数学を楽しむ

トランプで同じ数字が重なる確率

西山　豊

1. キングとクイーン

イギリスの知人スティーブ・ハンブルから確率についてのメールが届く．トランプをよくシャッフルして開くと，キング(K)とクイーン(Q)が隣り合わせになっているか，または1枚おいて隣り合わせになっている確率が70%を超えるというのである．実際にトランプで試してみた．確かにKとQが隣り合わせになっている場合(KQ, QK)と，1枚おいて隣り合わせになっている場合(K|Q, Q|K)を調べてみると高確率で，7割は超えているように思えた．これは数学ではどのように説明すればいいのだろうか．

あれこれ考えているところ，彼から別のメールで，トランプをよくシャッフルして調べてみると，同じ数字のカードが並んでいる確率が96.35%であるというサイトを見つけたというのである[1]．96.35%とは，かなりの高確率である．トランプでゲームをする場合はランダムの度合いが常に問題となる．十分にシャッフルできているかをチェックする方法として，同じ数字が並んでいないかがひとつの目安である．もし並んでいたら，そのカードの1枚を別の場所に移動することがよくある．しかし移動しても数字が並ぶことが時々ある．

サイトには，この確率の計算方法が示されていなかった．そこであらためて96.35%の真偽について考えてみることにした．

2. 8のカードが重なる確率

最終目標は，少なくとも1組の数字が重なる確率を求めることであるが，それを求めるために順を追って考えていこう．

はじめに，1つの数字を固定する．たとえば数字が8のカードが重なる確率を求めてみよう．

トランプは全部で52枚ある．52枚の並び方は順列の公式より52!となる，この値は，

$$52! \approx 8.066 \times 10^{67}$$

となる．

8のカードは4枚あり，52枚の中で，この4枚が占める位置の組合せは公式より${}_{52}C_4$となり，計算すると，

$${}_{52}C_4 = \frac{52!}{(52-4)!\,4!} = 270725$$

となる．

52枚のカードから8のカード(4枚)を除くと48枚が残る．それらを|で表すとする(図1)．

||||||||||||||·········||

||||8|||8||8||·········|8||||

図1　8のカードが重ならない確率

8のカードを1枚ずつ|と|の間に入れるとすると，8のカードは隣り合わすことがない．8のカードを入れることができる場所は左右両端を含めると$48+1=49$か所ある．この49か所の中から4か所選ぶ組合せは${}_{49}C_4$であり，計算すると

$${}_{49}C_4 = \frac{49!}{(49-4)!\,4!} = 211876$$

となる．

以上の結果より，8のカード(4枚)がまったく隣り合わさない確率は

$$\frac{{}_{49}C_4}{{}_{52}C_4} = \frac{211876}{270725} = 0.782624$$

となる．少なくとも2枚の8のカードが隣り合わす確率は，この補集合として求まるから，

$$P_1 = 1 - 0.782624 = 0.217376$$

が，8のカードが重なる確率となる．

8のカードに注目して，カードが重なる確率を計算したが，約22%は意外と低い値である．トランプは数字が1(A)から13(K)まで，絵柄が4種類の合計52枚ある．8のカード(4枚)が重ならない確率は0.782624であった．少し粗っぽい計算ではあるが，1のカード(4枚)，2のカード(4枚)，…，Kのカード(4枚)すべてが重ならない確率は，

$$0.782624^{13} = 0.041323$$

で約4.13%となり，少なくとも1組の数字が重なる確率は，この補集合であるから，

$$P_2 = 1 - 0.782624^{13} = 0.958677$$

で約95.87%となる．

3. すべてを書きあげる

この95.87%は，サイトに掲載されていた96.35%とはオーダー的にはあっている．ただし，95.87%は大雑把な計算である．なぜなら1から13までの13種類の数字のカードは完全に独立しているのではなく，52枚の中で互いに関連しあっている．1のカードが重ならない確率と2のカードが重ならない確率を独立事象として掛け合わせることは厳密にはできない．

52枚のカードの並べ方は全部で

$$52! \approx 8.066 \times 10^{67} \text{ 通り}$$

であり，このなかで数字が重なっている場合を調べればよいわけだが，10^{67}は天文学的な数字であり，いくらパソコンが進化したとはいえ，プログラムでの処理は不可能であると思えた．

そこで，52枚のカードで考えるのではなく，4枚のカードから始めることにした．1(Ace)のカードが4枚の場合，この並べ方は，AAAAの1通りしかなく，数字が重なっているから，重なる確率は1である．1(Ace)のカード4枚と2のカード4枚の合計8枚の場合はどうだろうか．1(Ace)のカードの並べ方は${}_8C_4 = 70$通りあり，重ならないのは

2A2A2A2A と A2A2A2A2

の2通りしかなく，これ以外の68通りはすべて重なる．したがって重なる確率は

$$68/70 = 0.9714$$

となる．

1のカード4枚と2のカード4枚と3のカード4枚の合計12枚の場合はどうだろうか．12枚の中から1のカード4枚を選ぶ組合せは${}_{12}C_4 = 495$通りであり，残る8枚の中から2のカード4枚を選ぶ組合せは${}_8C_4 = 70$通りであるので，12枚のカードの並べ方は

$${}_{12}C_4 \times {}_8C_4 = 495 \times 70 = 34650 \text{ 通り}$$

となる．3種類のカードが重ならない場合は1092通りあり，これらを除くと確率は

$$(34650 - 1092)/34650 = 0.9685$$

となった．

ここで，1092通りをどのようにして求めたかを説明しておこう．組合せの${}_{12}C_4 = 495$通りのリストと，${}_8C_4 = 70$通りのリストは，「組合せリスト」を作成する公開されたプログラムを使った．このプログラムのアルゴリズムは，エレガントで素晴らしいものであるが，これについての説明は別の機会とする．

最初に表計算ソフトを使い${}_{12}C_4 = 495$通りの1のカードの位置をリストアップした．その後${}_8C_4 = 70$通りの2のカードの位置をリストアップした．カードは全部で12枚であるから，12の配列を用意しておき，1のカードを埋め，その後，前から空いている配列に2のカードを埋める．そして最後に空いている配列に3のカードを埋めるという手順である．

このようにしてできあがった34650通りのパターンについて，表計算ソフトのマクロを使って数字が重なっていないか逐次検査すると，重ならないパターンが1092通りであることがわかった．重なるパターンは$34650 - 1092 = 33558$であり，重なる確率は

$$33558/34650 = 0.9685$$

であり，96.85%となった．

カードの種類が3個で枚数が12枚であっても34650通りものパターンがあり，手作業は無理で，プログラムの力を借りることになったが，34650通りすべてを調べる必要がなく，実際は，

$$\frac{{}_{12}C_4}{3} \times \frac{{}_8C_4}{2} = 165 \times 35 = 5775$$

通りの6分の1のパターンでよいことがわかった．

以下同様にして，1から4のカードで16枚の場合は，パターン総数が63063000通りで，重なるパターンが60797976通りで，重なる確率は96.41%であり，1から5のカードで20枚の場合は，重なる確率は96.14%であった(表1)．

カードの組合せは幾何級数的に増加していく．52枚のカードの場合は，

$$\frac{{}_{52}C_4}{13}\times\frac{{}_{48}C_4}{12}\times\cdots\times\frac{{}_{8}C_4}{2}$$

$$=20825\times16215\times\cdots\times35=1.478\times10^{40}$$

通りのパターンについて，すべてを検査しなければならない．$52!\approx8\times10^{67}$ に比べれば少ないが，これだけのパターンを書き上げる記憶容量も計算時間も現在のパソコンでは無理であり，カード枚数20枚が限界であることを知った．

カード20枚ですべてを書き上げて求めた確率は96.14％であるが，サイトに掲載されていた確率は96.35％であり，この値より小さいことになる．カード枚数が52枚なら確率はもっと小さくなるのではと予測した．

枚数	パターンの総数	重なる場合数	確率
4	1	1	100％
8	70	68	97.14％
12	34650	33558	96.85％
16	63063000	60797976	96.41％
20	305540235000	293752962960	96.14％

表1　数え上げによる方法

4. 乱数によるシミュレーション

すべての場合を書きあげ，それぞれ同じ数字が並んでいないかをチェックする方法は，カード枚数が20枚（＝5種類×4枚）が限界であることを知り，52枚のカードについてはどうすればよいのか思案に暮れた．私の頭の中で，この方法は絶対使いたくないと決めていた方法にシミュレーションによる方法があったが，この方法を検討せざるを得なかった．

パソコンには一様乱数を生成する関数が用意されている．周知のように一様乱数は開区間(0, 1)の実数値（小数）として生成される．これが利用できないかということだ．ずいぶん前の話であるが，ビンゴ・ゲームを，乱数を用いてシミュレーションし，確率を計算したことがあるので，これが生かせないか考えてみた(2)．

ビンゴ・ゲームは，各自が5×5のマス目に1から25までの数字を記入したカードを用意しておく．そして1から25までの番号がついたボールがピッチング・マシンの中に入っていて，出てくるボールの番号を各自が○印をつけて，縦，横，または斜めに1列並べば勝つというゲームである．

私は，r 個のボールで1列そろう確率 P_r を計算すると，

$$P_r=\frac{12\times{}_{20}C_{r-5}}{{}_{25}C_r}$$

となり，$r=8$ のときの理論値が $P_8=0.1265\times10^{-1}$ で，乱数による実験値は 0.1258×10^{-1} であり，比較的よくあっていたのを経験している．

ビンゴ・ゲームの場合は，(0, 1)区間の一様乱数からボールの番号1から25を生成した．トランプの場合は，52枚のカードがあるので，とりあえず，開区間(0, 1)の一様乱数（実数）から閉区間[1,52]の乱数（整数）を生成することを考える．(0,1)で生成される一様乱数を x としておく．

最初は52個の数字から1個選ぶ．1回出た数字は使えないから，2回目は51個の数字から1個選ぶことになる．3回目は50個の数字から1個選ぶ．これらの数字を I_1, I_2, I_3 としておくと，これらは，

$$I_1(\text{整数})=x(\text{実数})\times52+1$$

$$I_2(\text{整数})=x(\text{実数})\times51+1$$

$$I_3(\text{整数})=x(\text{実数})\times50+1$$

のようにして，一様乱数 x から求めることができる．右辺の値（実数）から小数点以下を切り捨てることによって左辺の値（整数）が計算される．この方法は，それほど問題なく処理される．

たとえば，一様乱数 x が，0.55817, 0.08649, 0.91924などと求められたとき，

$$0.55817\times52+1=30.02470$$

$$0.08649\times51+1=5.41087$$

$$0.91929\times50+1=46.96441$$

となり，小数点以下が切り捨てられて，整数の乱数30, 5, 46が計算される．

さて，このようにして得られた乱数の整数値を実際のカードの番号1から52に置き換えるのは次のように考えるとよい．乱数（整数）が $I_1=30$, $I_2=5$, $I_3=46$ となったとしよう．ここでの30, 5, 46は

$$1\leqq30\leqq52$$

$$1\leqq5\leqq51$$

$$1\leqq46\leqq50$$

である．

52個の箱を用意しておき，1番から52番の番号をつけておく．カードにも1番から52番の番号を書いておく．最初の乱数は30であるから30番目の箱に1番目のカードを入れる．そして，その箱はカードを入れたので蓋をする．入れることのできる箱は1個減り51個となる．次の乱数は5であるから，開いている箱で前から5番目のところに2番目のカードを入れる．そして，その箱は蓋をする．入れることができる箱は50個になる．

次の乱数は46であるから，開いている箱で前から46番目の箱を探す．すでに5番目と30番目の箱はカードが入っていて蓋がしてあるから，その箱を飛ばすと，実際の場所は46＋2＝48番目の位置にある箱ということになる．そして48番目の箱に3番目のカードを入れて，その箱は蓋をする．このような手順で52枚のカードを52個の箱にすべて入れることができる．そこで1番目の箱から蓋を開けてカードを取り出し，カードの番号を読み上げていくと，それが1から52までの乱数となる．

このように一様乱数から52枚のカードの乱数に関連付けるのは，2段階の手順がいるのは事実である．トランプは52枚のカードがあるが，4枚ずつが同じ数字である(A, 2, 3, …, K)．それで，1から52の乱数を，

(A, A, A, A, 2, 2, 2, 2, 3, 3, 3, 3, …, K, K, K, K)

に対応することにする．この作業はプログラムではそれほど難しくはない．

このようにして生成された乱数が実際のシミュレーションに適しているか検証の作業が必要になる．

冒頭では，8のカードが重なる確率は，

$$P_1=1-\frac{{}_{49}C_4}{{}_{52}C_4}=1-0.782624=0.217376$$

であったので，これを試すことにした．シミュレーションの試行回数と実験値，有効桁の関係は表2のようになった．1000万回の試行回数では有効桁が約4桁である．

試行回数	確率の実験値	有効桁
100000(＝10^5)	0.21629	2桁
1000000(＝10^6)	0.217102	3桁
10000000(＝10^7)	0.2173721	4桁
理論値	0.217375566…	

表2　乱数による方法の検証

試行回数を単純に増やせば理論値に近づくことが予想できるが，一様乱数の周期（$2^{31}\approx2.1\times10^9$）を超えることができないし，パソコンの限界もあるので，試行回数の1000万回（＝10^7）は妥当な数であろう．試行回数を1000万回として，数え上げの方法で確率を正確に求めたカードが8枚から20枚のケースについて，理論値と乱数による実験値を比較してみた．これによると有効桁が3桁～4桁であることがわかる．乱数による方法がほぼ有効であることが確認できたので，カード枚数が24枚～52枚についても実験値を求めてみた（表3）．

枚数	数え上げ	乱数
4	100％	
8	97.14％	97.12％
12	96.85％	96.86％
16	96.41％	96.40％
20	96.14％	96.13％
24		95.95％
28		95.82％
32		95.70％
36		95.64％
40		95.56％
44		95.51％
48		95.49％
52		95.44％

表3　数え上げと乱数による確率の比較

トランプのカード52枚について，粗い計算で求めた確率は95.87％であった．乱数による確率は95.44％となり，近い値が出た．インターネットのサイトでは96.35％となっていたので(1)，私の計算結果をまとめて問い合わせたが未だ返事がない．掲載された記事は1996年の15年前のものなので，担当者が移動しているかもしれない．いずれにしても，同じ数字が重なる確率は95％～96％であるということになる．

トランプ・ゲームなどでカードがよく繰られているかをチェックする場合，同じ数字のカードが続いていないかを調べることがよくあるが，この方法はあまり有効ではないようだ．同じ数字のカードが見つかったとき，その1枚のカードを別の場所に移動しても，また重なることがよくある．それは，今回，確率の計算をしたように，数字が並ぶ確率は約95％という高確率であるということだ．したがって，重なる数字がないようになるまでシャッフルを繰り返すなど，神経質にならなくてもよいという結論になる．

参考文献

(1) Random Card Shuffling Probabilities,（米国ドレクセル大学の数学フォーラム）
http://mathforum.org/library/drmath/view/52153.html
(2) 西山豊「ビンゴ・ゲーム」『数学セミナー』1983年4月，Vol.22, No.4, 61－66,（『サイエンスの香り』（日本評論社）に所収）

（にしやま　ゆたか／大阪経済大学）