高校生のための現代数学講座
「確率の考え方」
講義(1)　楠岡成雄

東京大学
玉原国際セミナーハウス
2005 年 10 月 22 日

確率について

不確実なものは、なるべく避けたいと思うかも知れない。しかし、乱数を人工的に発生させ不確実性を積極的に利用していくということが、実際に行われている。このことを紹介することがこの講義の目的である。

ゼロ和ゲーム

じゃんけんをするとき、くせがあってそのくせを読みとられてしまうと、長くじゃんけんを続けると負ける率が高くなる。くせのないようにじゃんけんを行うには、でたらめに出す手を選ぶという方法がある。

じゃんけんを２人でして、負けた方が勝った方にポーカーチップを渡すというゲームをすることにしよう。もし、どの場合も負けた方は勝った方にポーカーチップを１枚だけ渡すとすると、「グー」、「チョキ」、「パー」はどれも対等なので、でたらめに同じ確率１／３で出せば、多分負けないであろう。実際、利得表を作ると次のようになる

		自分		
		グー	チョキ	パー
相手	グー	0	-1	1
	チョキ	1	0	-1
	パー	-1	1	0

たとえばもし、さいころを振り１,４が出ればグー、２,５が出ればチョキ、３,６が出たらパーを出すことにすると、

相手がグーを出すとき利得の期待値は

$$0 \times (1/3) + (-1) \times (1/3) + 1 \times (1/3) = 0$$

相手がチョキを出すとき利得の期待値は

$$1 \times (1/3) + 0 \times (1/3) + (-1) \times (1/3) = 0$$

相手がパーを出すとき利得の期待値は

$$(-1) \times (1/3) + 1 \times (1/3) + 0 \times (1/3) = 0$$

となる。大数の法則により、何度もじゃんけんを繰り返すと、利得の平均は期待値に近づくことが知られている。したがって、この戦略を用いれば勝つこともないが負けることもないことがわかる。

ではグーで勝ったときはポーカーチップを３枚、チョキ・パーで勝ったときは６枚もらえるとしたらどうだろう。この時の利得表は

		自分		
		グー	チョキ	パー
相手	グー	0	-3	6
	チョキ	3	0	- 6
	パー	-6	6	0

となる。先と同じようにどの手も１／３の確率で出すと、

相手がグーを出すとき利得の期待値は

$0 \times(1/3) + (-3)\times(1/3) + 6\times(1/3) = 1$

相手がチョキを出すとき利得の期待値は

$3 \times(1/3) + 0 \times(1/3) + (-6)\times(1/3) = -1$

相手がパーを出すとき利得の期待値は

$(-6)\times(1/3) + 6 \times(1/3) + 0 \times(1/3) = 0$

となるので、相手はチョキを出せば勝てることになる。

では、グーを確率 ｐ で、チョキを確率 ｑ で、パーを確率 ｒ で出すことにすると、ｐ＋ｑ＋ｒ＝１であり、

相手がグーを出すとき利得の期待値は　　－３ｑ＋６ｒ

相手がチョキを出すとき利得の期待値は　　３ｐ－６ｒ

相手がパーを出すとき利得の期待値は　　－６ｐ＋６ｑ

となる。　ｐ＝２／５，ｑ＝２／５，ｒ＝１／５とおくと、期待値はすべて０となる。この戦略で長くじゃんけんをすると、グー、チョキ を多く出すということを知られてしまうが、それでも勝つことも負けることもない戦略となる。

乱数による暗号

暗号は現在いろいろなところで用いられており、日常的なものとなりつつある。しかし、かっては、主に軍事や外交において電波で情報を送るために用いられていた。ここでは簡単のために、アルファベットで書かれた文章を情報の送信者から受信者へ情報を送ることを考える。アルファベット２６文字に「空白」も文字と考えて加えると、２７文字となり、３の３乗なので、０，１，２の３文字で３進数的に表記することができる。

A 001	B 002	C 010	D 011	E 012	F 020	G 021	H 022	I 100	J 101	K 102	L 110	M 111	N 112
O 120	P 121	Q 122	R 200	S 201	T 202	U 210	V 211	W 212	X 220	Y 221	Z 222	空白 000	

暗号の作り方には昔から色々なものが考案されているが、ここでは置き換えによる暗号の作り方と乱数暗号について解説する。

（置き換えによる方法）

今、次のような表を考える。

A N	B L	C P	D Q	E O	F J	G K	H I	I V	J W	K U	L Y	M Z	N X
O S	P T	Q R	R D	S E	T C	U G	V H	W F	X A	Y B	Z M	空白 空白	

この表によって　NUMATA　は　XGZNCN　に置き換わり、その３進数化は

２２０　０２１　２２２　１１２　０１０　１１２

となる。このような暗号は短い文に１度きり使う限りは解読されないが、何度も用いたり長文に用いるとその規則を見破られ可能性が高い。そのため傍受されると暗号が破られる可能性が高い。

（乱数に基づく暗号）

乱数に基づく暗号は 1917年に G.Vernam により考案されたものでバーナム暗号とも呼ばれる。この暗号は解読表（乱数表）を知らなければ、絶対に解読することが出来ない。乱数暗号は以下のようにして作る。

	N	U	M	A	T	A
① ３進数化	１１２	２１０	１１１	００１	２０２	００１
② 乱数	００１	１２２	２２２	０２１	００２	２１０
③ 暗号	１１０	００２	０００	０２２	２０１	２１１
④ 乱数の補数	１００	１２１	０１２	１２１	２１１	１２０

３進和

	0	1	2
0	0	1	2
1	1	2	0
2	2	0	1

ここで②の乱数はさいころを振り、１,４が出たら１，２,５が出たら２，３,６が出たら０という風にして作った数列である。①、②から③は３進和（２つの数を足して３以上になったら３を引く）で作ったものである。④の乱数の補数は②の乱数との３進和が０となるような数である。この時、①は③と④の３進和で求めることができる。従って、暗号の受信者は④の数の列を知っていれば、暗号を解くことができる。

③の数の列には規則性が全くない。それは②で現れる２４個の数字の列がどの列も同じ確率（１／３）^{24}で現れるので、③の暗号に現れる２4個の数字の列も同じ確率で現れるためである。

	H	A	K	A	T	A
① ３進数化	００２	００１	１０２	００１	２０２	００１
② 乱数	２０１	２００	２２０	０１２	１２１	１１０
③ 暗号	２００	２０１	０２２	０１０	０２０	１１１

たとえば、原文が ＮＵＭＡＴＡであったか、ＨＡＫＡＴＡ であったかはベイズの定理を用いて推測すると、同じような確からしさとなり、区別できない。乱数暗号はあまり用いられなくなったが、文字列がどれだけの情報をもつかという考え方はシャノンによる情報量という概念に発展していった。これは通信の理論ではなくてはならないものとなっている。また、量子通信という新技術ができ、乱数暗号は近い将来に量子暗号として復活すると予測されている。

サイ投げで１次方程式を解く

９つの未知数A,B,C,D,E,F,G,H,I を未知数とする９元の一次方程式

(1) B+0+1+D = 4A　　(2) C+1+A+E = 4B　　(3) 0+1+B+F = 4C

(4) E+A+1+G = 4D　　(5) F+B+D+H = 4E　　(6) 0+C+E+I = 4F

(7) H+D+0+1 = 4G　　(8) I+E+G+1 = 4H　　(9) 1+F+H+0 = 4I

を考えよう。見るのも嫌になるかも知れないが、この方程式は図　に関係した方程式である。A,B,C,D,E,F,G,H,I の値それぞれが、それと線で結ばれている４つの値の平均と等しいということを表したのが(1)〜(9)の式である。

この方程式のＤの値を知りたい。どうすればよいか。

９元の一次方程式は頑張れば手で解けるかもしれない。実際の応用では、未知数の数は 10^{15} に及ぶ場合があり、スーパーコンピュータを用いても普通のやり方では解けない。サイ投げで解く方法については講演の中で述べる。

実験１

まず、簡単な　方程式から考える

(1) 0 + B = 2A　(2) A + C = 2B　(3) B + D = 2C　(4) C + 1 = 2D

これの答えは簡単に　A = 1/5,　B= 2/5,　C=3/5,　D=4/5 となることがわかるのであるが、サイ投げでＢの値を求める方法を実験してみる。

実験２

次に上の問題をサイ投げで解く方法を実験してみる。

乱数

上のような問題でサイ投げで乱数を発生させようとすると、何万回もさいころを振る必要があり、事実上不可能である。サイ投げの結果現れるような数の列を乱数と呼ぶ。乱数を高速に発生させるにはどうすればよいか、いろいろな方法が考えられている。放射性物質が発生させる放射線をカウントしたり、熱の生成する雑音を使用したりという方法が考えられているが、このような方法で得られる　０,１　よりなる乱数列で性質の良いものはせいぜい１秒間に数百万程度とされており、精度の高い計算にはあまり十分の早さではない。このため、コンピュータによる乱数発生というものが研究されている。ただしこれは厳密な意味での乱数には決してならないので擬似乱数と呼ばれている。擬似乱数の発生には、整数論をはじめとする色々な数学のアイデアが用いられている。