ソフトウェア活用法

ランダムウォークとガウス（正規）分布

2009/6/22 長島雅裕（4 階 413）

1. ブラウン運動

1905 年、アインシュタインは 3 つの論文を発表した。その内容は、「特殊相対性理論」「光電効果（光量子説）」「ブラウン運動」の理論である。アインシュタインは光電効果でノーベル賞を取った。これは現代の量子力学の基礎をなすものである（高校物理でも出てくる）。特殊相対性理論は光速に近い速さで運動する物体に関する理論である。

さて、ブラウンは水の中の花粉粒子を顕微鏡で見るとランダムに運動していることを発見した。これをブラウン運動と呼ぶ。しかし、その運動の起源は謎であった。アインシュタインは、水の分子のうち高速のものがランダムに花粉に当たり運動量を与えることでランダムな運動が起こるとし、それを理論にまとめた。これは原子・分子の実在の証明にもつながった。さらに、ミクロな分子レベルの運動と、温度や密度などのマクロな物理量とをつなぐ役割も果たしており、マクロ系の物理学をミクロから基礎付ける上で欠かせない重要な理論となっている。

今回は、ブラウン運動を手始めに「ランダム」さに基づく「規則性」に注目して、エクセルを使ってみよう。

2. ランダムウォーク（乱歩、酔歩）

さて、ランダムに動く花粉の運動を、2次元面に限定されているとする（xy方向のみ）として、エクセルで実現してみよう。

1. A1,B1 セルに、それぞれ 0 を入れる。これは、花粉の出発点を$(x,y)=(0,0)$にする、ということである。
2. 乱数を与える関数 RAND()を使って、次のステップでの位置を与える。花粉の運動はマイナス方向も考えられるので、0.5 を引いて、-0.5 から 0.5 まで動けるようにしよう。そのため、A2 セルには =A1+RAND()-0.5 と書く。同様に、B2 セルには=B1+RAND()-0.5 と書く。
3. さらに、また黒枠をドラッグして、10 ステップランダムに動かしてみよう。(x,y)=(A11,B11)が 10 ステップ進んだ後の粒子の位置となる。
4. 軸の範囲を「軸の書式設定」を使って固定しておこう。

これを図にする。**グラフウィザードボタンを押し、散布図**を選択する。すると、ランダムに動く花粉の様子がなんとなく見て取れる。これを 2 次元での**ランダムウォーク**という。

もっと長時間動かしてみよう。さらに A10,B10 セルの黒枠をドラッグし、一気に 1000 ステップまでやってみよう。ランダムに動きつつ、ひとところにしばらく留まったり、一方向に動いたり、乱雑な花粉の運動の様子が再現されている。F9 を押せば乱数が振りなおせるので、何回か F9 を押してみて、乱雑さを味わってみよう。

なお、1000 ステップ進んでも、たいていの場合、花粉の位置は±10 程度のあたりにいる。これは、ランダムウォークの場合、N ステップたつと、$\sqrt{N}$ ステップ分だけ一方向にすすんだのと（統計的に）ほぼ等しいからである（行きつ戻りつしながら進むので）。いま 1 ステップでは平均的に 0.3 程度（=標準偏差）進む（後述）ので、1000 ステップでは大体 $0.3\times\sqrt{1000}≒9.5$ になる。

3. ランダムウォークの統計

次にランダムウォークの統計を考えてみよう。2 次元で扱うのは面倒なので、さらに簡単化して、1 次元に落とそう。x 軸上を左右にランダムに動くとするのである。

まず、実際に**沢山の花粉を用意して、その軌跡を見てみよう**。

1. A 列に、0～10 までの数字を書く（A1～A11 セル）。これがステップ数になる。
2. B1 セルに 0 と書き、B2 セルには=B1+RAND()-0.5 と書こう。
3. そのまま下へドラッグし、とりあえず 10 ステップ分進めよう。
4. 次に B1～B11 セルを選択し、右方向へドラッグしよう。とりあえず 10 セル分作ろう。これは、各列が別々の花粉に対応しているので、10 個の花粉を用意したことに対応する。
5. これをグラフにする。折れ線グラフを選び、横軸に時間（ステップ数）、縦軸に位置となるようにする。x=0 の点から、時間がたつにつれて（右に行くにつれて）花粉の位置がバラけるのが分かる。そして、多くの花粉が、最終的には x＝0 付近におり、一部の花粉が離れたところへ運ばれているのが分かるだろう。

では、花粉の位置を統計的に予測できるだろうか？

「統計」を議論したいので、もっと多くのステップ、もっと多くの花粉で調べてみよう。

まず、黒枠を下にドラッグし、ステップを増やそう。100 ステップ用意しよう。次に右にドラッグし、花粉の数を増やそう。B から始めると、CW が 100 個目になる。これで、100 個の花粉が 100 ステップ進んだときに、最終的に x 軸上のどの位置にいるかを与えたことになる。

上で作った軌跡の図に、新たに計算したものを足しておこう。グラフをクリックするとセルに青枠が出るので、これを広げ、今回計算した部分がふくまれるようにする。すると、先ほどよりも、明らかに 0 付近に多くの粒子が集中し、わずかな花粉のみが 0 から遠く離れたところに位置するのがわかるようになる。

では、次に最後（100 **ステップ目）での粒子の位置のヒストグラム**を作ろう。

1. ヒストグラムの横軸、すなわちデータをどう分割して個数を数えるかを指定する必要がある。ここでは、A110 あたりから、縦にデータ範囲を指定して書いておく。たとえば、-15.5, -14.5,...:15.5 と縦に並べて書く。
2. 「分析ツール」を使ってヒストグラムを作る。「ツール」メニューから、「分析ツール」を選ぶ。もし分析ツールがなければ、「アドイン」を選び、分析ツールにチェックを入れよう。「分析ツール」→「ヒストグラム」を選ぶ。「入力範囲」は、最後のステップを横にドラッグし、全花粉粒子の最後の位置になるようにする。「データ区間」はさきほど入力したヒストグラムのデータ範囲である。「出力オプション」で、データ範囲のすぐ横あたりに出力させるようにしよう。これで OK をクリックすると、出力オプションで指定したセルにヒストグラムの数値データが出力される。
3. グラフウィザードを使い、今度は棒グラフで示してみよう。「下のデータ」から、「項目軸ラベルに使用」で「データ区間」のセルが選ばれるようにしておくこと。すると、大体 0 付近に粒子が多数あり、0 から離れるに従って粒子数が減ることが分かる。
4. F9 を押して新たな分布の集団を作り、上と同様の操作を繰り返してみよう。多少のバラつきはあるものの、だいたい 0 付近に粒子が集まり、離れるにつれて数が減るという傾向は変わらないはずである。

4. √*N*の法則

さて、思考実験としてもっと単純な場合を考えよう。1 ステップごとに、+1 または-1 進むとする。大勢の人が *N* ステップ進んだ場合の位置の標準偏差を求めよう。

まず平均値であるが、+1 と-1 が毎回等確率で出るので、大勢で平均すると 0 になるはずである。

次に分散である。分散は（自乗の平均）－（平均の自乗）で表される。いま平均は 0 なので、自乗しても 0 である。自乗の平均であるが、1 ステップの自乗が $(\pm1)^2=1$ なので、*N* ステップ進むとどの人も $N\times(\pm1)^2=N$ になる。全員が *N* なので、平均しても *N* である。よって、分散は N-$0=N$ である。

標準偏差は分散の平方（√）なので、結局√*N* となる。つまり、1 ステップ当たりランダムに±1 進むランダムウォークの場合、*N* ステップ進むと平均としては 0 であるが、0 のまわりに大体√*N* 程度拡がることを示している。

つまり、ランダムなプロセスの場合（母集団から *N* 個あるいは *N* 人ピックアップする場合など）、期待値に対して√*N* 程度の誤差が通常生じることを示している（割合にすると、$\sqrt{N}/N=1/\sqrt{N}$）。

5. ガウス（正規）分布と中心極限定理

さて、ここで統計の一般論の話をごく簡単にする。*n* 個のランダムな値の和は、*n* の値を大きくしていくと、ある形の分布に従う。これを正規分布、あるいはガウス分布と呼ぶ。別の観点から言うと、ある母集団から *n* 個の標本をサンプルとして取り出した時、その *n* 個のサンプルの平均は正規分布に近づく。これを**中心極限定理**と呼び、非常に強力な数学的定理である。

以下、今回のケースに即して考える。*i* ステップ目で与えられた-0.5～0.5 の乱数を *Xi* とすると、*n* ステップ目での値は

$$X_n=\sum_{i=1}^{n} X_i$$

となる（無論、ここでは $X_0=0$ である）。各 *Xi* の期待値（平均値）は、*x* に分布関数 *p(x)*をかけて積分すれば得られるが、今回は区間[-0.5,0.5)で一様であるので $-0.5<x<0.5$ で $p(x)=1$、それ以外で 0 であるため、

$$E(X)=\bar{X}=\int_{-0.5}^{0.5} xp(x)\,dx=0$$

分散（自乗平均-平均の自乗）は

$$V(X)=\int_{-0.5}^{0.5} x^2\,p(x)\,dx-\left(\int_{-0.5}^{0.5} x\,p(x)\,dx\right)^2=0.083$$

となる。標準偏差 σ は分散のルート（$\sigma=\sqrt{V}$）であり、約 0.29 である。

正規分布は、一般に次のように書かれる（導出は統計学の教科書を参照して欲しい）:

$$f(X_n)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(X_n-\bar{X})^2}{2\sigma^2}\right)$$

これは積分すると1になるように規格化されている。つまり確率分布ともみなすことができる。

では、実際に上で作ったヒストグラムが正規分布になっているかどうか、確かめよう。注意すべき点は、（1）100

ステップ進んでいるため、分散に100をかける（あるいは標準偏差に$\sqrt{100}$をかける：$\sqrt{N}$の法則）、(2) 花粉100個のヒストグラムなので、分布を積分すると100になるようにする、つまり上の正規分布に100をかける、ということである。

データ区間がB110から下に書かれているとする（違う位置に書いた場合は適宜変更すること）。その横のほうのセルに、=100*EXP(-0.5*B110^2/(100*0.083))/SQRT(2*PI()*(100*0.083))と書いてみよう（正規分布の100倍を表している）。ここでPI()はπを与える関数である。<u>最初に100を掛けたのは、花粉の個数が100個だから</u>である。<u>0.083に100がかかっているのは、100ステップだから</u>である。また<u>今回は区間の幅を1に取っているが、幅を0.5にしたら0.5を書ける必要がある</u>ことに注意。これを例によって下のほうへドラッグし、データ区間に対応する範囲で数値を求めておく。

次にグラフをクリックし、「元のデータ」を使って正規分布の数値もプロットするようにする。このままでは棒グラフが二つでわかりにくい。そこで、「グラフの種類」をクリックし、「ユーザー設定」で折れ線—縦棒を選ぶ。これで、実際の花粉の位置の分布は縦棒で、理論予測（正規分布）は折れ線で描かれ、見やすくなる。

実際に見比べてみてどうであろうか？F9を押して何度かヒストグラムを作り直し、正規分布に近づいていることを実感して欲しい。

6. 終わりに

この5回の講義を通じて、エクセルの使い方の基礎の基礎をマスターしたはずである。エクセルに限らず、ソフトウェアは道具であり、道具は使えば使うほどその使い方が上達する。エクセルについても講義では教えきれないくらいの多くの機能がある。自分でいろいろ使ってみて、試行錯誤しながら自らスキルアップする能力を身につけてほしい。

7. 課題（それぞれ番号・名前を忘れずに）

1. 2次元のブラウン運動（ランダムウォーク）のグラフを1枚印刷し提出する。
2. 3で作成した100ステップ、100粒子の軌跡の図を印刷し提出する。
3. 100ステップ目でのヒストグラムと正規分布を重ねた図を印刷し提出する。
4. 余裕があれば、10ステップ目でのヒストグラムと正規分布も重ねた図を描く。3では1ステップの分散0.083を100倍したが、今回は10ステップなので10倍することに注意。分布が狭くなる（ステップが進むごとに分布が広くなる）ことを確認せよ。

※長島担当のこの5回分について批判や感想など書いてもらえると、
今後に役立てられるので有難いです。
当然、成績とは関係ありません。