数であることによっている。

init_genrand(seed) は疑似乱数を初期化する。seed は乱数の「種」、すなわち乱数列の初期状態を与える。その名の通り、**初期化は通常の場合、最初に一度だけ行えば良い。** seed を決めると、その後発生する乱数列は一意に決まる。整数 seed の値を変えると、異なる乱数列が得られる。

実行するごとに異なる乱数列を得たい場合、例えば seed=time(0); とすると良い。実行時の時刻（を秒に換算したもの）が seed になるので、実行時の時刻によって異なる乱数列が得られる。

課題

疑似乱数を用いて、人間とジャンケンをするプログラムを作成せよ。

18.3 モンテカルロ計算

確率現象のシミュレーションなど、乱数を用いた計算を一般にモンテカルロ法と呼ぶ。（有名なカジノであるモンテカルロに因む。）

練習問題

最も簡単な例として次のような方法で円周率を近似的に求めてみよう。

- xy 平面上で (0,0),(1,0),(1,1),(0,1) を頂点とする正方形の中の一点を一様乱数によりランダムに生成する。
- この点が原点から距離 1 以内にあるかどうかを調べる
- 多数の点に関してこれを行ない、距離 1 以内にある確率を調べる。
- この確率は 1/4 円の面積に等しいので、確率を 4 倍すれば円周率を推定することができる

18.4 中心極限定理と統計誤差

モンテカルロ法では、確率的なシミュレーションを行うので厳密に正しい値を得ることはできない。無限回のシミュレーションを行えば厳密な値が得られるだろうが、もちろん実際には有限回のシミュレーションしか行えないので必ず誤差が発生する。このように、有限個のサンプルから推定することに伴う誤差を統計誤差と言う。このことは、前節の練習問題で求めた円の面積について考えても明らかであろう。

従って、モンテカルロ法においては、得られた結果について統計誤差を評価することが重要である。多くの計算では、多数の試行について得られた結果の平均が求める値の推定値となる。すなわち、j 回目の試行で得られる結果は X_j と言う確率変数で与えられる。N 回の試行で得られる結果の平均は

$$A_N = \frac{X_1 + X_2 + \ldots + X_N}{N}$$

となる。この A_N も確率変数である。ここで、簡単のため、各試行が等価であり、各試行で得られる結果が独立であると仮定しよう。(つまり、毎回同じ確率現象のシミュレーションを行い、得られた結果は次回の結果に影響を与えないとする。) このとき、独立な確率変数の和である A_N については統計学で非常に重要となる中心極限定理が成立する。それによると、以下のことが成立する。

> 毎試行の際の結果を表す確率変数 X_j の平均を $\bar{X}$、標準偏差を σ とする。(仮定により、平均と標準偏差は j によらず一定である。) このとき、A_N は $N \to \infty$ の極限で平均 $\bar{X}$、標準偏差 $\sigma_A = \sigma/\sqrt{N}$ の正規分布に近づく。

なお、確率変数 R が平均 $\bar{R}$、標準偏差 σ_R の正規分布に従うとは、$r < R < r + dr$ となる確率が微小な dr に対して

$$\mathrm{Prob}(r < R < r + dr) = \frac{1}{\sqrt{2\pi}\sigma} \exp - \frac{(r - \bar{R})^2}{2\sigma^2} dr$$

となることである。統計学の詳細、中心極限定理の証明等は本講義の範囲を超えるので興味があれば専門書を参照すること。正規分布は統計学で最も重要な分布の一つであり、多くの性質が良く知られている。たとえば、平均値からの差が標準偏差以下である確率は約 68 パーセント、標準偏差の 2 倍以下である確率は約 95 パーセントである。

N 回の試行から求めた A_N の確率変数としての平均は、求めるべき真の値 $\bar{X}$ に等しい。(すなわち、N 回の試行で A_N を求めることを更に多数回くりかえせば、A_N の更なる平均は $\bar{X}$ に漸近する。) しかし、実際に得られる A_N は確率的に分布する。この分布の標準偏差 σ_A が A_N の統計誤差の目安を与える。N が十分大きければ A_N の分布は中心極限定理によって正規分布で近似できる。

実際には真の値 $\bar{X}$ は不明であり (わかっていたらモンテカルロシミュレーションをする必要はない!) 計算の結果得られるのは A_N の値である。そこで、逆に $\bar{X}$ は得られた A_N を中心に標準偏差 σ_A の正規分布にしたがって確率的に分布していると考えられる。たとえば、求めるべき値 $\bar{X}$ が A_N から距離 σ_A 以内にある確率が約 68 パーセントであることになる。このように、標準偏差 σ_A が統計誤差の目安を与える。

あるシミュレーションに対して、σ は一定であるから、統計誤差 $\sigma_A = \sigma/\sqrt{N}$ は試行回数 N の平方根に反比例して減少する。すなわち、統計誤差を 1/10 にするためには 100 倍、1/100 にするためには 1 万倍の計算を行う必要があることになる。

なお、中心極限定理はあくまでも一定の仮定のもとに成立する定理であることに注意すること。例えば、実際の計算では場合によっては**各試行が独立でない場合もある。この時は、中心極限定理を単純に適用することはできない。**

練習問題

$(0,1)$ 区間の一様乱数を N 個合計した量の確率分布をいろいろな N に対してモンテカルロシミュレーションによって調べよ。得られた結果を中心極限定理と比較せよ。

ヒント: 実は N はそれほど大きくなくても中心極限定理と比較できる結果を得る。これは、コンピュータ上で近似的に正規分布に従う乱数を発生させる時にも用いる方法である。

練習問題

円の面積をモンテカルロ法で求める問題について、統計誤差を推定し真の値とのずれと比較せよ。

19　さまざまな応用

練習問題:ランダムウォーク

d 次元正方格子 ($d=2$ なら正方格子、$d=3$ なら立方格子) がある。すなわち、格子点は x_j を全て整数として $(x_1, x_2, \ldots, x_d)$ で与えられる。各点の周りに、$2d$ 個の格子点が隣接している。

この格子の上をランダムに運動する粒子を考える。この粒子は、各ステップごとに、$2d$ 個の隣接格子点の一つを等確率で選んで飛び移る。

最初原点にいた粒子が、その後ある時点で原点に帰ってくる確率をシミュレーションによって求めよ。

ヒント

上のランダムな運動を多数回シミュレーションすると確率を推定することができる。粒子が帰ってきたらその時点で計算を停止すれば良いが、帰ってこない場合は無限に運動が続いてしまう。そこで、適当な判断基準を設けて、たとえば原点からある距離 R 以上離れたら戻ってこないと判断する。ただし、本来は $R \to \infty$ とすべきなので、実際に R を変化させても結果が変わらない程度に大きく取るべきである。

練習問題: ビンゴゲーム

あなたはパーティーの幹事をすることになった。ここで余興としてビンゴゲームが計画されている。ビンゴゲームとは次のようなルールに従ったゲームである。

各参加者は5x5の枡目を持ったカードを受けとる。

- 中央のマスには最初から穴を開ける。
- 他のマスには、1から75までの数がランダムに入っている。
- ただし、第k列には $15(k-1)+1$ から $15k$ までの数しか入らない。$(k=1,2,3,4,5)$
- 同じ数が一つのカードの複数のマスに入ることはない。

皆がカードを受けとった後、司会者はビンゴマシーンを用いて1から75までの数字の入ったボールをランダムに一つずつ取り出し、出てきたボールに書いてある数字を読み上げる。各参加者は自分のカードにその数字があればそのマスに穴を開ける。持っているカードの穴がタテ、ヨコまたはナナメのいずれかに5つつながれば「ビンゴ」(上がり)である。

N 人の参加者のうち、最初にビンゴになった M 人には賞品が用意されているとする。ただし、同時に何人もがビンゴとなって、その中で順位をつける必要がある場合はジャンケンで決めることにする。すなわち、N 人の参加者のうち M 人以上がビンゴになった時点でゲーム終了である。

パーティーの時間が限られているので、あらかじめどれくらいの時間がかかるかを予想しておきたい。例えば一つのボールを取り出して数字を読み上げ参加者が対応するのに30秒かかるとすると、全てのボール(1-75)を取り出すには40分近くもかかってしまう。しかし、実際には全てのボールを取り出す前にゲームが終了するだろう。

予想される適当なN,Mを決めて(例えば $N=100$,$M=10$)、ゲームが終了するまでに何個のボールを取り出す必要があるかをモンテカルロシミュレーションで調べよ。(取り出すボールの個数の確率分布を調べよ。)

練習問題

二次元正方格子を考え、各辺が独立に確率 p でつながっているとする。このとき、つながっている辺をたどって到達できる辺の集合をクラスタと呼ぶ。あるクラスタに含まれている辺の数をクラスタの大きさと定義しよう。

$p = 0$ ならばどの辺もつながっていないのでクラスタも存在しない。一方、$p = 1$ ならば全ての辺がつながっているので、格子全体が一つのクラスタになる。一般には、格子には複数のクラスタができる。

p を変化させていくと、ある値 p_c で一種の相転移が起こり、$p > p_c$ では (無限に大きな系では) 無限に大きなクラスタが出現する。$p < p_c$ では無限に大きなクラスタは出現しない。

(1)

p を変化させた時の相転移を、シミュレーションで確かめよ。もちろん、シミュレーションでは無限に大きな系は扱えない。そこで、有限の大きさ $L \times L$ の系でシミュレーションを行い、クラスタの大きさの分布が、L を大きくした時の変化を調べよ。

(2)

実は、上の問題では正確に $p_c = 1/2$ である。これを示せ。

(3)

$p = p_c = 1/2$ での最大のクラスタの形状を議論しよう。最大クラスタの形を実際に表示してみると、複雑な形をしていることを確かめよ。これは、以下に見るように、いわゆるフラクタル図形と呼ばれるものの一種である。

格子の大きさ $L \times L$ の時の最大のクラスタの「直径」は L で与えられる。この最大クラスタのサイズ (クラスタに含まれる辺の数) を s とする。s は面積に相当する量なので、通常の 2 次元の図形ではその形状によらず

$$s \propto L^2$$

である。最大のクラスタのサイズ s をいろいろな系の大きさ L に対して調べると、

$$s \propto L^D$$

という関係が成立する。これをシミュレーションで確かめ、D の値を求めよ。

実は $D < 2$ になっている。このことは、クラスタが図形として非整数 (フラクタル) 次元を持っていることを意味している。このような図形をフラクタル図形と言う。(ただし、D は 2 に近いので、十分な計算を行わないとクラスタのフラクタル性を確認することはできない。)