サイコロで暗号を解く？

白井朋之 (九州大学数理学研究院)

2009 年 7 月 31 日

概 要

皆さんはランダム (random) という言葉でどのようなことを想像されるでしょうか？CD プレーヤーの再生モードの中にランダム再生というのがあったり,「世論調査でランダムに有権者を抽出して…」などと使われたりしますので，そのイメージは持っておられるかと思います．多分「無作為，乱雑，出鱈目 (でたらめ)，予測不能」などの意味で使われることが多いのではないかと思います．出鱈目の「目」はサイコロの目のことだという説もあるようですが，実際，サイコロはコインと同様にランダムネス (ランダムの名詞形) を生み出す，とても手頃な道具です．確率論はそんなランダムな現象を扱う数学の一分野ですが，その確率論の中で重要なモデルの一つにランダムウォーク (酔歩) と呼ばれるものがあります．酔歩とありますように，酔っぱらいの千鳥足をモデルにしたものですが，以前『ウォール街のランダム・ウォーカー』という本がベストセラーになったこともありますので，もしかしたらそちらでご存知の方もいらっしゃるかもしれません．今回はこのランダムウォークとその親戚であるマルコフ連鎖を題材にして，ランダムであるがゆえに見えてくる現象や有用な方法論について，いくつかの話題を取り上げてお話しする予定です．思いもかけないところで確率論の考え方が役に立ちそうだということを少しでもお伝えできたらいいなと思っています．

最後に，以下はある有名な英文小説の冒頭の一文をある方法で暗号化したものですが，この暗号をサイコロを用いて解くことができるでしょうか？

```
onvqrkgockprtvwwvwtkdxktrdkfrbikdvbrakxskcvd
dvwtkpikurbkcvcdrbkxwkdurkpowmkowakxskuofvwt
kwxduvwtkdxkaxkxwqrkxbkdgvqrkcurkuoakyrryrak
vwdxkdurkpxxmkurbkcvcdrbkgockbroavwtkkpedkvd
kuoakwxkyvqdebrckxbkqxwfrbcodvxwckvwkvdkkowa
kguodkvckdurkecrkxskokpxxmkkkduxetudkonvqrkk
gvduxedkyvqdebrckxbkqxwfrbcodvxwkk
```

目 次

1　はじめに

「二人がお金を賭けて公平なゲームを行ない，先に 3 勝した方がその賭け金をすべて受けとるという約束をしたが，ある事情で 2 勝 1 敗の時点でゲームを中断せざるを得なくなった．ゲームの賭け金を公平に分配するにはどうしたらよいか？」この問いについてフェルマー (P. de Fermat) とパスカル (B. Pascal) がかわした書簡がランダムネスを数学的に扱う確率論の萌芽といわれています．ランダムネスとは何かという哲学的な問題は今も議論のあるところですが，1933 年にコルモゴロフ (A. N. Kolmogorov) が『確率論の基礎概念』を著したのが，集合と公理系から出発するいわゆる現代数学としての確率論の始まりと言えるかもしれません．

ゲームや博打ではランダムネスが重要な一要素で，そのランダムネスを生成する手頃な道具の一つがサイコロです．実際，双六やバックギャモンなどはそれを使って遊ぶゲームですね．サイコロが昔から人々にとって身近なものであったことは，以下のような有名な言葉の中にでてくることからも想像されます．

- 「賽は投げられた (Alea jacta est)」(J. Ceasar)
- 「賀茂川の水，双六の賽，山法師，是ぞわが心にかなはぬもの」(白河法皇，平家物語)
- 「神はサイコロを振らない (Der Alte würfelt nicht)」(A. Einstein)

それでは，賽を投げながら話を進めていきましょう．

2　サイコロとランダムネス

まずは手始めにサイコロを使って様々なランダムネス (乱数) をつくる方法について考えてみましょう．サイコロはご存知のように，立方体の表面に 1 から 6 の数字 (実際は図 1 のような絵ですね) が記されていて，それを投げて上面に来た部分を「目」といいます．公平なサイコロであれば，

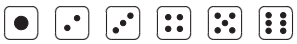

図 1: サイコロの目

1 から 6 の目が出る確率は 1/6 になる，というのは自然でしょう．もちろん歪んだサイコロだと「1 の目の方が 6 の目より出易い」ということもあるかもしれませんし，もしかしたらサイコロの穴の数や投げ方の違いが微妙に影響して公平でないかもしれません[1]．しかし，以降ではそういう問題には触れず，サイコロは歪みもなく確率 1/6 で公平に各目があらわれるものとします．

サイコロと同じく，もう一つよく使われるランダムネス発生器はコインです．公平なコインであれば表裏が等確率 1/2 で出るはずです．サッカーなどの始めにコイントスして先攻後攻を決めますね．実はコインと同じランダムネスはサイコロで再現することができます．実際，テレビの時代劇でよくでてくる丁半博打のように，サイコロの目が偶数か奇数かによってコインの表裏に相当することを決められるのです[2]．つまり，サイコロがコインを再現するという意味で「サイコロのランダムネスの方がコインのランダムネスより大きい」ということもできます．

[1]実際に，Persi Diaconis というスタンフォード大の教授は，サイコロではないですが，コインを正確に投げる機械をつくり実験をしたそうです．結果は,「まったく同じように投げられたコインは同じ方の面が出る」だそうです．機械でなく人間がやったらまったく同じ投げ方にはならないわけですが，それでも何らかのくせがあって完全には公平ではないだろうと想像されます．

[2]丁半博打では二つのサイコロを投げてその和が偶数か奇数かで丁半を定めます．

さて，今度は次のような問を考えてみましょう．

問 1. 1 から 12 を等確率 1/12 で実現したければどうすればよいでしょうか？

今度はサイコロを一回投げただけでは難しそうです．実際，サイコロ一つでは 6 つの可能性しか生みだすことができませんので，12 の可能性を等しく作り出すのは不可能なのです．ではどうすればよいでしょうか？

解 1. ちょっとインチキですが，正 12 面体のサイコロを使う．

解 2. 一つのサイコロを 2 回続けて投げれば

$$(1,1)\ (1,2)\ (1,3)\ (1,4)\ (1,5)\ (1,6)$$
$$(2,1)\ (2,2)\ (2,3)\ (2,4)\ (2,5)\ (2,6)$$
$$(3,1)\ (3,2)\ (3,3)\ (3,4)\ (3,5)\ (3,6)$$
$$(4,1)\ (4,2)\ (4,3)\ (4,4)\ (4,5)\ (4,6)$$
$$(5,1)\ (5,2)\ (5,3)\ (5,4)\ (5,5)\ (5,6)$$
$$(6,1)\ (6,2)\ (6,3)\ (6,4)\ (6,5)\ (6,6)$$

の 36 通りが等しい確率 1/36 となるので，これを 3 個ずつ 12 の組に分ける．その組に 1 から 12 の番号を対応させればよい．

解 3. コインとサイコロを一つずつ投げて，コインが表なら $a = 1$, 裏なら $a = 0$ とし，さらにサイコロの目を b として $X = 6a + b$ とおくと，X は 1 から 12 まで等確率であらわれます．

このように方法は何通りもありますし，この他にもたくさん可能性があります．それでは，以下の問題ならどうしますか？

問 2. 1 から 23 を等確率 1/23 で実現するにはどうすればよいでしょうか？

今度は少し難しいですね．解 1 の方法では正 23 面体のサイコロが必要ですが，そのようなものはありません．解 2 の方法も，23 が 36 の約数ならば組にわける方法が使えますが，23 は 36 を割り切らないのでうまくいきません．解 3 の方法では，23 ではなく 24 であれば $24 = 6 \times 2 \times 2$ なので，サイコロを一回投げて (6 通り)，さらにコインを二回投げれば (4 通り) うまくいくのですが...

それでは，このようなことはサイコロだけではできないのでしょうか？答えは「少し手間はかかるができる」です．少なくとも次のようなことが言えます．

♠ n を自然数とする．サイコロを $\lceil \log_6 n \rceil$ 個同時に投げる試行[3]を繰り返せば，$\{1, 2, \ldots, n\}$ を等確率 $1/n$ で実現することができる．ただし，$\lceil x \rceil$ は x 以上の最小の整数をあらわす[4]．

n	6	36	216	1296	7776	46656	279936	1679616	10077696	60466176	$\cdots$
$\log_6 n$	1	2	3	4	5	6	7	8	9	10	$\cdots$

表 1: $\log_6 n$

以下では，この結果をブラックボックスとして，n を代入すれば $1, 2, \ldots, n$ が一様に (等確率 $1/n$ で) 得られる乱数発生器が手に入ったということにしましょう．この乱数発生器を使えば様々な乱

[3]サイコロ 1 個を k 回投げることと，k 個のサイコロを同時に 1 回投げることは同じことなので，k 個のサイコロを投げるという言い方にします．

[4]高校数学などでは x を越えない最大の整数をあらわす記号として，しばしばガウス記号 $[x]$ を使いますが，数学の分野によってはガウス記号のかわりに $\lfloor x \rfloor$ を使うこともあります．$\lfloor x \rfloor$ は切り下げですが，♠ にあらわれた $\lceil x \rceil$ は切り上げに対応します．

数が発生させられます．例えば，1 を確率 1/8, 2 を確率 3/8, 3 を確率 1/3, 4 を確率 1/6 で乱数を発生させたければ，このブラックボックスに $n = 24$ を代入して，1 から 24 を一様に発生させて，1〜3 ならば 1，4〜12 ならば 2，13〜20 ならば 3，21〜24 ならば 4 とすればよいのです．

3　酔っ払いは家に帰れるか？

さて，双六はサイコロを振って遊ぶ古くからあるゲームですね．ここでは少し変則的な双六を考えてみましょう．

0	1	2	$\cdots$	$k-1$	k	$k+1$	$\cdots$	$n-1$	n

図 2: 酔っ払いの問題・ギャンブラーの破産問題

図 2 のような双六盤 $\{0, 1, 2, \ldots, n\}$ を使います．初め位置 k に双六の駒があり，各場所で公平なサイコロを投げて，偶数が出たら右に，奇数が出たら左に駒を動かし，左端 0 に到着するか，右端 n に到着したら「あがり」とします．このとき，以下のような問題を考えてみます．

問 3. k から駒がスタートしたとき,「あがり」までに必要な回数は平均的に何回くらいか？また，右端であがる確率はどれくらいか？

位置 0 を居酒屋，n を家とすると，自分の家の方向がわからなくて行ったり来たりする酔っ払いは家に帰れるか？という問題をモデル化したものということができます．また別の見方をすると，一回のゲームで勝つと 1 万円手に入り，負けると 1 万円失うという公平なギャンブルをするとき，n 万円で勝ち抜けする確率 (余事象を考えれば 0 円になって破産してしまう確率) はいくらか，という問題とも言えます．ですから，このモデルは「ギャンブラーの破産問題」とも言われています．

3.1　実験的観察

サイコロを持っている人は用意してください．持っていない人はコインでも構いません．

実験. $n = 6$ のときの双六盤で，以下のようなルールで双六を行なうことにします．

0	1	2	3	4	5	6

図 3: $n = 6$ のときの双六盤

- 出発点を $\{1, 2, 3, 4, 5\}$ のいずれかに決めて，サイコロを振って偶数の目 $2, 4, 6$ ならば右に，奇数の目 $1, 3, 5$ ならば左に．(コインの人ならば，コインを投げて表ならば右に，裏ならば左に．)
- 0 か 6 に到達する (あがる) まで上の操作を続けます．
- 0 か 6 に到達するまでに何回サイコロを振ったか，それから 0 と 6 のどちらに到達したかを記録して以下のような表にします．

回数	3	7	11	25	13	3	5	39	$\cdots$	7	3	5
到着点	0	6	6	0	6	6	0	0	$\cdots$	0	0	6

表 2: 100 回ゲームを行なったときの例

例えば，実験を 100 回行なって表 2 のようになったとします.
それらの平均をとって

$$\text{あがるまでの平均回数} = \frac{3+7+11+25+13+3+5+39+\cdots+7+3+5}{100}$$

$$6 \text{ であがる確率} = \frac{6 \text{ で終った回数}}{100}$$

があがりまでの平均回数と 6 であがる確率といえます.

この双六ゲームは，$S = \{0, 1, 2, 3, 4, 5, 6\}$ 上のランダムウォーク (酔歩) といい，ランダムウォークが動きうるこの集合 S のことを状態空間 (state space) といいます.

3.2 理論的考察

では，理論的にはどのような確率がでてくるか考えてみましょう．次のように問題設定をします.

問 4. 状態空間 $S = \{0, 1, 2, \ldots, n\}$ 上の k から出発したランダムウォークを考える.

(1) 初めて 0 か n に到着したときに，それが n である確率 p_k.

(2) 初めて 0 か n に到着するまでにかかる時間の期待値 t_k.

ここでは先の双六ゲームと同じように，特に $n = 6$ の場合を考えてみます．問題の性質から各確率 $p_0, p_1, \ldots, p_5, p_6$ の間には以下に見るようにシンプルな関係式が成り立つことがわかります.

1°) (境界条件) まず，6 から出発した人は初めから 0 または 6 に到着していますから試行をする必要はありません．そしてそれが 6 である確率はもちろん 1 ですから，$p_6 = 1$ です．同様にして，p_0 は既に 0 か 6 に到着しているので試行は必要ありませんが，それが 6 である確率は 0 ですので，$p_0 = 0$ です．このように，端っこ (境界) での確率がみたすべき

$$p_0 = 0, \quad p_6 = 1$$

のことを境界条件といいます.

2°) (1 次 (線形) 方程式) 例えば，p_3 を実験された方はお気付きになったと思いますが，3 から出発して一回サイコロを振ると 1/2 の確率で右 (4) か左 (2) に行きますね．そうすると，そこから先は 4 から出発して p_4 を計算する人，または 2 から出発して p_2 を計算する人とまったく同じ状況にあることがわかります．違いは初めの一回で偶数がでるか，奇数がでるかの違いだけなのです．このことから,

$$p_3 = \frac{1}{2}p_2 + \frac{1}{2}p_4$$

という関係があることがわかります．次に 4 からスタートした人について同様の考察をすると，1 歩目で 3 か 5 に移動した後は，p_3 もしくは p_5 の計算と同じになりますので,

$$p_4 = \frac{1}{2}p_3 + \frac{1}{2}p_5$$

となります.

関係式を導くのに使った考え方は，一回サイコロを投げた後は，投げる前にどこにいたかは忘れて現在の地点からあがるまでの確率を考えればよいというものです．この過去を忘れる性質はマルコフ性とよばれています.

3°) よって，1°) , 2°) の考察をまとめると確率 p_k は次のような 7 元連立 1 次方程式をみたすことがわかります.

$$
\begin{aligned}
&p_0 = 0, \quad p_6 = 1 \\
&p_1 = \frac{1}{2}p_0 + \frac{1}{2}p_2 \left(= \frac{1}{2}p_2 \right) \\
&p_2 = \frac{1}{2}p_1 + \frac{1}{2}p_3 \\
&p_3 = \frac{1}{2}p_2 + \frac{1}{2}p_4 \\
&p_4 = \frac{1}{2}p_3 + \frac{1}{2}p_5 \\
&p_5 = \frac{1}{2}p_4 + \frac{1}{2}p_6 \left(= \frac{1}{2}p_4 + \frac{1}{2} \right)
\end{aligned}
$$

この方程式を解くことはそれなりに手間がかかりますが，例えば消去法によって必ず解くことができます．実際に解いてみると，次のようなとても綺麗な解になることがわかります.

$$
p_0 = 0, \quad p_1 = \frac{1}{6}, \quad p_2 = \frac{2}{6}, \quad p_3 = \frac{3}{6}, \quad p_4 = \frac{4}{6}, \quad p_5 = \frac{5}{6}, \quad p_6 = 1
$$

上の結果から一般の n の場合にも容易に答が想像できますね．実際，次のような答えになることがわかります.

$$
p_k = \frac{k}{n}, \quad (k = 0, 1, 2, \ldots, n)
$$

このことから，n であがる確率は出発点 k と 0 との距離に比例するという著しい結果が得られました.

それでは，期待値 t_k についてはどうなるでしょうか？この場合もほぼ同様なのですが，もう少し複雑になりますので，ここでは，結果だけを書くに留めます．興味ある方は次節に解説がありますので，参照してください．答えは，

$$
t_k = k(n - k), \quad (k = 0, 1, 2, \ldots, n)
$$

となります．つまり，例えば n が偶数でちょうど真ん中 $k = n/2$ からスタートすると，

$$
t_{n/2} = \frac{n}{2}\left(n - \frac{n}{2} \right) = \frac{n^2}{4}
$$

となります．ギャンブラーの問題だとすれば，初めの所持金を $n/2$ とすると，ちょうどその 2 乗くらいの時間の間ゲームを楽しむことができ，ゲームが終ったときには 1/2 の確率で擦っているか，1/2 の確率で所持金が倍になっています.

$n = 6$ の双六のときには，表 3 にあるように，3 から出発すると平均的に 9 回であがるということです．もちろん，これは平均的に 9 回ということですから，3 回であがりのこともありますし，39 回かかってしまうかもしれません.

出発点	0	1	2	3	4	5	6
平均回数	0	5	8	9	8	5	0

表 3: $n = 6$ のときのあがりまでの平均回数

3.3 期待値 t_k の計算

n であがる確率のときと同様の考察により，k から出発したときのあがりの回数の平均 t_k は次の方程式をみたすことがわかります．

$$\begin{aligned} t_0 &= t_n = 0, \\ t_k &= \frac{1}{2}(t_{k+1} + 1) + \frac{1}{2}(t_{k-1} + 1) \\ &= \frac{1}{2}t_{k+1} + \frac{1}{2}t_{k-1} + 1 \quad (k = 1, 2, \ldots, n-1) \end{aligned}$$

この連立方程式は色々な解き方があると思いますが，少しかわった考え方をしてみましょう．まず，解の候補 $t_k = k(n-k)$ が何らかの方法で見つかったとしましょう[5]．一旦答えが見えてしまえばこの方程式の解となっていることは簡単に確かめられますが，それで終りではありません．実際，2 次方程式 $x^2 = 1$ の解 $x = 1$ を見つけておしまいと喜んではいけないのと同様です．しかし，もし別口でこの方程式が一つしか解を持たないということを示すことができれば，解が一つ見つかっているのでおしまいです．解の一意性といいます．

それではちょっと高級な考え方ではありますが，解の一意性を示してみましょう．もしも，同じ方程式と境界条件をみたす二つの解 u_k, v_k があったとします．その二つの方程式を並べて

$$\begin{aligned} u_0 = u_n = 0, \quad u_k &= \frac{1}{2}u_{k+1} + \frac{1}{2}u_{k-1} + 1 \quad (k = 1, 2, \ldots, n-1) \\ v_0 = v_n = 0, \quad v_k &= \frac{1}{2}v_{k+1} + \frac{1}{2}v_{k-1} + 1 \quad (k = 1, 2, \ldots, n-1) \end{aligned}$$

両辺引き算して，$f_k = u_k - v_k$ とおいて f_k の方程式に直すと，

$$\begin{aligned} f_0 &= f_n = 0, \\ f_k &= \frac{1}{2}f_{k+1} + \frac{1}{2}f_{k-1} \quad (k = 1, 2, \ldots, n-1) \end{aligned}$$

をみたします．さて，2 行目の両辺を少し式変形すると

$$f_k - f_{k-1} = f_{k+1} - f_k \quad (k = 1, 2, \ldots, n-1)$$

となりますので全部並べて書くと，

$$f_1 - f_0 = f_2 - f_1 = f_3 - f_2 = \cdots = f_{n-1} - f_{n-2} = f_n - f_{n-1} =: C$$

となり，共通の値を C とします．$f_k - f_{k-1} = C$ をすべての $k = 1, 2, \ldots, n-1$ について和をとると

$$f_n - f_0 = nC$$

[5] この解がどうやって見つかったかは問わないでください．試行錯誤して見つかったかもしれませんし，一瞬の閃きでわかったかもしれません．とにかく，理由はともあれ見つかったとします．

が得られます．よって，境界条件 $f_0 = f_n = 0$ より自動的に $C = 0$ となりますが，これは，$f_0 = f_1 = \cdots = f_n = 0$ を意味します．もとの u_k, v_k の言葉に直すと，

$$u_0 = v_0 = 0,\ u_n = v_n = 0,\ u_k = v_k\ (k = 1, \ldots, n-1)$$

を意味しますので，二つの解は完全に一致しなければなりません．これで解が二つ存在したとしても実は同じものでしかありえない，つまり解が一意 (唯一) であることがわかりました．

4　連立方程式を解いてみる

4.1　グラフの上のランダムウォーク

さて，ランダムウォークのモデルの一般化を考えてみましょう．さきほどは酔っ払いが一直線上しか動けませんでしたが，例えば京都の街は格子状ですし，放射状になっている街もあるでしょう．そこで街の形を簡略化したモデルであるグラフというものを考えてみることにします．ここでいうグラフは，放物線 $y = x^2$ をあらわすグラフと言うときのグラフとは違うものです．まずは例を見てみるのがわかりやすいでしょう．

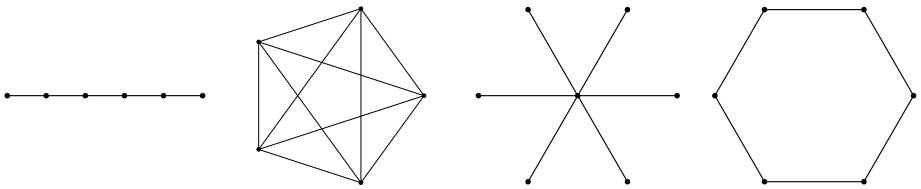

図 4: グラフの例

図 4 の左から順にパスグラフ (P_6)，完全グラフ (K_5)，星グラフ ($K_{1,6}$)，サイクルグラフ (C_6) などといいます．グラフとは図のように，点 (vertex) の集合 V と辺 (edge) の集合 E からなり，しばしば，$G = (V, E)$ のように書きます．例えば，図 5 のグラフであれば，

$$V = \{0, 1, 2, 3, 4, 5, 6\},\ E = \{01, 12, 15, 23, 34, 45, 46\}$$

となります．辺は向きをもたないとしますので，12 と 21 は同じものとみなして，代表して 12 だけを E に加えています．

各点 v から出る辺の本数を次数 (degree) といい，$\deg(v)$ とあらわします．例えば，パスグラフ (P_6) では，左端の点から $v_0, v_1, \ldots, v_5$ とすれば

$$\deg(v_0) = \deg(v_5) = 1, \quad \deg(v_1) = \deg(v_2) = \deg(v_3) = \deg(v_4) = 2$$

となります．完全グラフ (K_5) ではすべての点 v で $\deg(v) = 4$ となっています．

また，グラフ $G = (V, E)$ において，V の部分集合 ∂ を特別視して境界とみなす (呼ぶ) とき，境界付きグラフ (G, ∂) とあらわします．例えば，図 5 のグラフで

$$V = \{0, 1, 2, 3, 4, 5, 6\},\ E = \{01, 12, 15, 23, 34, 45, 46\},\ \partial = \{0, 6\}$$

と 0 と 6 を特別視したものは境界付きグラフ (G, ∂) です．境界 ∂ を指定しても，見かけ上は何も変化はありませんが，考えているグラフを双六盤と思えば「あがり」の場所を指定していると思ってください．

定義. グラフ $G=(V,E)$ 上のランダムウォークとは，以下の操作を繰り返し行なう確率モデルのことをいう．

> 点 v にランダムウォーカーがいるとき，$\frac{1}{\deg(v)}$ の確率で v から出る辺を一つ選び，その v とは異なる方の点にジャンプする

さらに，境界付きグラフ (G,∂) の上のランダムウォークとは，境界点でない点から出発して，上の操作を繰り返して行ない，境界点のいずれかに到着したらそこで終了するものとする．

さて，図 5 の境界付きグラフの上のランダムウォークについて，以前と同様の問題を考えてみましょう．0 と 6 を境界として，k から出発して 6 に先に到着する確率 p_k を計算してみます．方程

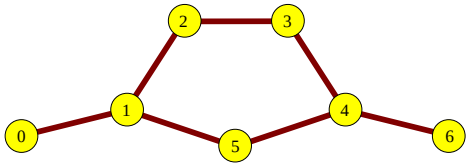

図 5: 境界付きグラフ．$\partial=\{0,6\}$

式は前と同様の議論によって，

$$
\begin{aligned}
&p_0=0,\quad p_6=1\\
&p_1=\frac{1}{3}p_0+\frac{1}{3}p_2+\frac{1}{3}p_5\\
&p_2=\frac{1}{2}p_1+\frac{1}{2}p_3\\
&p_3=\frac{1}{2}p_2+\frac{1}{2}p_4\\
&p_4=\frac{1}{3}p_3+\frac{1}{3}p_5+\frac{1}{3}p_6\\
&p_5=\frac{1}{2}p_1+\frac{1}{2}p_4
\end{aligned}
$$

となり，やはり連立一次方程式ですから，消去法を用いれば必ず解くことができます．結果だけを一応書いておきましょう．

$$p_0=0,\ p_1=\frac{5}{16},\ p_2=\frac{7}{16},\ p_3=\frac{9}{16},\ p_4=\frac{11}{16},\ p_5=\frac{1}{2},\ p_6=1.$$

ここまでの説明で，ランダムウォークの境界への到達確率と連立 1 次方程式は密接に関係していることがわかっていただけたでしょうか？

問 5. 他の境界付きグラフ上でランダムウォークを考えて，同様の計算をしてみよ．

4.2 連立方程式をサイコロで解いてみる

以下の問を考えてみましょう．

問 6. 以下の 9 元連立 1 次方程式を解け.

$$\begin{cases} 4a = 1+b+d & 4b = a+c+e & 4c = 1+b+f \\ 4d = 1+a+e+g & 4e = b+d+f+h & 4f = 1+c+e+i \\ 4g = 2+d+h & 4h = e+g+i & 4i = 1+f+h \end{cases}$$

少し時間はかかりますが，消去法で解くと答えは

$$a = \frac{17}{32},\ b = \frac{45}{112},\ c = \frac{115}{224},\ d = \frac{81}{112},\ e = \frac{9}{16},\ f = \frac{73}{112},\ g = \frac{179}{224},\ h = \frac{53}{112},\ i = \frac{17}{32}$$

と得られるわけですが，この節ではサイコロでこの連立方程式を解くことを考えてみましょう.

前節までの考察で，連立 1 次方程式はランダムウォークと関係があるということがわかりました．そのことを考慮すると，実は問 6 の連立方程式を解くことは図 6 のような境界付グラフの上のランダムウォークの境界への到達確率を求めることに帰着されます．ただし，0 と 1 が書いてある点が境界に対応します.

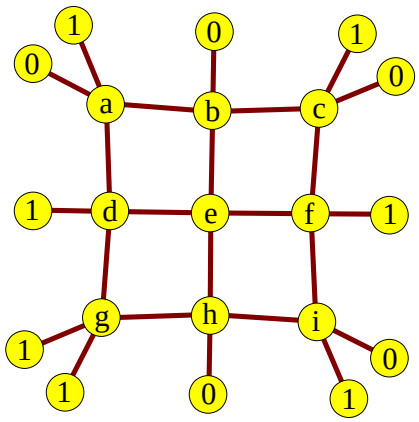

図 6: 対応する境界付きグラフ (双六盤)

出発点 k を決めて境界付きグラフ上のランダムウォークを 100 回発生させ，境界に到着したときに 1 であったものの回数を数えて 100 で割ったものをここでは k の酔歩解とよびます．各出発点 k に対して酔歩解を求めて表 4 に示し厳密解と比較しています． 表からもわかるように，もち

	a	b	c	d	e	f	g	h	i
厳密解	0.53125	0.40178	0.51339	0.72321	0.5625	0.65178	0.79910	0.47321	0.53125
酔歩解	0.54	0.39	0.49	0.75	0.54	0.68	0.78	0.46	0.52

表 4: 厳密解と酔歩解 (厳密解は下 5 桁まで)

ろん，酔歩解は手計算でやった厳密な解とは一致しない近似的な解ですし，計算のたびに変化します．しかし，こんな単純な方法でそれなりに近い値が得られていることは驚くべきことではないでしょうか？

なぜこのような厳密な解が求まる問題で近似的な解を求める必要があるのでしょうか？もちろん問 6 ぐらいの問題ならば厳密に解いてしまえばいいでしょう．しかし，例えば，ある研究の過程でどうしても連立 10000000000 次方程式を解く必要があったとします．それも特にその中で知りたいのは 10000000000 個の変数のうちの一つの変数 v についてだけだったとしましょう．このときに，膨大な連立方程式を厳密に解いてすべての解を求めた後で v を知ることと，近似的な方法で v の値を求めるのはどちらが速いでしょうか？どちらも実際にはそれなりに時間がかかるかもしれませんが，状況によっては「とにかく厳密ならばいい」というものでもなさそうだと感じていただければ十分です．このように，解くのにきわめて時間がかかる大規模な問題や厳密に解くことができない問題をランダムネスを使って近似的に解く方法は，しばしば「モンテカルロ法」と呼ばれ様々な問題に応用されています．モンテカルロとはカジノで有名なモナコにある地区の名前だということはご存知だと思いますが，1940 年代にロスアラモス研究所で原爆開発にこのようなランダムネスを用いる方法を使った科学者達がつけた名前だそうです．最近では，モンテカルロ法を取り入れたコンピューター囲碁のソフトもあるそうです．

4.3 連立方程式を電気回路で解いてみる

さて，ここでがらりと話を変えて電気回路の話をしましょう．図 7 のような直列の電気回路を考えて各点での電圧を計算してみます．各抵抗はすべて等しく r とし，0 での電圧は 0，6 での電圧は 1 とします．電圧の計算にはオームの法則とキルヒホッフの法則を使います．隣りあう点 k と

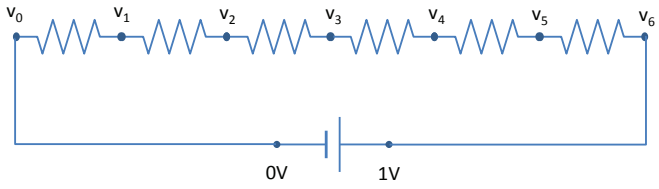

図 7: (直列) 電気回路

$k+1$ の間を流れる電流 I_k は k によらず一定値 i です．小難しく言えば点 k に流入する電流と k から流出する電流は等しいというキルヒホッフの法則です．また，二点間の電圧の差 V は二点間の抵抗 R と間を流れる電流 I の積で与えられるというオームの法則 $V = RI$ を考慮すると，以下の方程式が立ちます．

$$v_0 = 0, \quad v_6 = 1$$
$$v_6 - v_5 = v_5 - v_4 = v_4 - v_3 = v_3 - v_2 = v_2 - v_1 = v_1 - v_0 = ri$$

例えば，2 行目の左端の方程式 $v_6 - v_5 = v_5 - v_4$ を v_5 について整理すると

$$v_5 = \frac{1}{2}v_4 + \frac{1}{2}v_6$$

となります．同じ変形を 2 行目の左から 5 個の等号についてすべて行なう (つまり ri を消去する) と，

$$
\begin{aligned}
&v_0 = 0, \quad v_6 = 1 \\
&v_1 = \frac{1}{2}v_0 + \frac{1}{2}v_2 \left(= \frac{1}{2}v_2 \right) \\
&v_2 = \frac{1}{2}v_1 + \frac{1}{2}v_3 \\
&v_3 = \frac{1}{2}v_2 + \frac{1}{2}v_4 \\
&v_4 = \frac{1}{2}v_3 + \frac{1}{2}v_5 \\
&v_5 = \frac{1}{2}v_4 + \frac{1}{2}v_6 \left(= \frac{1}{2}v_4 + \frac{1}{2} \right)
\end{aligned}
$$

を得ます．お気づきと思いますが，これは $n = 6$ のときの双六ゲームで得られた連立方程式とまったく同じものです．よって，各電圧 v_k はランダムウォークのときに得られた確率 p_k とまったく一致し，

$$v_0 = 0, \quad v_1 = \frac{1}{6}, \quad v_2 = \frac{2}{6}, \quad v_3 = \frac{3}{6}, \quad v_4 = \frac{4}{6}, \quad v_5 = \frac{5}{6}, \quad v_6 = 1$$

となります．実はこの事実を一般化して次の定理が成り立つことが知られています．

定理 1. (G, ∂) を境界付きグラフとして，境界 ∂ は V_1 と V_0 の二つの部分に分けられているとする．このとき，G の点 x から出発したランダムウォークが初めて ∂ に到達したときにそれが V_1 の点である確率を p_x とする．このとき，p_x はグラフ G の各辺を同じ抵抗として，V_1 の各点で 1 ボルト，V_0 の各点をアースした (0 ボルト) 電気回路の点 x での電圧 v_x に等しい．

この定理によれば，問 6 の連立方程式の解を得るためには，図 6 の境界付きグラフに対応する電気回路 (図 8) を実際につくって，電流を流して各点の電圧を測ればよいことがわかります．これ

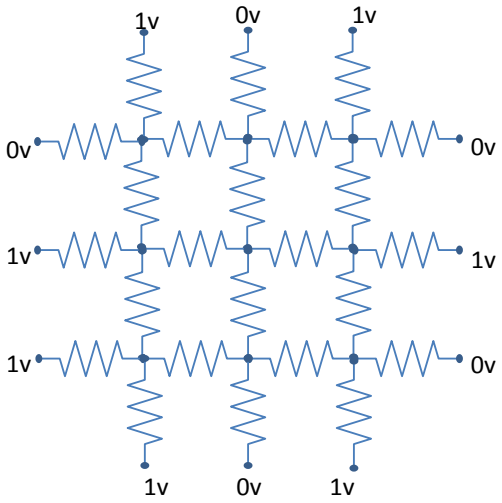

図 8: 図 6 に対応する電気回路

ならば一瞬にしてすべての電圧，つまり，ランダムウォークの到達確率が物理的誤差を除けばわかるわけですから，実際に電気回路をつくれる程度のサイズのグラフのときにはとても強力な方法であることがわかると思います．自然は連立方程式の答えを知っているわけです．

5　マルコフ連鎖とその例

前節では，グラフ上のランダムウォークを考察しました．これをもう少しだけ一般化したマルコフ連鎖というものを考えます．マルコフ連鎖とは大雑把に言えば，状態空間とよばれる可能な状態の集合と，ある状態から別の状態へ移る確率のルールが与えられた確率的モデルのことをいいます．例で考えるのがわかりやすいので，いくつか例をあげます．

5.1　ランダムナイトムーブ

チェスボードは 8×8 のマス目になっており，ナイトは八方桂馬とよばれるように将棋の桂馬と同様の動きをしますが，将棋と違って横にも後にも桂馬跳びができます．ランダムナイトムーブ

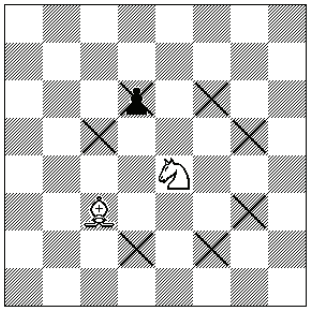

図 9: ナイトの動き方

とは，あるマス目にナイトがあるときに次の一手を可能なムーブの中から等確率で選ぶことによって手を進めていく確率モデル (マルコフ連鎖) のことをいいます．64 個のマス目がナイトの位置をあらわしますので，その 64 個のマス目がここでは状態空間となります．つまり，ランダムナイトムーブによってナイトの位置 (状態) がランダムに時々刻々とかわっていくわけです．ちょっとかわった酔っ払いということもできます．

問 7. ある点から出発したナイトは，すべての点に到達することができるか？

5.2　15 パズル

ランダムナイトムーブはある意味で変形双六とみなすことができますが，次はもう少しわかりにくい例を考えてみましょう．

図 10 のような 4×4 のマス目の中に 1 から 15 のピースがはいっています．一つだけある空白を利用してピースを動かしていき，

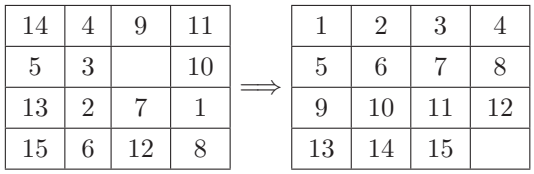

図 10: 15 パズルの一つの状態と最終状態

左のばらばらになった状態から右の整列した状態に持っていくというゲームです[6]．これはパズルですから普通は頭を使ってやるものですが，横着な人はサイコロを使ってやることもできます．つまり，各状態において動かせるピースが何種類かありますので，それらの中から一つのピースを等確率で選んで空白の位置に動かします．それを繰り返すことによって得られるのがマルコフ連鎖です．例えば，図 10 の左図では $3, 7, 9, 10$ が動かせますので，等確率 $1/4$ でどれかのピースを選んで空白の位置に移動します．

先程のランダムナイトムーブの例では可能な状態数はたったの 64 でしたが，このパズルの状態数は $16! = 20922789888000$ 通りという膨大な数になります．ですからすべての可能なパターンを手で紙の上に描くなんてことは諦めなければなりませんし，コンピューターを使っても到底無理です．

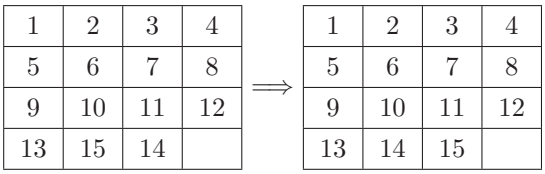

図 11: これは解けるでしょうか？

問 8. 図 11 の左図から出発して右図に到達させよ．

5.3 トランプのシャッフリング

買ったばかりのトランプのカードは $\spadesuit A$ から $\diamondsuit K$ までの 52 枚のカードがきれいに整列しています．トランプゲームを始める前に普通はシャッフルしてばらばらな順序にしますね．これをマルコフ連鎖としてモデル化してみましょう．よく使われるシャッフルはリフルシャッフルといい，カードの山を二つの山に分けて，それらをパラパラっと交互に挟みこみ下に落していくというというものです．ここでは説明を簡単にするために，もっとも原始的な (普通はこんなシャッフリングは絶対に使いませんが) 方法について述べます．

まず，一番上のカードを手に取り，残った 51 枚のカードのどこかに $1/52$ の確率で差込みます．$1/52$ の確率で一番上に置いてしまうと元の状態と同じということになります．簡単のために，52 枚のカードには図柄のかわりにそれぞれ 1 から 52 の番号が書いてあると思って，図 12 を参照してください．

このマルコフ連鎖の状態空間は 52 枚のカードの可能な配列の全体ですから，何と

$$52! = 80658175170943878571660636856403766975289505440883277824000000000000$$

[6] `http://www.afsgames.com/15puzzle.htm` にいくと実際にゲームができます．

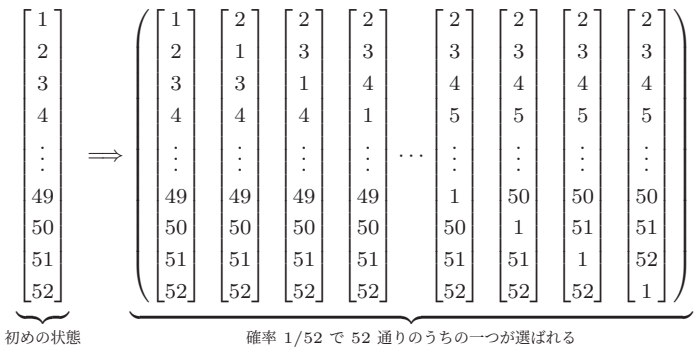

図 12: 1 ステップでの推移の様子

通りあります．さきほどの 15 パズルに比べても膨大な数ですね．このマルコフ連鎖を実際に実行した例が図 13 です．簡単のために，1 から 10 までが書いてある 10 枚のカードで同様の実験をしています．ちなみにカード 10 枚のときの状態空間の要素数は $10! = 3628800$ です．

図 13 では何回目くらいでばらばらになったように見えるでしょうか？n 枚のカードのとき，理論的には $n \log n$ 回くらいで「ばらばら」になることが知られています[7]．$n = 10$ のときには 23.0529 回，$n = 52$ のときには 205.465 回ということになります．ということで，この方法でトランプをシャッフリングする人はまずいないでしょう...

6　サイコロで暗号を解くには？

最後に，この節ではある種の暗号をランダムネスを用いて解くことを考えてみましょう．まず，初めにこの講義の内容紹介で挙げていた暗号について，どのような暗号化を行なったかを明らかにしておきましょう．表 5 は 1 ページの概要にあるものと同じ文章の暗号化ですが，異なる鍵で暗号化していますので違ったものに見えます．

```
fyeaxjqfzjoxceddedcjbujcxbjhxgpjbegxrjukjzeb
bedcjopjsxgjzezbxgjudjbsxjofdljjfdrjukjsfhed
cjdubsedcjbujrujudaxjugjbqeaxjzsxjsfrjnxxnxr
jedbujbsxjouuljsxgjzezbxgjqfzjgxfredcjjotbje
bjsfrjdujneabtgxzjugjaudhxgzfbeudzjedjebjjfd
rjqsfbjezjbsxjtzxjukjfjouuljjjbsutcsbjfyeaxj
jqebsutbjneabtgxzjugjaudhxgzfbeudjj
```

表 5: ある小説の冒頭の暗号化

6.1　シーザー暗号

シーザー (J. Ceaser) は「賽は投げられた」と言ったとされる古代ローマの指導者でしたが，暗号でも有名です．シーザー暗号とはもっとも原始的な暗号の一つですが，アルファベット $a, b, c, d, \ldots, x, y, z$

[7] もちろん，数学的には「ばらばら」の意味をはっきりさせる必要がありますが，ここでは省略します．

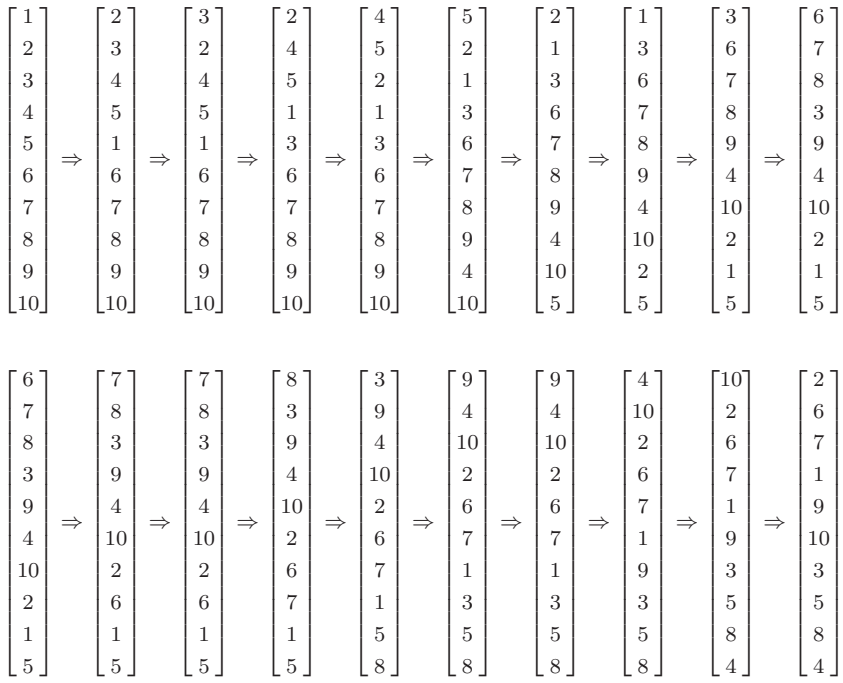

図 13: 10 枚のカードのシャッフリングの実験例

を何文字かずつずらして別のアルファベットを対応させて文章を暗号化したと言われています．つまり，$a \to d,\ b \to e,\ c \to f, \ldots, x \to a,\ y \to b, z \to c$ のようにです．表にすると以下のようになります．

$\sigma =$

Before	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z
After	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	a	b	c

表 6: シーザー暗号の鍵 σ

この規則に従って this is a pen を暗号鍵 σ によって暗号化すると，

$$\text{this is a pen} \Longrightarrow \text{wklv lv d shq}$$

となります．このずらし方をあらわす対応表を暗号鍵といい，σ とあらわすことにします．この σ を用いて

$$\sigma(\text{this is a pen}) = \text{wklv lv d shq} \quad \text{または} \quad \text{wklv lv d shq} = \sigma(\text{this is a pen})$$

と数式のようにあらわすことにしましょう．逆に復号するには，暗号鍵と逆方向にずらす鍵を持ってくればよいわけです．そのような鍵を復号鍵といい，σ^{-1} とあらわします． 真の文章 $\mathbf{x}$ を暗号

$\sigma^{-1} =$

Before	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z
After	x	y	z	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w

表 7: 復号鍵 σ^{-1}

鍵 σ によって暗号化した文章が $\mathbf{y} = \sigma(\mathbf{x})$ で，それに復号鍵を施すと

$$\sigma^{-1}(\mathbf{y}) = \sigma^{-1}(\sigma(\mathbf{x})) = \mathbf{x}$$

となって復号されるわけです．実際にやってみると

$$\sigma^{-1}(\text{wklv lv d shq}) = \text{this is a pen}$$

となります．あらゆる文章を集めた集合を Z とあらわすと，暗号鍵 σ も復号鍵も Z から Z への写像で，σ と σ^{-1} は逆写像の関係にあると言えます．

容易にわかるように，シーザー暗号は 26 通りのずらし方しかバリエーションがありませんので[8]，暗号 “wklv lv d shq” を解読せよと言われたとき，もしもシーザー暗号であることをあらかじめ知ることができたとすると，26 通りすべての鍵 η について

$$\eta(\text{wklv lv d shq})$$

を試してみれば，そのうちのどれかが復号鍵 σ^{-1} のはずですから，復号が可能だと言えるわけです．その意味でシーザー暗号の発明は画期的ではありましたが，暗号としては弱いと言えるでしょう．

6.2 置換暗号

シーザー暗号をもっと複雑化した暗号を考えてみましょう．さきほどのシーザー暗号では，a から z をいくつか巡回的にずらしてできたアルファベット列を a から z の下に置いて対応表をつくって鍵としたわけです．ここでは，巡回的にずらしたものだけではなく，任意の並び方を使ってよいということにしましょう．さらに，文章中のピリオド・コンマ・空白などアルファベット以外の記号はすべて記号 $\diamond$ で代用しています．このような暗号化をここでは置換暗号と呼ぶことにします．表 5 の暗号は置換暗号によって暗号化しています．ちなみに表 8 の σ がその暗号鍵です．もちろん暗号鍵がわかっていればその復号鍵である σ^{-1} はわかりますから，表 8 の復号鍵 σ^{-1} で表 5 は解読可能です．

$\sigma =$

Before	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	$\diamond$
After	f	o	a	r	x	k	c	s	e	v	l	y	m	d	u	n	w	g	z	b	t	h	q	i	p	$\diamond$	j

$\sigma^{-1} =$

Before	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	$\diamond$
After	c	t	g	n	i	a	r	v	x	$\diamond$	f	k	m	p	b	y	w	d	h	u	o	j	q	e	l	s	z

表 8: 暗号鍵 σ と復号鍵 σ^{-1}

さて，ここでは次の問題を考えることにしましょう．

問 9. ある文章が置換暗号を用いて暗号化されているとわかっているとします．もちろん鍵 σ はわかっていません．このとき，暗号解読ができるでしょうか？

この暗号化ではシーザー暗号と違って $27! = 10888869450418352160768000000$ 通りの鍵の種類があるので，総当たりで調べるということは事実上不可能です．それでは，この暗号は安全なのでしょうか？

結論から先に言うと，これから紹介する確率的な方法で暗号が「ほぼ」解けてしまうことがあります．ですから，置換暗号はシーザー暗号に比べて鍵の可能性が膨大で総当たりは無理なのですが，だからと言って安全ともいえないわけです．

[8] 1 通りは何もずらさないことに対応しますので，25 通りと言った方がよいかもしれません

6.3 てっぺんをさがす

準備として，一旦，暗号の話から離れて次のような問題をマルコフ連鎖で解くことを考えてみましょう．

問 10. S をグラフとして，$f : S \to \mathbb{R}$ をある関数とする．このとき，f が最大値 (もしくは最小値) を与える点を見つけよ．

この問題では，グラフ S の点 x に人が居るとして，その人は x に隣りあう点 y の f の値 $f(y)$ は観測可能であると仮定しておきます．例えば，ある地域にガソリンスタンドが何軒かあって，それをグラフの点だと思います．今いるガソリンスタンド x からすぐ近くのガソリンスタンド y でのガソリンの値段 $f(y)$ は双眼鏡を使うなどしてわかるとします．このとき，f の最小値を探す問題は，なるべく安いガソリンスタンドを捜す問題と同等になります．ここで，S は有限ですが，例えば 27! 個の暗号鍵の全体のように巨大な集合を想定しています．ですから，すべての可能性を全部探索していくことは到底不可能な状況にあります．ガソリンスタンドの問題で言えば，箱崎近辺のガソリンスタンドなら虱潰しもできますが，日本全国となるととても無理ですね．

そこで，ここでは近似的な最大値 (のもしくは最小値) を見つけるために状態空間 S 上のマルコフ連鎖を考えます．関数 f は一般には負の値を持ってもいいのですが，簡単のために，適当に底あげすることによって非負の値を持っているとします．次のようなマルコフ連鎖を考えます．

> 点 x から隣りあう点を等確率で一つ選ぶ．例えば，y が選ばれたとすると，その点の $f(y)$ の値を観測して
> (i) $f(x) < f(y)$ ならば y に進む．
> (ii) $f(x) \geq f(y)$ ならば，$p_{x,y} := \frac{f(y)}{f(x)} \leq 1$ の確率で y に進み，$1 - p_{x,y}$ の確率で x に留まる．

このマルコフ連鎖を続けていくと，f の値の大きい方に進んで行きそうなことは予想されます．

さて，この方法を聞いてどうしてランダムにする必要があるのかと疑問に思われた方はおられないでしょうか？ある点 x から出発して，その点の周りをみて関数の値の一番大きい方向に動く，次の点でも周りの関数の値を調べて一番大きい値の方向に動く... とこの操作を繰り返していくと最大値を発見できるのではないかと思われませんか？実はこの操作を繰り返していくうちに f の極大点 x に到着してしまうと，x の周りの f の値は $f(x)$ より値が小さいものばかりなので (極大の定義)，その点から動くことはできなくなってしまいます．しかし，極大値が関数の最大値とは限りませんから，止まった点がてっぺんであると言いきるのは少し安易すぎます．

このような状況においてランダムネスの効果があらわれるのです．マルコフ連鎖の定義から，極大点 x でまわりの f の値が $f(x)$ よりも小さいとしても (case (ii))，小さい確率で極大点を抜けだすことができるのです．そのわずかな極大から脱出する確率が別の極大を探すために大きな役割を果すことになります．

6.4 統計と暗号解読

以前，シェイクスピアの戯曲が実はフランシスコ・ベーコンによるものではないかという説があったそうですが,「計量文献学」的な統計手法を用いて，そうではないと結論を出した学者がいました．この手法やその結論が正しいかどうかという問題はここでは脇に置いて，その考え方のエッセンスを説明しましょう．まず，シェイクスピアのものだとわかっている文献と，ベーコンの文献

のそれぞれの単語の統計をとって，例えば，出てくる単語の長さの平均値を比較します．もしその値に有意に違いがあれば，それは違う人物の手になるものと判定していいのではないか，という考え方です．もっと，詳しく品詞や個々の単語の出現頻度，単語のつながり方など詳しく見ていけばもっと精度はあがるはずだ，と考えるのは自然でしょう．贋作を発見するためにもこのような統計的手法はしばしば使われているようです．

少し話題はそれますが，筒井康隆氏の書いた『残像に口紅』をという小説があります．各章ごとに少しずつ使ってよい「かな」を減らしてどれくらい文章が書けるかという，ある種の実験小説です．使えるかなが減っていくに従って文章の豊富さが失なわれていくはずですが，何も知らずにその文章を読むとどの文字が使われていないかはすぐにはわからないかもしれません．文庫本の後書きには，当時東京女子大の水谷静夫教授の学生であった泉麻子氏がその文章の統計的な考察をして卒業論文を書いたという話が紹介されています[9]．

さて，また話は少しかわりますが，俳句は 50 種類の音を 17 文字並べたものに過ぎないですから (濁点や字余りなどを考えなければ) 全部で

$$50^{17} = 762939453125000000000000000000$$

通りの可能性があります．これが地球上にあるすべての俳句です．この数を「こんなにたくさん」と考えるか,「高々これだけ」と考えるかは人それぞれでしょう．いずれにしても人間が捻り出す俳句，例えば (濁点を除いて)

なつくさや つはものともか ゆめのあと

はこの中の一つなのです．それでは，サイコロはどのような俳句を捻りだすでしょうか？50 音からサイコロを使ってランダムに 17 文字を発生させてみると，

れいしてえ をのさようよみ るらひのね

のような俳句ができました．サイコロで意味の通じる俳句を作ることは至難のわざのようです．これはこれでリズムはいい感じですが (五七五のリズムの力？)，やはり NHK 俳句に投稿すると酷評されるでしょうね．

それでは何故，ランダムに作った俳句は普通の俳句と違うように感じるのでしょうか？一つは日本語に普通にあらわれる 50 音のつながり方と違うつながり方が多く出てくるからだと考えられます．例えば最後の「るらひ」なるつながり方は日本語では見かけません．その結果，サイコロの詠んだ俳句は人の詠んだ俳句とは違う響きを持っているように感じられるわけです．この何かが違うという感覚を数量化することができれば，暗号解読に使えるのではないかと考えるのは自然ではないでしょうか．

ここまで俳句を例にした考察は英語の文章でもまったく同様にあてはまります．日本語に比べると英語は 26 個のアルファベットとコンマやピリオドなどの記号だけで文章が構成されていて構造がとてもシンプルなので，以降は英語の文章を考察していくことにします．ここでは，英語の文章が自然な文章かどうかを判定するために，アルファベットのつながり方の統計を使うことにします．

1°) 英語の標準的な文章を一つ選んで一列にアルファベットが並んでいるとして，さらに，大文字・小文字の区別をなくし，空白，?, ! などの記号はすべて記号 $\diamond$ とみなします．

2°) 次に 2 文字のアルファベットのつながり方に関して統計をとります．例えば this is a pen. ならば，2 文字ずつ前から

[9] その論文によれば，5 箇所ほど使ってはいけないはずの「かな」が使われたところがあったそうです...

	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	◇
a	0	10	13	28	0	0	11	0	28	0	10	58	21	65	0	4	4	32	23	54	0	9	4	6	20	0	36
b	16	2	0	0	10	0	0	0	10	0	0	12	0	0	12	0	0	2	2	0	20	0	0	0	9	0	3
c	23	0	2	0	22	0	0	15	6	0	9	10	0	0	48	0	2	6	0	17	6	0	0	0	0	0	4
d	14	1	0	4	34	0	4	0	24	0	0	0	2	0	13	0	0	7	3	0	4	4	0	0	1	0	123
e	30	2	17	46	16	7	2	0	4	0	4	12	9	41	1	7	2	96	64	18	0	7	7	2	10	0	228
f	4	0	0	0	6	4	0	0	17	0	0	10	0	0	36	0	0	13	0	0	2	0	0	0	0	0	42
g	4	0	0	0	21	0	0	18	4	0	0	2	0	4	4	0	0	14	5	0	0	0	1	0	0	0	47
h	36	0	0	2	122	0	0	0	33	0	0	2	0	0	13	0	0	2	0	12	4	0	0	0	2	0	38
i	6	4	32	27	17	6	16	0	0	0	2	36	9	114	28	1	0	12	57	50	0	12	0	0	0	2	8
j	2	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	1	0	0	0	0	0	0
k	0	0	0	0	10	0	0	0	7	0	0	0	0	2	0	0	0	0	4	0	0	0	0	0	0	0	17
l	24	4	0	11	40	0	0	0	27	0	0	51	1	1	24	0	0	0	12	4	10	0	2	0	9	0	44
m	28	2	0	1	25	0	0	0	23	0	0	0	5	0	14	8	0	12	4	0	2	0	0	0	4	0	15
n	13	0	12	55	32	2	63	0	18	0	4	2	2	2	19	0	0	0	21	45	0	4	2	0	1	0	69
o	4	11	15	0	1	44	4	0	2	0	9	6	22	95	9	8	0	66	10	20	36	11	12	0	0	2	38
p	19	0	0	0	17	0	0	4	14	0	0	11	0	0	12	6	0	34	0	0	0	0	0	0	2	0	7
q	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0
r	35	0	1	7	87	2	4	0	31	0	0	0	6	8	27	2	0	4	36	8	4	2	2	0	5	0	84
s	15	0	8	4	38	0	0	19	30	0	0	5	3	1	21	15	0	0	19	28	12	0	0	0	2	0	165
t	21	0	0	0	43	0	0	134	38	0	0	3	0	0	32	0	0	11	20	11	8	0	2	0	8	0	99
u	0	6	2	12	7	0	4	0	4	0	0	12	4	13	0	11	0	13	21	16	0	0	0	0	0	0	4
v	8	0	0	0	30	0	0	0	14	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	2	0	0
w	25	0	0	0	17	0	0	23	34	0	0	0	0	4	9	0	0	0	4	0	0	0	0	0	0	0	5
x	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	4
y	0	0	0	0	10	0	0	0	0	0	0	0	4	0	2	0	0	0	3	2	0	0	0	0	0	0	62
z	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
◇	108	56	68	41	23	67	16	53	67	5	2	32	55	22	98	60	0	29	73	147	16	6	89	0	8	0	107

表 9: ニューヨークタイムズの記事から取った統計

(t, h), (h, i), (i, s), (s, ◇), (◇, i), (i, s), (s, ◇), (◇, a), (a, ◇), (◇, p), (p, e), (e, n), (n, ◇)

のように読んでいって，それぞれのアルファベットの後にはどのアルファベットが出現しやすいかの統計をとって 27×27 の行列にあらわします．表 9 はニューヨークタイムズ紙のある記事を使って統計を取りました．例えば，表 9 の行列の m 行 i 列の 23 は，元にした文章の中で m の次に i が 23 回あらわれたことを意味します．また，f 行 v 列は 0 ですが，これは fv と続く単語がその記事の中にはないことを意味します．

さて，ここからは少し難しいですが説明を試みます．表 9 に与えられた 27×27 行列を M，$\mathbf{x} = x_1 x_2 \dots x_n$ を長さ n のオリジナルの文章，真の暗号鍵を σ とします．つまり，暗号化された文章は $\mathbf{y} = \sigma(\mathbf{x})$ です．以後，もちろん $\mathbf{x}$ と暗号鍵 σ は未知とし，暗号化された文章 $\mathbf{y}$ のみが表 5 のように与えられているとします．

3°）考えるマルコフ連鎖の状態空間 S は可能な鍵 η の全体とします．つまり，$|S| = 27!$ の巨大な集合です．

4°）状態 η から状態 η' への推移のルール：鍵 η は 27 個のアルファベットが並んでいますが，その中から二つランダムにアルファベットを選んで交換したものを η' とします．

例えば，表 10 では，二つのアルファベット d と u が選ばれて，鍵 η によって d と u の移る先である (r, t) を (t, r) の順に交換した鍵が η' です．この操作を繰り返すことにより，鍵空間 S の

中でマルコフ連鎖が動きまわります.

$\eta =$

Before	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	◇
After	f	o	a	r	x	k	c	s	e	v	l	y	m	d	u	n	w	g	z	b	t	h	q	i	p	◇	j

$\eta' =$

Before	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	◇
After	f	o	a	t	x	k	c	s	e	v	l	y	m	d	u	n	w	g	z	b	r	h	q	i	p	◇	j

表 10: 鍵 η から鍵 η'

このマルコフ連鎖をもとに，復号鍵 σ^{-1} を見つけることができれば暗号化された文章 $\mathbf{y}$ は復号できるわけです．ですから，状態空間 S 上の関数 $f : S \to \mathbb{R}$ で σ^{-1} において最大値をとるような関数をうまく定義すれば，前節のてっぺんを捜す方法が使えるはずです.

5°) 関数 $f : S \to \mathbb{R}$ は以下のように定めます：暗号化された文章 $\mathbf{y}$ と鍵 η に対して $\mathbf{z} = \eta(\mathbf{y}) = z_1 z_2 \dots z_n$ となるとき，

$$f(\eta) = \prod_{i=1}^{n-1} M_{z_i z_{i+1}} = M_{z_1 z_2} M_{z_2 z_3} \cdot \dots \cdot M_{z_{n-1} z_n}$$

と定義します．例えば，$\mathbf{x}$ =this を暗号化した $\mathbf{y} = (\sigma(\mathbf{x}) =)$ wklv が与えられているとして，ある鍵 η によって，$\mathbf{z} = \eta(\mathbf{y}) =$ znoy となったとき，関数 $f(\eta)$ の値は

$$f(\eta) = M_{\mathrm{zn}} M_{\mathrm{no}} M_{\mathrm{oy}} = 0 \times 19 \times 0 = 0$$

となります．$M_{\mathrm{zn}}, M_{\mathrm{no}}, M_{\mathrm{oy}}$ などの値は表 9 を使っています．関数 f によって znoy を点数化すると 0 点となり，普通の英語の単語らしくないと判定するわけです．一方，η がたまたま復号鍵 σ^{-1} のときは，$\mathbf{z} = \eta(\mathbf{y}) = \sigma^{-1}(\mathbf{y})$ はオリジナルの文章 this になり，

$$f(\sigma^{-1}) = M_{\mathrm{th}} M_{\mathrm{hi}} M_{\mathrm{is}} = 134 \times 33 \times 57 = 252054$$

と高得点になるので，この単語は英語の単語としておかしくないと判定するわけです.

つまり，この関数 f のココロは

> 関数 f の値は，復号鍵 σ^{-1} のときは極めて大きい値をとり，そうでないときには小さい値 (多くの場合は 0) をとる.

この関数 f を用いて，前節で述べたマルコフ連鎖を用いて最大値を見つける問題を実際に実行すればよいのです[10].

基礎になっているのはマルコフ連鎖ですので，もちろん毎回実行結果は違います．あくまでもこの方法は確率的な暗号解読なのでいつでも必ずうまく行くというわけではありませんが，それなりにうまく解読ができることもあります.

実際に上に述べた方法で復号鍵の空間の中でマルコフ連鎖を走らせて，1000 回毎の復号鍵 η をとりだし，表 5 の暗号化された文章 $\mathbf{y}$ に対してその η を施した $\eta(\mathbf{y})$ を表示したものが以下の表 11 です．(表示されているのは文章の一部で全部ではありません．)

[10]原理的にはそうですが，実際にはもう少し工夫が必要です.

0: fyeaxjqfzjoxceddedcjbujcxbjhxgpjbegxrjukjzebbedcjopjsxgjzezbxgjudjbsxjofdljfdrjukjsfhedcj
1: ugado cut bohassash ie hoi ponm ianol ek taiiash bm fon tation es ifo busy usl ek fupash
2: upico fut bodissisd re dor hong rinol ek tirrisd bg mon titron es rmo busy usl ek muhisd
3: avico wat bohissish re hor pong rinol ed tirrish bg mon titron es rmo basy asl ed mapish
4: avice wat mepirrirp so pes leny sined ob tissirp my hen titsen or she mark ard ob halirp
5: apice was beginning to get dery tirel of sitting by her sister on the bank anl of hading
6: alice was beginning to get very tired of sitting by her sister on the bank and of having
7: alice was beginning to get very tired of sitting by her sister on the bank and of having
8: amice was beginning to get very tired of sitting by her sister on the banl and of having
9: alice was beginning to get very tired of sitting by her sister on the bank and of having

表 11: マルコフ連鎖の実行例 (1000 回毎の様子)

表 11 の様子を見ると，一行目の暗号化された文章の中にあるアルファベット j が空白 (とピリオドやカンマの記号) だということがまず解読されて，文章の切れ目がはっきりして各単語の長さが決定されます．その後個々の単語が自然なものへ変化していっていることがわかります．もし日本語で同じことをやろうとすると，(i) 英語と違って各単語間の切れ目がわかりにくい，(ii) 漢字があるので複雑，などの理由からもう少し工夫が必要だと思います．

6.5 解題

表 5 にあった暗号は，以下の文章を暗号化したものでした．

> Alice was beginning to get very tired of sitting by her sister on the bank, and of having nothing to do: once or twice she had peeped into the book her sister was reading, but it had no pictures or conversations in it, “and what is the use of a book,” thought Alice, “without pictures or conversation?
>
> — Lewis Carroll, “Alice’s Adventures in Wonderland”

この文章は『不思議の国のアリス』の 1 章 “Down the rabbit-hole” の冒頭の一節ですが，作者は有名なルイス・キャロルです．ルイス・キャロルというのはペンネームで，本名は Charles Lutwidge Dodgson といい，実は数学者・論理学者でもありました．数学では行列式に関する本を著し，凝集法なる行列式の計算方法を考案しています．ルイス・キャロルは言葉遊びが好きで暗号についてもいくつかの研究があります.『不思議の国のアリス』には多くの掛詞が使われており翻訳者泣かせで知られています．また『不思議の国のアリス』とともに人気のある『鏡の国のアリス』にはマザーグースのハンプティ・ダンプティが登場しますが，ハンプティ・ダンプティとアリスとのやり取りの中でいわゆる「かばん語」なる複数の単語をつなげた言葉遊びがでてきます．難解で有名なジェームス・ジョイスの『フィネガンズ・ウェイク』を翻訳した柳瀬尚紀氏がルイス・キャロルの翻訳を多く手掛けていますので，その言葉遊びをどのように訳しているか原著と比較して読んでみると面白いかもしれません．

参考文献

[1] 村上征勝：真贋の科学—計量文献学入門，朝倉書店，1994 年.

[2] 筒井康隆：残像に口紅を，中公文庫，1995 年.

[3] ルイス・キャロル (柳瀬尚紀訳)：不思議の国のアリス，ちくま文庫，1988 年.

[4] L. Carroll: Alice's Adventures in Wonderland, Project Gutenberg, available at `http://www.gutenberg.org/etext/11`

[5] P. Diaconis: The Markov Chain Monte Carlo Revolution, Bull. Amer. Math. Soc. **46**(2009) 179–205. Available at Prof. Diaconis's homepage.

[6] P. G. Doyle and J. L. Snell: Random Walks and Electoric Networks, available at `http://front.math.ucdavis.edu/0001.5057`

[7] A. N. Kolmogorov: Grundbegriffe der Wahrscheinlichkeitsrechnung. Berlin: Julius Springer, 1933. (Translation) Foundations of the Theory of Probability (2nd ed.). New York: Chelsea, 1956. English version is available at `http://www.mathematik.com/Kolmogorov/index.html`

[8] D. Levin, Y. Peres and E. Wilmer: Markov Chains and Mixing Times, A.M.S., 2008.