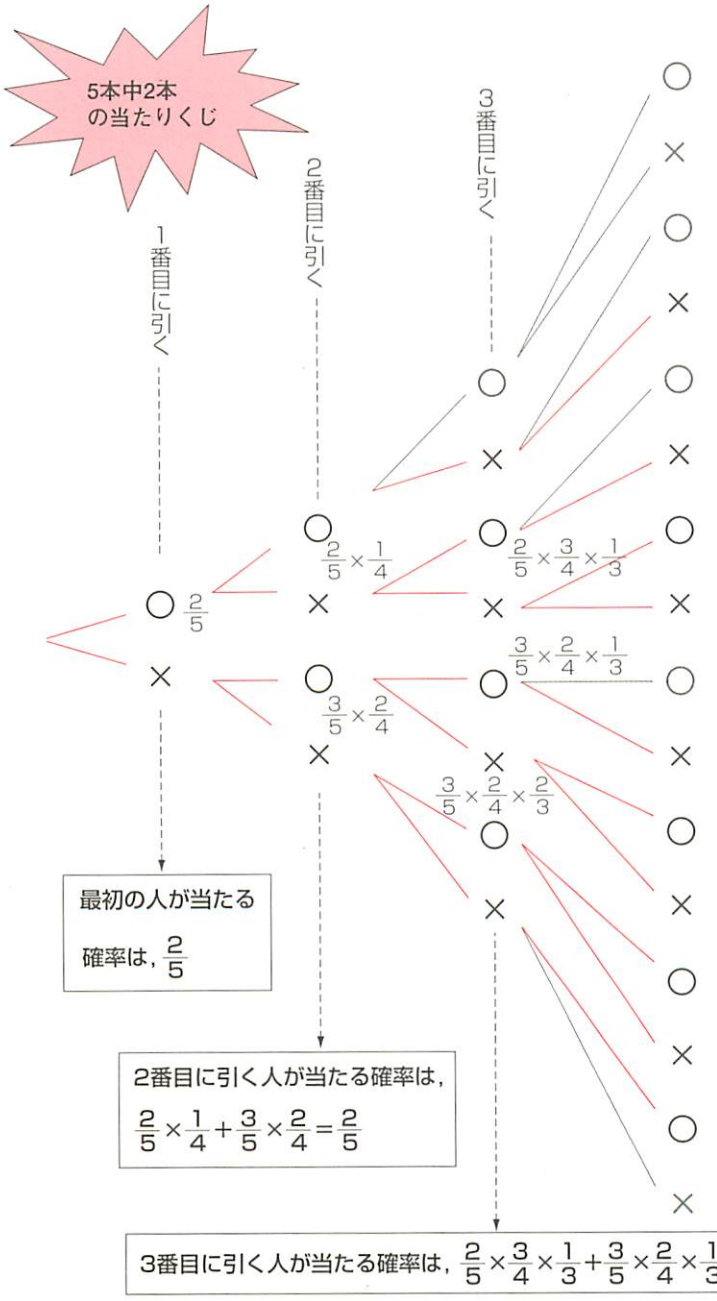
くじの当たりはずれは引く順序に無関係



§35 家族計画で男女比は変わらない

ここが要点！ 昔は男の子が欲しいという親が多かったが，最近は，「少なくとも１人は女の子」という考えに変わってきたようです。そこで，「女の子が産まれるまで生み続けるが女の子が産まれたらもう子供を作らないことにする。ただし，子供は３人まで」という家族計画で，すべての夫婦が家族計画を立てるものとします。この計画に従った場合，人口に占める女の子の割合が増えるように思われます。しかし，実際には自然にまかせて子供を作った場合と出生児の男女比は変わらないのです。

やさしい解説

将来，老後の面倒を見てもらうなら長男の嫁よりも実の娘，だから女の子が欲しい……。親の勝手な都合ですが，男の子が生まれるか女の子が産まれるかは，親にはどうしようもないことで神のみぞ知ることです（最近はちょっと事情が違うようですが）。しかし，産むことをやめることはできそうです。そこで，冒頭にあげた家族計画を立てて女の子をできる限り授かろうとしてみました。この前提で子供を出産した場合，子供の人数の期待値$\overline{S}$をまず求めてみましょう。ただし，女子の産まれる確率を0.5とします。

子供が１人の確率は女が産まれたときで$\frac{1}{2}$，子供が２人の確率は男，女と産まれたときで$\left(\frac{1}{2}\right)^2$，子供が３人の確率は男，男，女と産まれたときの$\left(\frac{1}{2}\right)^3$と男，男，男と産まれたときの$\left(\frac{1}{2}\right)^3$を加えた確率。したがって，

$$\overline{S}=1\times\frac{1}{2}+2\times\left(\frac{1}{2}\right)^2+3\times\left(\left(\frac{1}{2}\right)^3+\left(\frac{1}{2}\right)^3\right)=\frac{14}{8}$$

また，女の子の人数の期待値$\overline{G}$は$\overline{S}$と同様に考えて，

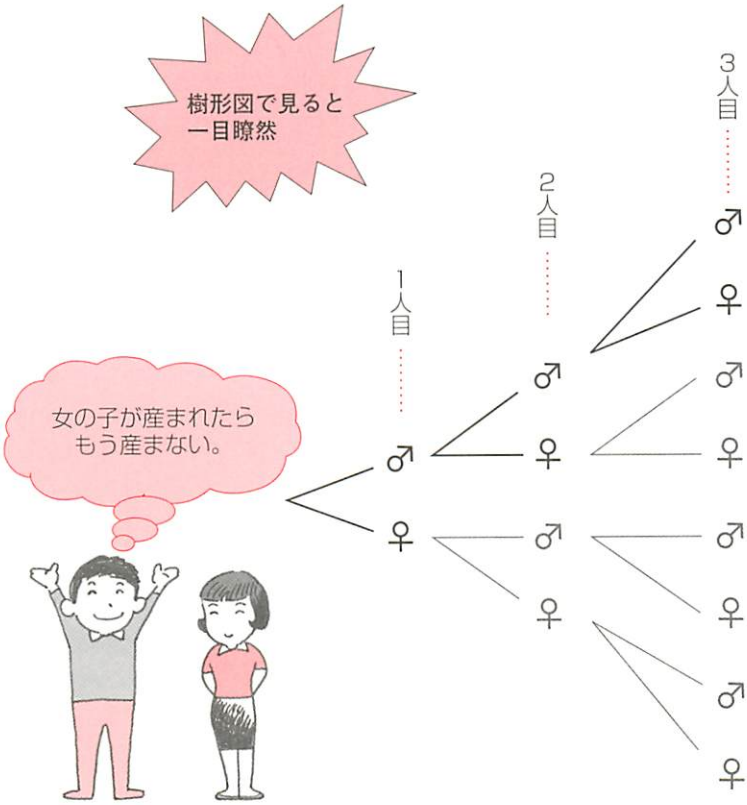
$$\overline{G}=1\times\frac{1}{2}+1\times\left(\frac{1}{2}\right)^2+1\times\left(\frac{1}{2}\right)^3=\frac{7}{8}$$

となります。したがって，子供全体に占める女子の割合は，

$$\frac{\overline{G}}{\overline{S}}=\frac{7}{14}=\frac{1}{2}$$

となり，男女比は変わりません。

家族計画を行っても子供の男女比は変わらない



参考 **n人まで生み続けたら**

「女の子が産まれるまで生み続けるが，女の子が産まれたらもう子供を作らないことにする。ただし，子供はn人までとし，女子の産まれる確率をpとする」という条件で考えると，産まれてくる子供の人数の期待値$\overline{S}$と女の子の人数の期待値$\overline{G}$は，それぞれ次のようになります。ただし，$q=1-p$

$$\overline{S}=1\times p+2\times qp+3\times q^2p+\cdots\cdots+n\times(q^{n-1}p+q^n)$$
$$=1+q+q^2+\cdots\cdots+q^{n-1}$$
$$\overline{G}=1\times p+1\times qp+1\times q^2p+\cdots\cdots+1\times q^{n-1}p$$
$$=p(1+q+q^2+\cdots\cdots+q^{n-1})$$

したがって，$\dfrac{\overline{G}}{\overline{S}}=p$ が成立します。

§36 意外に多い同じ日の誕生日

ここが要点！ 20数人集まれば，この中に同じ誕生日（月日が一致）の人が少なくとも1組いる確率は0.5を超えます。40人集まれば，確率はほぼ0.9になります。年間，ほぼ，365日あるという前提からすると，多くの人がこの確率は大きすぎるのではと首をかしげます。でも，理論的には正しいのです。

やさしい解説

太郎君を含む20人がいた場合，それぞれが独立に365通りの産まれ方をすると考えていいわけですから，このくらいの人数では太郎君と同じ誕生日の人はいないであろうと考えてもおかしくありません。

実際，太郎君が20人の中の特定の次郎君と同じ誕生日である確率は1/365です。したがって，他の人との可能性を考えても1/2を超えることはないと思われます。しかしここでは，特定の太郎君が他の誰かと同じ誕生日の人がいるかどうかを問題にしているわけではありません。問題をよく読んでください。「同じ誕生日（月日が一致）の人が少なくとも1組いる確率」なのです。

この「少なくとも」という条件が，確率をかなり高めることになるのです。それでは，n人の人がいるとき，この中に同じ誕生日（月日が一致）の人が少なくとも1組いる確率pを正しく求めてみることにしましょう。

それには，まず，n人の人の誕生日がすべて異なる確率qを求めてみることにします。すると，$p=1-q$より，pが求められます。

ここで，n人の誕生日がすべて異なる確率qは，次のようになります。

2番目の人が1番目の人と違う誕生日の確率

$$q=\frac{365}{365}\times\frac{364}{365}\times\cdots\times\frac{365-n+1}{365}$$

$$=\frac{365\times364\times\cdots\times(365-n+1)}{365^n}$$

$$=\frac{{}_{365}\mathrm{P}_n}{365^n}$$

n番目の人が1から$n-1$番目の人と違う誕生日の確率

したがって，

$$p=1-\frac{{}_{365}\mathrm{P}_n}{365^n}$$

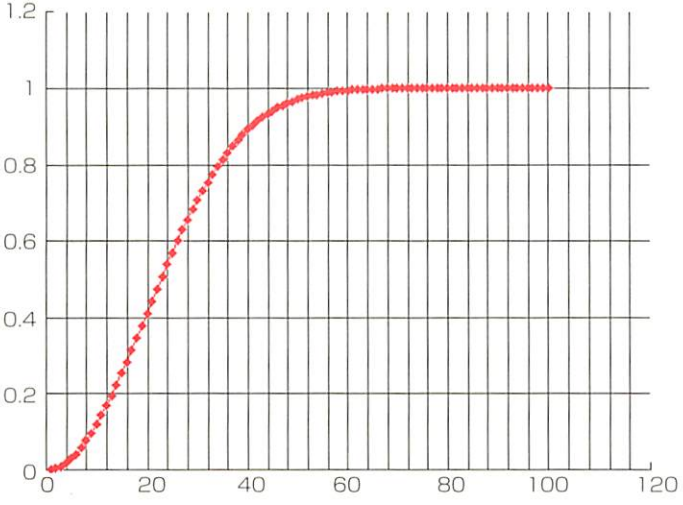
となります。人数nによって，pの値がどう変化するかは右ページを見てください。

意外に多い同じ誕生日の人

1 人数と同じ誕生日の人が少なくとも 1 組存在する確率の関係

人数	確率	人数	確率	人数	確率	人数	確率
1	0.0000	26	0.5982	51	0.9744	76	0.9998
2	0.0027	27	0.6269	52	0.9780	77	0.9998
3	0.0082	28	0.6545	53	0.9811	78	0.9999
4	0.0164	29	0.6810	54	0.9839	79	0.9999
5	0.0271	30	0.7063	55	0.9863	80	0.9999
6	0.0405	31	0.7305	56	0.9883	81	0.9999
7	0.0562	32	0.7533	57	0.9901	82	0.9999
8	0.0743	33	0.7750	58	0.9917	83	1.0000
9	0.0946	34	0.7953	59	0.9930	84	1.0000
10	0.1169	35	0.8144	60	0.9941	85	1.0000
11	0.1411	36	0.8322	61	0.9951	86	1.0000
12	0.1670	37	0.8487	62	0.9959	87	1.0000
13	0.1944	38	0.8641	63	0.9966	88	1.0000
14	0.2231	39	0.8782	64	0.9972	89	1.0000
15	0.2529	40	0.8912	65	0.9977	90	1.0000
16	0.2836	41	0.9032	66	0.9981	91	1.0000
17	0.3150	42	0.9140	67	0.9984	92	1.0000
18	0.3469	43	0.9239	68	0.9987	93	1.0000
19	0.3791	44	0.9329	69	0.9990	94	1.0000
20	0.4114	45	0.9410	70	0.9992	95	1.0000
21	0.4437	46	0.9483	71	0.9993	96	1.0000
22	0.4757	47	0.9548	72	0.9995	97	1.0000
23	0.5073	48	0.9606	73	0.9996	98	1.0000
24	0.5383	49	0.9658	74	0.9996	99	1.0000
25	0.5687	50	0.9704	75	0.9997	100	1.0000

2 人数と確率の関係を示したグラフ



§37 サイコロの起源は占いの道具

ここが要点！ サイコロはギャンブルやゲームなどに使用されていますが，その起源は占いの道具として誕生したといわれています。サイコロの歴史は古く，紀元前3000年ころから使われているそうです。古代には動物の骨で作られたようですが，現在では，ハイテクを駆使した高精度のサイコロも作られています。

やさしい解説

サイコロの歴史は長く，古代には動物の骨などで作られていたようです。

サイコロがどこで生まれたかについては定説がありませんが，紀元前3000年ごろにはすでにエジプトに存在していたようです。ここでのサイコロは，すでに1の裏が6，2の裏が5というように，反対面の数字の合計がどこをとっても7になるように作られていました。このころ，インダス文化においてもサイコロが存在していたのですが，これは反対面の数字の合計が7にはならないものでした。

日本には中国経由で入ってきたようですが，それは反対面の数字の合計が7になるように作られたエジプトと同じものです。当初，サイコロは賭事に使われるというよりも，占いの道具として使われていたようです。つまり，神意が出た目によって告げられると考えられていたのです。

また，7という数字は，昔から宗教的に縁起のよい数字とされていました。ヨーロッパでは古代イスラムの時代より7は聖数とされていました。日本でも吉数と呼び，七福神や七尊人などがあります。

サイコロはいろいろな時代に多くの人々に親しまれたためか，これを使った言葉が数多く残されています。以下に有名な言葉を紹介しましょう。

「賽は投げられた」（カエサル）

法を破ってイタリアへ入る際に，カエサルがいったとされる言葉。

「鴨川の水，双六の賽，山法師，これぞ朕の如意ならざるもの」（白川法王）

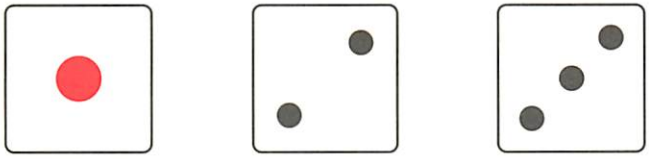
「神はサイコロを振らない」（アインシュタイン）

すべての現象は，何らかの規則性を持って起きているというほどの意味。

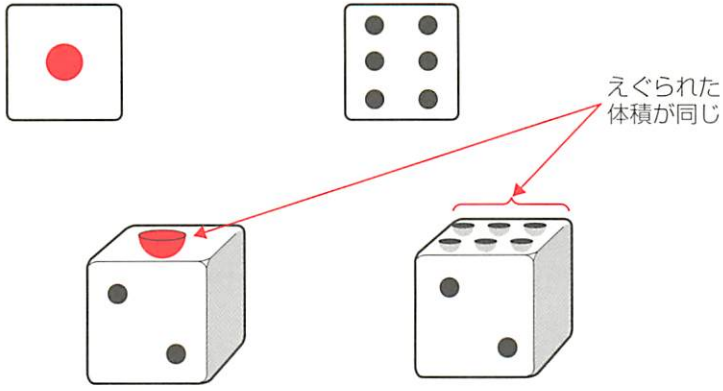
サイコロにはさまざまな工夫が

1 目の配置には対称性

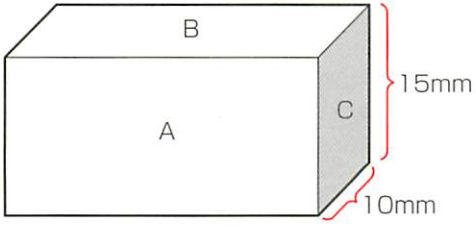
……えぐったり，塗ったりで面の重さに偏りが生じないように！



2 目の描き方に等重量性



3 直方体のサイコロの確率は？



（A，B，Cの面積比は3：2：1）

実験例
- A面の出る相対度数は，$\frac{759}{1000}$
- B面の出る相対度数は，$\frac{235}{1000}$
- C面の出る相対度数は，$\frac{6}{1000}$

§38 ジャンケンは少人数で

ここが要点！ 何かを決めるときジャンケンがよく使われます。しかし人数が多くなると，通常のグウ，チョキ，パーの３種類出すジャンケンはあまり使われません。なかなか決着がつかないことが経験的にわかっているからでしょう。

やさしい解説

まず，４人でジャンケンをしたときに，勝者が誰も決まらない（あいこ）になる確率を求めてみましょう。そのためには「誰か勝者が決まる確率」を求めて１から引けばいいわけです（**余事象の定理**）。

１－（誰か勝者が決まる確率）　…①

そこで，勝者が決まる場合の数を求めてみます。

(1) １人だけ勝つ場合

３（グウ，チョキ，パーのいずれか）

$\times {}_4C_1$（どの１人か）＝12通り

(2) ２人だけ勝つ場合

３（グウ，チョキ，パーのいずれか）

$\times {}_4C_2$（どの２人か）＝$3\times 6=18$通り

(3) ３人だけ勝つ場合

３（グウ，チョキ，パーのいずれか）

$\times {}_4C_3$（どの３人か）＝$3\times 4=12$通り

ここで，４人の出し方は$3^4=81$通り。したがって，誰か勝者が決まる確率は，

$$\frac{3\times {}_4C_1+3\times {}_4C_2+3\times {}_4C_3}{3^4}$$

$$=\frac{42}{81}=\frac{14}{27}$$

したがって，式①より４人でジャンケンをしたとき，勝者が誰も決まらない確率は，

$$1-\frac{3\times {}_4C_1+3\times {}_4C_2+3\times {}_4C_3}{3^4}=\frac{13}{27}$$

同様に考えて，n人でジャンケンをしたときに，勝者が誰も決まらない確率は，次のようになります。

$$1-\frac{3\times {}_nC_1+3\times {}_nC_2+\cdots+3\times {}_nC_{n-1}}{3^n}$$

$$=1-\frac{3\,(2^n-2)}{3^n}\quad\cdots\cdots\cdots\cdots\cdots\cdots②$$

注）式②を導くにあたり，次式を利用しました。

${}_nC_0+{}_nC_1+{}_nC_2+\cdots+{}_nC_{n-1}+{}_nC_n=2^n$

式②により，あいこになる確率を人数別に求めると，次のようになります。

$n=10$　のとき　0.95

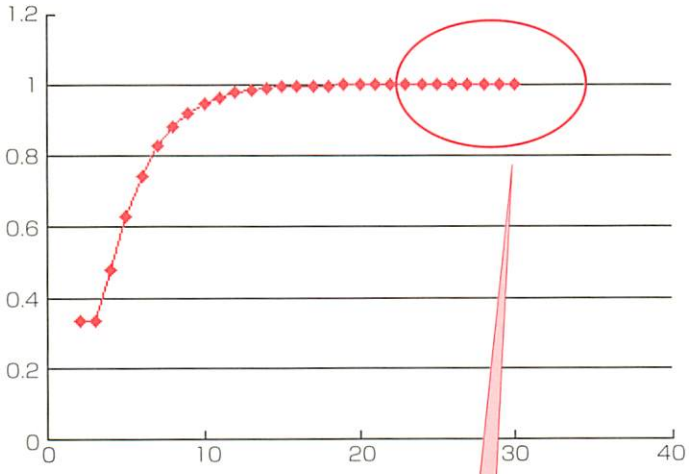
$n=30$　のとき　0.999984

$n=50$　のとき　1

人数が少し増えると，あいこの確率が急に増えることがわかります。もはや30人では，何回ジャンケンを繰り返してもほぼあいこになります。

ジャンケンは人数が多いときは不向き

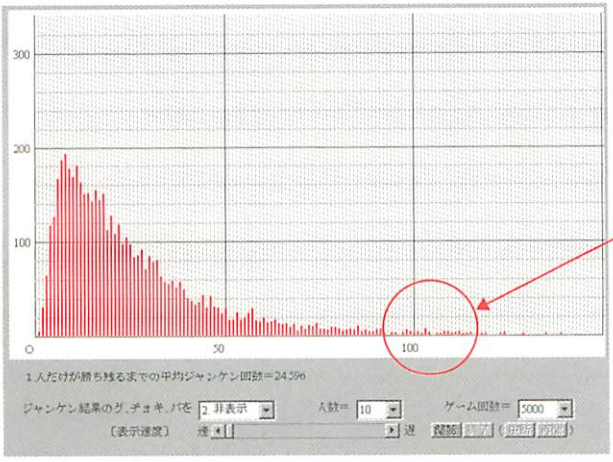
1 n人で1回ジャンケンして勝者が出ない確率



確率は $1-\frac{3(2^n-2)}{3^n}$

2 10人でジャンケンをして勝者が1人に決まった度数分布表

下のグラフは10人でジャンケンをして勝者が1人に決まったときのジャンケン回数の度数分布表(総度数は5000回)



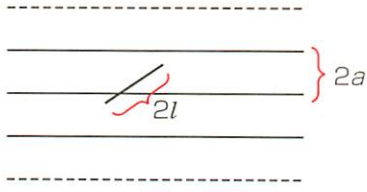
100回を超えることもある!!

§39 針を落とせば円周率

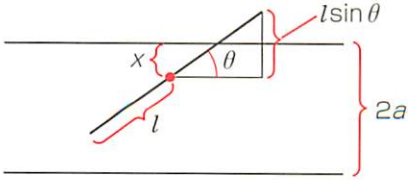
ここが要点！ 「針を投げれば円周率πの値がわかる」，これが「ビュフォンの針」といわれる発想です。針と円周率が結びつくなんて，とても不思議としかいいようがありません。

やさしい解説

平面上に平行線が幅$2a$の間隔でたくさん描かれているものとします。ここに長さ$2l$の針をデタラメに1本落としてみることにします。ただし，$l < a$とします。このとき，針が平行線に交わる確率pを求めてみましょう。



仮定$l < a$より針が2本以上の平行線と交わることはありません。すなわち，交わるとしても1本の平行線のみです。そこで，針の中心から，もっとも近い平行線までの距離をx，針と平行線のなす角をθとしてみます。



このとき，

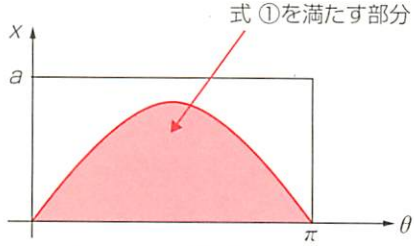
$0 \leqq x \leqq a$，$0 \leqq \theta \leqq \pi$

であって，針の位置はxとθによって完全に決定されます。針が平行線と交わるための条件は先の図からわかるように，

$x \leqq l \sin\theta$ ……①

が成立することです。

したがって，点(θ, x)が右図の長方形の中の斜線部分にあれば，針は平行線と交わります。



よって，求める確率は，

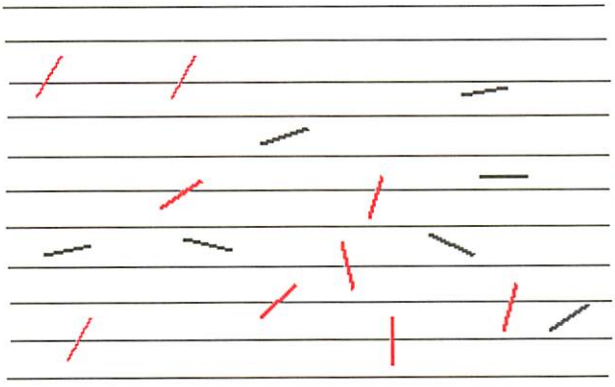
$$p = \frac{\text{斜辺部の面積}}{\text{長方形の面積}} = \frac{1}{a\pi}\int_0^{\pi} l\sin\theta\, d\theta = \frac{2l}{a\pi}$$

となり，この式から$\pi = \dfrac{2l}{ap}$ …②

となります。したがって，実際に針をたくさん投げることでpの近似値求めれば（統計的確率），円周率πの近似値を式②から求めることができます。

針を投げれば円周率πの値がわかる

1 横線に交わっている針の相対度数から円周率を求める



m本中n本の針が平行線と交わっていれば，

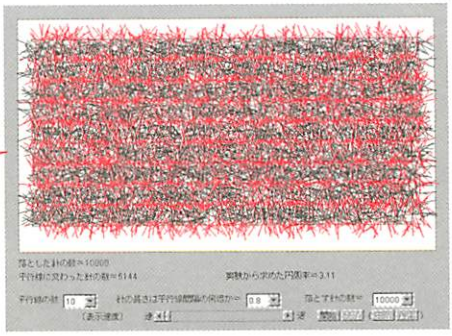
$p \fallingdotseq \dfrac{n}{m}$

よって，

$\pi = \dfrac{2l}{pa} \fallingdotseq \dfrac{2l}{a\dfrac{n}{m}} = \dfrac{2lm}{an}$

2 コンピュータシミュレーションで体験すると

針の長さを平行線の間隔の0.8倍にして，針を1000本落とした例。円周率は3.15ぐらい



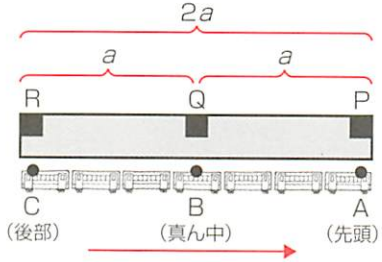
針の長さを平行線の間隔の0.8倍にして，針を10000本落とした例。円周率は3.11ぐらい

§40 乗車車両は何両目か

ここが要点！ 改札口がどこにあるかわからない駅で下車するとき，改札口までの到達時間を少なくするには，何番目の車両に乗り込めばいいのでしょうか。ただし，改札口はホームの端か中央にあるものとします。答えは，真ん中の車両ということになります。

やさしい解説

この種の問題では，改札口までの距離の期待値という考え方が有効です。改札口は，等確率で下図のP，Q，Rのいずれかの位置にあるものとします。



ここで，PQ間の距離をaとし，先頭A，真ん中B，後部Cの車両から改札口P，Q，Rまでの距離は0と見なして，改札口までの期待値を計算すると次のようになります。

(1) AまたはCに乗った人の期待値

$$0\times\frac{1}{3}+a\times\frac{1}{3}+2a\times\frac{1}{3}=a$$

(2) Bに乗った人の期待値

$$a\times\frac{1}{3}+0\times\frac{1}{3}+a\times\frac{1}{3}=\frac{2a}{3}$$

(1)，(2)より真ん中の車両に乗ったときが，改札口までの距離の期待値が一番小さいことがわかります。

注）存在確率の仮定（ベイズの確率，§24参照）

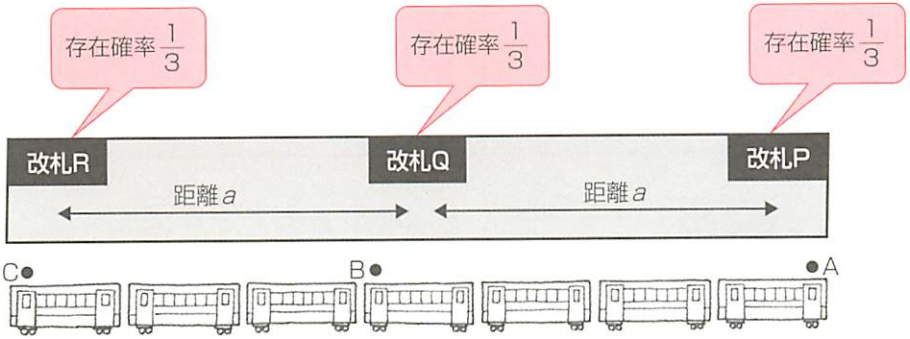
なお，ゲームにおいてはミニマックス（Mini-max）という考え方があります。これは，簡単にいえば「最悪の選択肢のうちで最善の対応を選ぶ」ということです。また，「相手が最善最強の手段で来るときに，なおかつ自らも最善の手段で対処しようとすれば，どうすればよいかという考え」ともみなせます。

この考え方だと，先程の乗り込み車両と改札口の問題では次のように考えられます。先頭車両や最後尾の車両に乗り込むと最悪の場合，ホームの端から端まで歩かねばなりません。

ところが，真ん中の車両に乗り込めば最悪の場合でもホームの半分の距離を歩くだけで済みます。そこで，真ん中の車両に乗り込むことがミニマックス解ということになります。

急ぐなら真ん中の車両

1 改札口が等確率でホームに存在すれば乗車車両は真ん中



Aから改札口までの距離の期待値は，

（距離AP×Pの存在確率）＋（距離AQ×Qの存在確率）＋（距離AR×Rの存在確率）

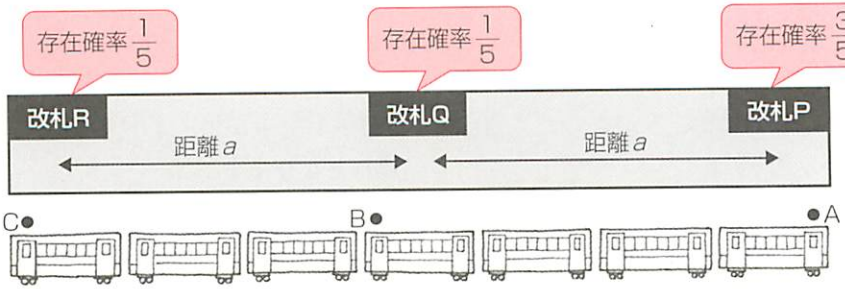
$= 0\times\frac{1}{3}+a\times\frac{1}{3}+2a\times\frac{1}{3}=a$

Bから改札口までの距離の期待値は，

（距離BP×Pの存在確率）＋（距離BQ×Qの存在確率）＋（距離BR×Rの存在確率）

$= a\times\frac{1}{3}+0\times\frac{1}{3}+a\times\frac{1}{3}=\frac{2a}{3}$

2 改札口が等確率でないと乗車車両は真ん中とは限らない



Aから改札口までの距離の期待値$= 0\times\frac{3}{5}+a\times\frac{1}{5}+2a\times\frac{1}{5}=\frac{3a}{5}$

Bから改札口までの距離の期待値$= a\times\frac{3}{5}+0\times\frac{1}{5}+a\times\frac{1}{5}=\frac{4a}{5}$

§41 封筒の宛名と中身が違う

ここが要点！ 「4通の手紙と宛先を書いた封筒がある。デタラメに1通ずつ封筒に入れるとき，どの手紙も違った宛先の封筒に入る確率はどのくらいか」，これがモンモールの問題です。不思議なことに，この問題をn通りに拡張して解いた答えの確率は，nを限りなく大きくしていくとネイピアの数e(=2.71･･･)の逆数に近づいていきます。

やさしい解説

1，2，3，……，nと書かれたn枚のカードを1列に並べ替えたとき，先頭から数えた順番と，カードの番号がどれも一致しない並べ方を**完全順列**といいます。このモンモールの問題は，まさに完全順列になる確率を求める問題と同じです。

したがって，手紙の問題を$n=4$の完全順列の場合で解くことにしましょう。

4枚のカード{1,2,3,4}の可能な並べ方は4!(=4×3×2×1)通りです。

○○○○ ……4!通り

1が1番目に並べられる場合の数は，2から4番目のところがまったく自由ですから3!通り。

①○○○ ……3!通り

したがって，1が1番目に並べられない数は4!から3!を引けばよい。

$\overline{①}$○○○ ……4!−3!通り…(イ)

次に，2が2番目に並べられる場合の数は，2が2番目に固定されているので3!通り。

○②○○ ……3!通り

このうち，1が1番目に並べられる場合の数は2!通り。

①②○○ ……2!通り

したがって，1が1番目に並ばず，2が2番目に並ぶ場合の数は3!−2!通り。

$\overline{①}$②○○ ……3!−2!通り…(ロ)

ゆえに，1が1番目に並ばず，2が2番目に並ばない場合の数は，(イ)−(ロ)より，

$(4!-3!)-(3!-2!)=4!-2\cdot3!+2!$

$\overline{①}\overline{②}$○○ ……4!−2･3!+2!通り

同様に考えていくと，すべて異なる場合の数は，

4!−4･3!+6･2!−4･1!+1･0!…(ハ)

通りとなります。

したがって，求める確率pは，式(ハ)を4!で割り次式となります。

$$p=1-1+\frac{1}{2!}-\frac{1}{3!}+\frac{1}{4!}$$

$$=0.375$$

完全順列になる確率は $1/e$ へ（eはネイピアの数）

1 $\{1,2,3,\cdots\cdots, k\}$ が完全順列になる確率 P_k

$k=1$　$p_1=1-1=0$

$k=2$　$p_2=1-1+\frac{1}{2!}=0.5$

$k=3$　$p_3=1-1+\frac{1}{2!}-\frac{1}{3!}=0.3333\cdots$

$k=4$　$p_4=1-1+\frac{1}{2!}-\frac{1}{3!}+\frac{1}{4!}=0.375$

$k=5$　$p_5=1-1+\frac{1}{2!}-\frac{1}{3!}+\frac{1}{4!}-\frac{1}{5!}=0.36667\cdots$

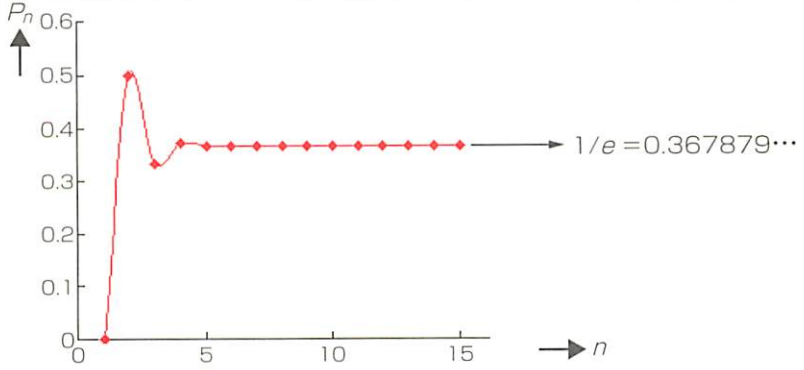
……

……

$k=n$　$p_n=1-1+\frac{1}{2!}-\frac{1}{3!}+\frac{1}{4!}-\frac{1}{5!}+\cdots+(-1)^n\frac{1}{n!}$

2 完全順列になる確率 P_n の変化

nを限りなく大きくすると，P_nは$1/e=0.367879\cdots\cdots$に近づく。



注1）e(=2.71828)はネイピアの数で，数学では円周率 π と並んで大事な数である。
2）完全順列は攪乱(かくらん)順列ともいう。
3）ネイピアの数eと完全順列になる確率が，円周率 π とビュフォンの針の確率（§39）が結びつく，実に不思議！

§42 酔っぱらいの行き着く先は

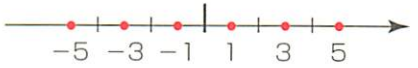
ここが要点！ 酔っぱらいは意識朦朧(もうろう)として，あっちへフラフラ，こっちへフラフラしています。しかし，よく観察すると結局，もといた位置を中心にふらついているだけで，さほど遠くへは行かないことがわかります。じつは，酔っぱらいの行き着く先は，理論的には二項分布に従うのです。

やさしい解説

酔っぱらいの足どりを詳しく調べてみることにしましょう。まずは酔っぱらいが数直線上でふらついているとしてみましょう。酔っぱらいは最初に原点にいたとして，そこから確率 p で右へ，確率 $q\,(=1-p)$ で左へ1歩移動するとします。移動後は，そこからまた同じ条件で右または左に1歩移動するとします。

このことを，たとえば5回繰り返した後は，酔っぱらいはどこに到達するのでしょうか。

原点から出発してランダムウォークを5回繰り返したときの行き着く先の座標は $\{-5,-3,-1,1,3,5\}$ の6通りです。



それぞれの座標に行き着く確率を求めると，次のようになります。

(1) 5に辿り着く確率

5歩中5歩とも右へ進むことになるので，確率は ${}_5C_5p^5q^0$

(2) 3に辿り着く確率

5歩中4歩は右へ，1歩は左へ進むことになるので，確率は ${}_5C_4p^4q^1$

同様にして，

(3) 1に辿り着く確率は，${}_5C_3p^3q^2$

(4) −1に辿り着く確率は，${}_5C_2p^2q^3$

(5) −3に辿り着く確率は，${}_5C_1p^1q^4$

(6) −5に辿り着く確率は，${}_5C_0p^0q^5$

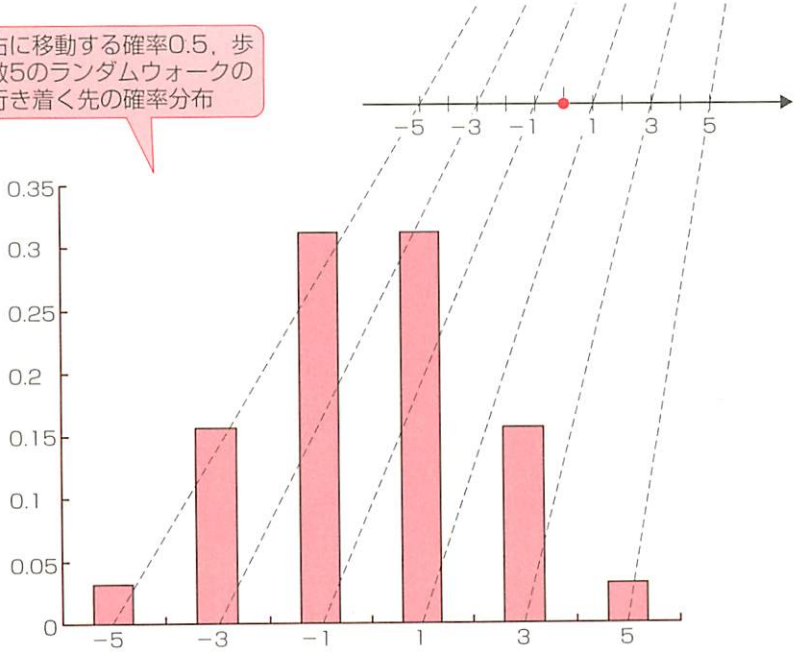
$p=q=0.5$ として確率を計算すると，原点周辺にたどり着きやすいことがわかります。

座標	確率
5	0.031
3	0.156
1	0.313
−1	0.313
−3	0.156
−5	0.031

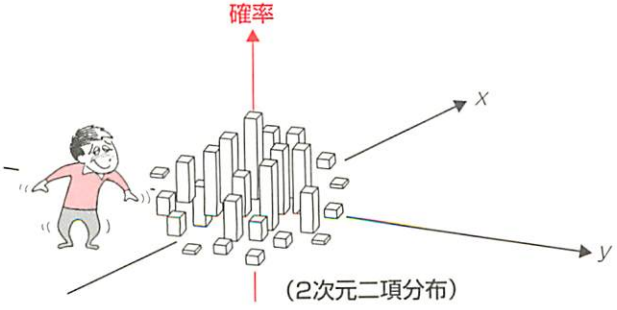
酔っぱらいは飲み屋の周辺をうろつく

1 1次元ランダムウォーク

右に移動する確率0.5，歩数5のランダムウォークの行き着く先の確率分布



2 2次元ランダムウォーク



（2次元二項分布）

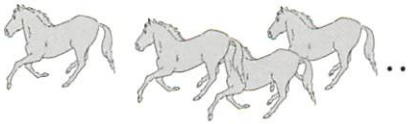
§43 野生動物の数は

ここが要点！ 広大な原野に住む野生動物の数を数えるのは，想像しただけでも大変なことです。もちろん１頭ずつ数え上げることは絶望的です。こんなときには確率の考え方で，その概数を探り当てることができます。

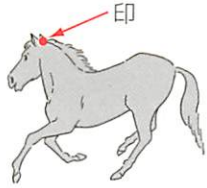
やさしい解説

野生動物の数を調べるいろいろな方法が開発されています。また，動物によって方法も違います。ここでは，それらの方法の中で簡単で基本的な**捕獲ー再捕獲法**と呼ばれるものを紹介します。これは，確率の考え方を使った方法で次の手順に従います。

まず，野生動物（例：馬）が生息している地域からランダムにn頭を捕獲します。捕獲の仕方はできる限り動物に傷をつけないように，また驚かせないようにしなければなりません。

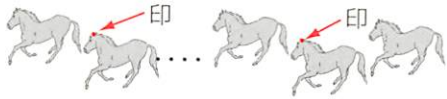


次に，捕獲したn頭の動物に何か適当な印をつけます。ここでも注意が必要です。印をつけたことによって，今後の生活に支障をきたしたり，他の動物に警戒心をおこさせたりしたら大変です。十分，慎重に行わなければなりません。



その後，印をつけたn頭の動物を再度，野に放ちます。

適当な時間の経過の後に，再び，ランダムに動物を捕獲します。その頭数をmとし，この中の印のある動物の数をa頭とします。



すると，野生動物全体の頭数をSとすると，相対度数に着目して次の式が成立します。

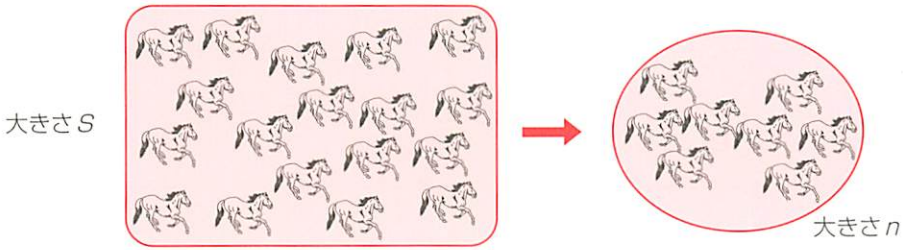
$$\frac{n}{S} \fallingdotseq \frac{a}{m} \quad \text{よって，} \quad S \fallingdotseq \frac{mn}{a}$$

ここで「$\fallingdotseq$」は「約」を意味します。

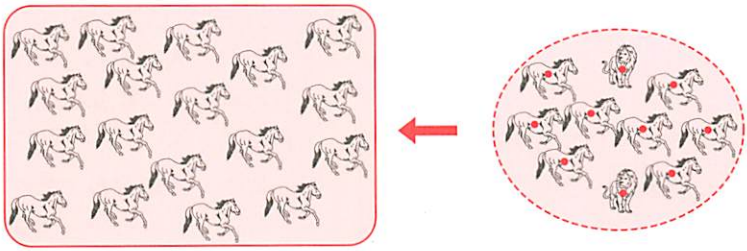
印のついた動物の数aによって，ここでの推定値は変動しますが，比率の推定の考え方を用いると，どのくらいの信頼度を持つかを算出することができます。

野生動物の数は「捕獲－再捕獲法」

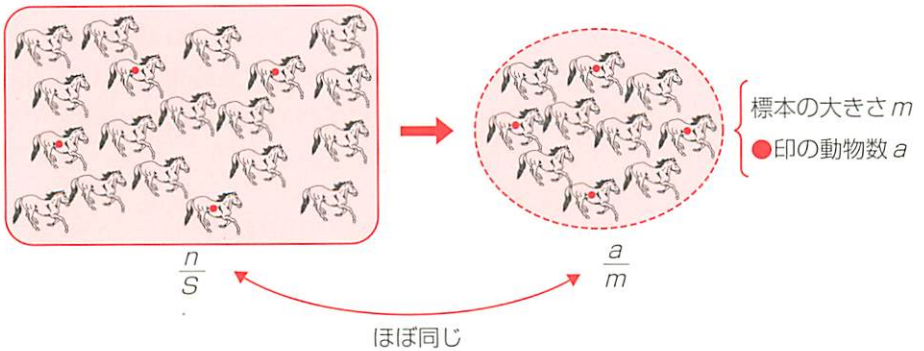
1 全体から大きさ n の標本を抽出する



2 標本の各個体に印をつけて復元する



3 再度，全体から大きさ m の標本を抽出する



4 全体の個体数は，約 $S \fallingdotseq \frac{mn}{a}$ である

§44 ミリオン分の1が1億円を産む

ここが要点！ 庶民のささやかな夢を語った言葉に「1億円当たったらなぁ」があります。そしてまた，この夢を実現する手段に「宝くじ」や「サッカーくじ(toto)」などがあります。そこで，この夢が真正夢になる確率を考えてみました。すると，何と100万分の1以下にすぎないのです。この確率，実感できますか？

やさしい解説

ある年の年末ジャンボ宝くじの1ユニット（くじ1000万本）についての当たりくじは，次のようになっていました。

1等	2億円	2本
1等の前後賞	5000万円	4本
1等の組違い賞	10万円	100本
2等	1000万円	2本
3等	100万円	40本
4等	1万円	1万本
…….	…..	….
…….	…..	….

通常，連番で何枚も買うことが多いので，1等に当たると3億円の賞金を入手することになります。このような人が2人出るし，2等，3等の人も多少いい思いをするわけですから，年末ジャンボ宝くじの場合，1枚300円のくじで1億円の夢が実現するのは，きわめてアバウトですが，約100万分の1ということになります。

それでは，サッカーくじではどうでしょうか。調べてみましょう。

Jリーグが開催する試合のうち，あらかじめセンターが指定した13試合の結果について，「勝ち」，「負け」，「その他の結果」のいずれかの予測をくじの購入者がします。このくじは1枚100円で，賞金は次のようになっています。

1等：13試合の予測がすべて的中した場合で，当選金は最高1億円，配当金は払戻金の50％

2等：……

3等：……

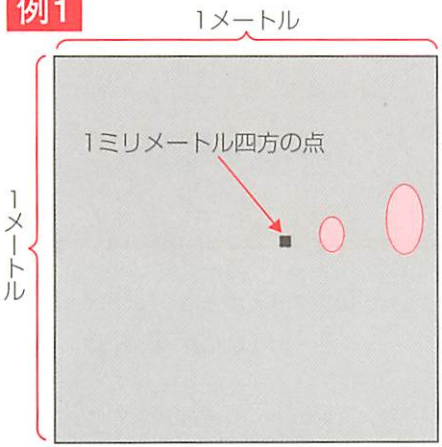
ここで，1等になる確率は，1試合について的中する確率が1/3で，13試合すべてに的中するのですから，次のようになります。

$$\left(\frac{1}{3}\right)^{13}=\frac{1}{1594323}$$

宝くじ1枚でサッカーくじを3枚買えるわけですから，これもきわめてアバウトに考えて，1億円の夢の実現確率は約100万分の1ということになります。

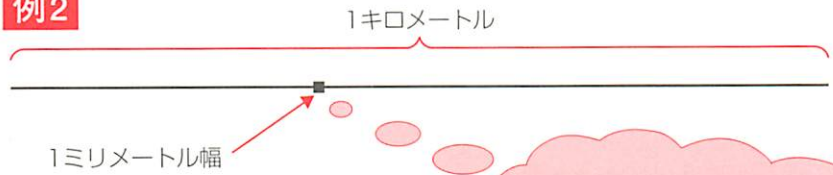
日本では200～300円で1億円がミリオン分の1の確率で当たる

例1



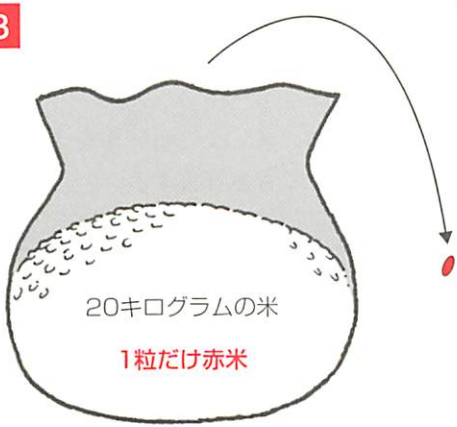
ミリオン分の1の確率は，1m四方の正方形にランダムに点を落としたとき，特定の1mm四方の正方形の中に入る確率

例2



ミリオン分の1の確率は，1kmのひもにランダムに点を落としたとき，特定の1mm幅の中に入る確率

例3



ミリオン分の1の確率は，20kgの米からランダムに1粒取り出すとき，特定の赤い1粒を取り出す確率

§45 最短コースにいろいろな数学

ここが要点！

次の３つは親密な関係にある。

（1）最短コースの数

（2）展開式の係数（二項係数）……パスカルの三角形

（3）二項分布

やさしい解説

$(a+b)^n$を展開すると，次のようになります。

$(a+b)^0=1$

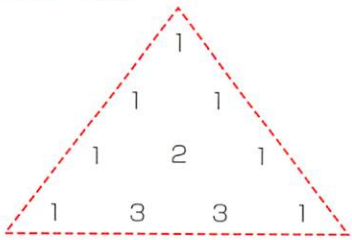
$(a+b)^1=a+b$

$(a+b)^2=a^2+2ab+b^2$

$(a+b)^3=a^3+3a^2b+3ab^2+b^3$

……

展開式の係数を書き出すと，次の三角形の関係が見えてきます。これが**パスカルの三角形**です。



上の2つの係数を足すと，下の係数になるという関係があります。

上記の展開式を組合せの記号${}_nC_r$を使って書くと，次のようになります。

$(a+b)^1={}_1C_0\,a+{}_1C_1\,b$

$(a+b)^2={}_2C_0\,a^2+{}_2C_1\,ab+{}_2C_2\,b^2$

$(a+b)^3={}_3C_0\,a^3+{}_3C_1\,a^2b+{}_3C_2\,ab^2+{}_3C_3\,b$

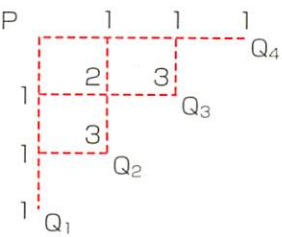
……

ここで，組合せ${}_nC_r$については，

${}_{n-1}C_r+{}_{n-1}C_{r-1}={}_nC_r$

が成り立ちます。これがパスカルの三角形の成立理由です。

また，パスカルの三角形は下図においてPからQ_1，Q_2，Q_3，Q_4までの最短コースの数に等しくなります。



なお，${}_nC_r$は二項分布を表現する大事な数でもあります。

×のとる値	0	1	2
確率	${}_nC_0p^0q^n$	${}_nC_1p^1q^{n-1}$	${}_nC_2p^2q^{n-2}$

……	r	……	n
……	${}_nC_rp^rq^{n-r}$	……	${}_nC_np^nq^0$

（二項分布表）

最短コース，二項係数，二項分布は仲よし兄弟

$(a+b)^0=1$

$(a+b)^1=a+b$

$(a+b)^2=a^2+2ab+b^2$

$(a+b)^3=a^3+3a^2b+3ab^2+b^3$

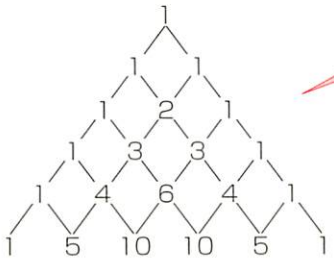
$(a+b)^4=a^4+4a^3b+6a^2b^2+4ab^3+b^4$

$(a+b)^5=a^5+5a^4b+10a^3b^2+10a^2b^3+5ab^4+b^5$

……

展開式の係数

パスカルの三角形



Pから各格子点までの最短コースの数

P	1	1	1	1	1
1	2	3	4	5	6
1	3	6	10	15	21
1	4	10	20	35	56
1	5	15	35	70	126
1	6	21	56	126	256
					Q

§46 「丁か半か」というけれど

ここが要点！ 丁とは２で割り切れる数，つまり偶数です。半とは２で割り切れない数，つまり奇数です。丁度の数が「丁」で，半端な数が「半」だと覚えておくといいでしょう。時代劇の賭博（とばく）では常用語でした。

やさしい解説

「丁か半か」は時代劇によく出てくる名セリフ。しかし，この意味を知っている人は意外に少ないようです。丁とは偶数，半とは奇数のことです。大昔の博徒でも，ちゃんと「偶数」，「奇数」の考えを持っていたとは驚きです。

しかし，これだけで「丁か半か」の意味が全部わかったわけではありません。時代劇での博徒はサイコロ２つを壺籠に入れてよく振り，畳に伏せて威勢よく「丁か半か」と叫びます。それではこのとき，何が丁で何が半なのでしょうか。

太郎君は，２つのサイコロの目の積が偶数か奇数かのことを「丁か半か」だと主張します。

花子さんは，２つのサイコロの目の和が偶数か奇数かで「丁か半か」の意味と捉えているようです。さて，どちらが正しいのでしょうか。

このことは，下図の表を見れば一目瞭然です。

表１は太郎君の考えです。丁が圧倒的に出やすいことがわかります。

	1	2	3	4	5	6
1	1(半)	2(丁)	3(半)	4(丁)	5(半)	6(丁)
2	2(丁)	4(丁)	6(丁)	8(丁)	10(丁)	12(丁)
3	3(半)	6(丁)	9(半)	12(丁)	15(半)	18(丁)
4	4(丁)	8(丁)	12(丁)	16(丁)	20(丁)	24(丁)
5	5(半)	10(丁)	15(半)	20(丁)	25(半)	30(丁)
6	6(丁)	12(丁)	18(丁)	24(丁)	30(丁)	36(丁)

〔表１〕

表２は花子さんの考えです。丁と半の出方が同等であることがわかります。

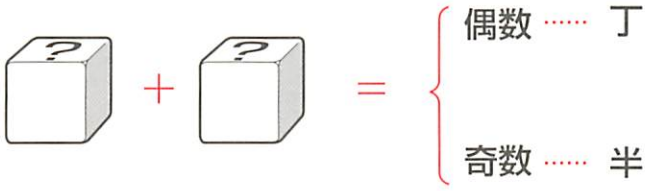
	1	2	3	4	5	6
1	2(丁)	3(半)	4(丁)	5(半)	6(丁)	7(半)
2	3(半)	4(丁)	5(半)	6(丁)	7(半)	8(丁)
3	4(丁)	5(半)	6(丁)	7(半)	8(丁)	9(半)
4	5(半)	6(丁)	7(半)	8(丁)	9(半)	10(丁)
5	6(丁)	7(半)	8(丁)	9(半)	10(丁)	11(半)
6	7(半)	8(丁)	9(半)	10(丁)	11(半)	12(丁)

〔表２〕

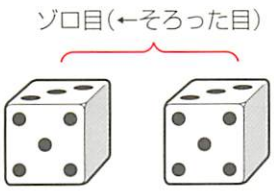
博打では丁か半が同等に出ないと話になりません。したがって，目の和に着目するのです。

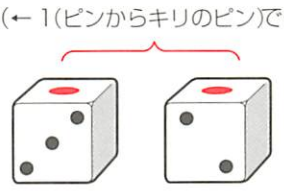
「丁か半か」は目の和が偶数か奇数か

1 丁か半か



2 ゾロ目とピンゾロ





3 サイコロ雑学

一か八か：一説には，「丁か半か」の「丁」「半」という字のそれぞれ上部をとったもの。また，「一か罰か」，すなわち「賽の目で一が出るか，しくじるか」による

ピンからキリ：「ピン」は「1」のことで，ポルトガル語ピンタが語源。「キリ」はやはりポルトガル語のクルス（＝十字架）」のこと。

裏目に出る：サイコロの1つの面とその裏側にあたる面との関係は，奇数と偶数の関係になっている。

三下：ダイス目が三以下の場合は勝てそうにない。そこから，目の出そうにない者をいう。

参考 **1の目はなぜ赤い**

もともと，サイコロの目はすべて黒かったそうです。あるサイコロ製造業者が目の丸をモチーフに1の目だけを赤色にして売り出したら大当たりし，それ以来，1の目が赤いサイコロが親しまれるようになったそうです。

§47 伝言ゲームとマルコフ過程

ここが要点！ 人から人へと耳打ちで伝言していくと，最後の人の受けとる伝言は，最初のものとは異なることはよく経験するところです。このことをマルコフ過程の理論で調べると，伝言が正しく伝わる確率は，初期の状態に関係なく決まってしまうことがわかります。

やさしい解説

「花子さんは美しい」という噂を人から人へ伝える場合を考えてみましょう。ただし，「美しい」と聞いたのに「美しくない」と次の人に伝えてしまう確率を1/5とします。また，「美しくない」と聞いたのに「美しい」と次の人に伝えてしまう確率を1/4とします。

この噂が，何人もの人を経て伝言されたら，最後の人はどのくらいの確率で「美しい」と聞かされ，または「美しくない」と聞かされるのでしょうか。

「花子さんは美しい」という伝言を聞くことをA，「花子さんは美しくない」という伝言を聞くことをBとします。このとき，次のようなマルコフ過程（§25参照）の系列ができます。

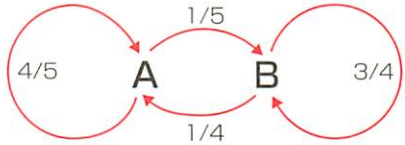
…，A，B，A，A，A，B，B，A，B，A，A，A，B，…

条件より，

AからBに推移する確率は1/5

BからAに推移する確率は1/4

したがって，マルコフ過程の推移図（§25参照）は次のようになります。



ここで，ある人がAである確率をα，Bである確率をβと，次の人がAである確率をx，Bである確率をyとすると，

$$x=\frac{4}{5}\alpha+\frac{1}{4}\beta,\ y=\frac{1}{5}\alpha+\frac{3}{4}\beta,$$

$$\alpha+\beta=1$$

となります。行列を使うと，

$$\begin{pmatrix}x\\y\end{pmatrix}=\begin{pmatrix}\frac{4}{5} & \frac{1}{4}\\ \frac{1}{5} & \frac{3}{4}\end{pmatrix}\begin{pmatrix}\alpha\\\beta\end{pmatrix}$$

と書けます。

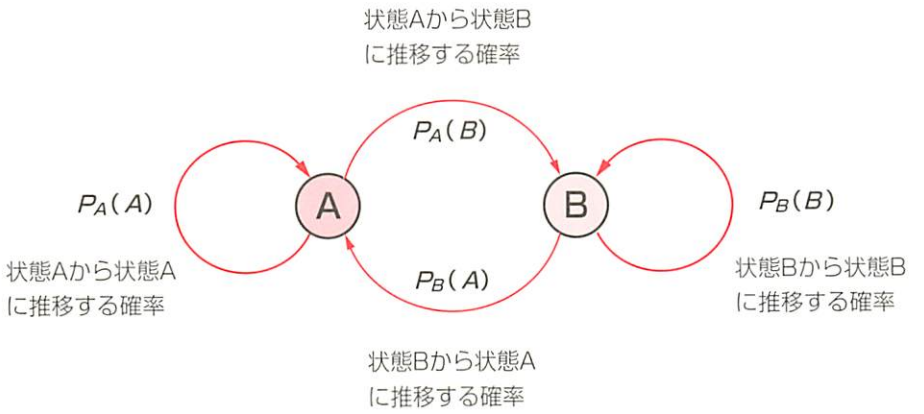
したがって，ある人からn人後の人がAである確率をx_n，Bである確率をy_nとすると，

$$\begin{pmatrix}x_n\\y_n\end{pmatrix}=\begin{pmatrix}\frac{4}{5} & \frac{1}{4}\\ \frac{1}{5} & \frac{3}{4}\end{pmatrix}^n\begin{pmatrix}\alpha\\\beta\end{pmatrix}\cdots①$$

ここで，nを限りなく大きくすると，式①よりαやβの値に関係なくx_nは5/9に，y_nは4/9に近づくことがわかります。つまり，噂が正しく伝わる確率は5/9ということになります。

マルコフ過程の確率関係は推移図で

1 シャノン線図（推移図）を表現した行列が推移行列



$P_A(A)$，$P_A(B)$，$P_B(A)$，$P_B(B)$を用いて行列を作る。

$$\begin{pmatrix} P_A(A) & P_B(A) \\ P_A(B) & P_B(B) \end{pmatrix} \quad \cdots\cdots\cdots \text{推移行列}$$

2 推移行列を使ってマルコフ過程の確率の推移状況が簡単に表現

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} P_A(A) & P_B(A) \\ P_A(B) & P_B(B) \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$$

ある時点で，状態がAである確率をα，状態がBである確率をβ，次の時点で，状態がAである確率をx，状態がBである確率をy

$$\begin{pmatrix} x_n \\ y_n \end{pmatrix} = \begin{pmatrix} P_A(A) & P_B(A) \\ P_A(B) & P_B(B) \end{pmatrix}^n \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$$

n回状態が推移した後にAである確率をx_n，状態がBである確率をy_n

§48 一口に「乱数」というけれど

ここが要点！ たくさんの数が羅列されているとき，それが乱数の列（乱数列）であるとは，次の２つの性質持っていることです。

（1）等確率性
（2）独立性

やさしい解説

「乱数」とは何かと聞かれると，多くの人々はデタラメの数と答えます。さらに「デタラメの数」とはどんな数かと聞かれると，「メチャクチャな数」と答えます。まさにハチャメチャな問答です。

数学では，数が羅列されたもの（数列）が乱数の列（乱数列）であるとは，次の2つの性質を持っているときと約束します。

(1) どの数でもその出現頻度は同じ……つまり「等確率性」

(2) 数字の出方が前の出方に影響されない……つまり「独立性」

これら2つの条件を満たす数列の例としては，サイコロを何回も振って出た目の数を書き出したものが考えられます。

6, 3, 5, 2, 2, 1, 4, 6, 3, 5, ……

なぜならば，サイコロはどの面が出ることも同様に確からしくなるように作られていると考えられます。したがって，(1) の「等確率性」が満たされます。また，ある回に出た目の結果は，次の回に振って出る目の数に影響を与えないと考えられます。つまり，(2) の「独立性」が満たされます。

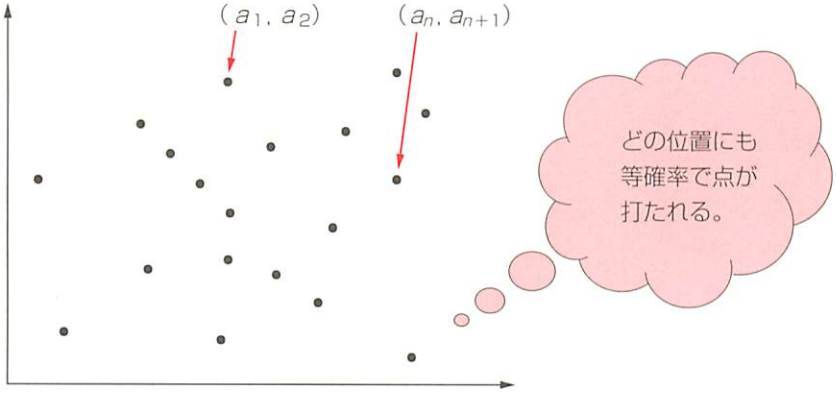
実際に，教科書やいろいろなテキストに載っている乱数表（下表）は，下図のような統計用乱数サイコロと呼ばれるものを使って作成されています。

80,86,47,12,25,14,48,50,96,25,21,37,30,76,64,95,90,40,56,69
49,19,98,41,96,12,29,49,39,19,32,44,42,25,62,12,23,18,46,69
84,57,90,47,87,34,98,63,22,30,38,21,68,62,15,52,30,37,14,30
27,33,78,45,49,22,15,37,99,21,77,35,20,95,38,74,47,85,44,14
60,43,60,50,22,50,19,91,94,49,80,42,38,82,73,60,61,46,15,61
89,57,25,69,24,59,86,83,47,22,51,53,64,59,17,64,79,98,25,94
51,29,60,33,17,14,13,45,24,35,54,71,89,38,37,82,21,64,13,48
60,64,69,98,52,88,32,57,46,73,70,81,84,29,30,89,29,95,48,67
12,73,32,12,27,89,16,54,79,44,30,71,50,97,81,96,79,59,16,68
54,22,37,35,82,55,82,72,62,98,63,79,83,70,73,48,45,88,64,29
13,62,33,70,64,49,90,15,93,76,83,53,83,53,14,17,97,68,44,40
42,28,61,57,68,88,95,28,81,33,79,71,83,49,22,87,77,36,28,87
75,67,97,72,74,91,16,48,94,50,30,39,27,95,73,61,30,15,44,98
95,93,68,16,86,72,36,49,21,12,89,54,50,64,98,54,24,34,80,14
37,65,34,22,47,31,74,94,56,93,30,33,86,35,15,87,28,93,47,34
61,55,12,62,27,36,62,39,61,90,90,68,55,15,89,63,52,22,39,98
83,32,35,96,39,41,16,69,77,12,82,75,56,65,39,27,25,43,77,25
15,80,43,75,73,58,27,63,30,29,79,66,61,82,75,69,84,36,95,63
29,60,72,43,22,65,18,93,79,28,23,86,90,67,85,84,82,46,50,67
61,90,78,85,98,84,80,47,13,87,81,94,42,21,83,72,61,84,28,71

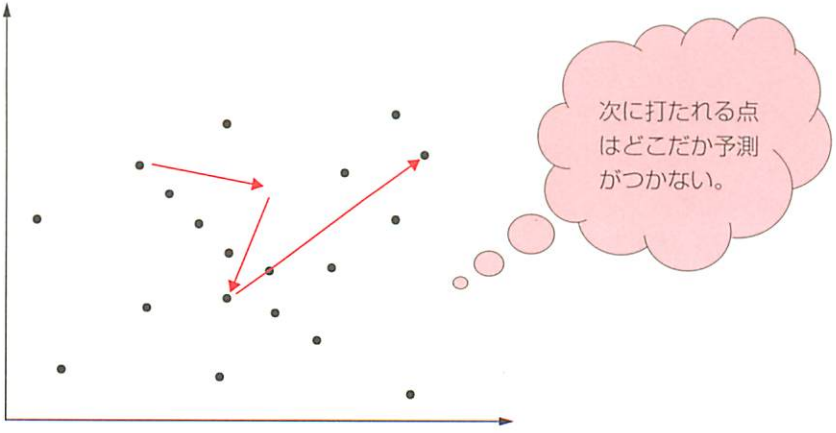
等確率性と独立性が乱数の条件

数列 $a_1, a_2, a_3, a_4, \cdots\cdots a_n, a_{n-1}, \cdots\cdots$ が乱数列であることを目で見て実感するには，この数列の奇数番目を x 座標，偶数番目を y 座標に持つ点 (a_n, a_{n+1}) を座標平面上にプロットするといい。ただし，各乱数 a_i の値は 0 以上 1 未満とする。

1 等確率性



2 独立性



§49 疑似乱数はこうやって作る

ここが要点！ 乱数を作る数式があるとすればおおいに矛盾です。しかし，結果からみると，それはあたかも乱数に見えることがあります。そこで，このような乱数を疑似乱数と呼ぶことにします。ここでは，疑似乱数の作り方を２つ紹介します。

（1）平均採中法
（2）混合合同法

やさしい解説

本物の乱数を大量に作成することは大変です。実際に，大量の乱数を必要とするときには，コンピュータを利用して乱数と見なせるような数（疑似乱数）を作成することになります。その際に使われるアルゴリズムはいろいろあります。ここでは単純でわかりやすい「平均採中法」と「混合合同法」を調べてみましょう。

(1) 平均採中法

① n桁の数字を与える
② 以下を繰り返す
　2乗して中央のn桁を取り出す

(例) $n=4$のとき

初期値 1275　2乗した数
1625625　→　2562
6563844　→　6384
40755456　→　7554
57062916　→　0629
……

乱数列が途中から0になってしまう可能性があります。

(2) 混合合同法

a，b，L（すべて正の整数）を与えて，以下の合同式を繰り返します。

$x_{n+1}=(ax_n+b)\pmod{L}$

つまり，n番目の乱数x_nをa倍してbを加えたものをLで割り，その余りを$n+1$番目の乱数x_{n+1}とするのです。

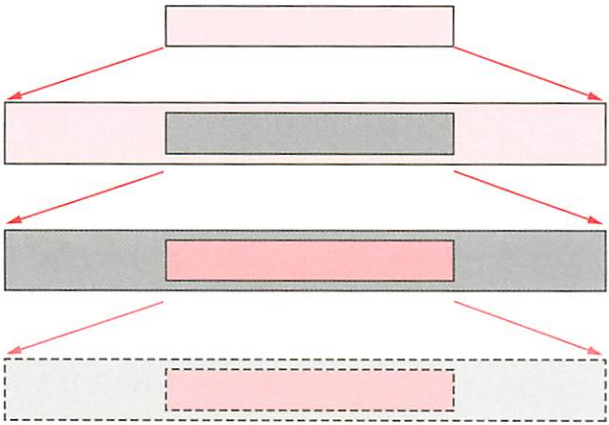
(例)

初期値　$x_1=75$
$a=17$　　$b=19$　　$L=100$
$17\times75+19=1294$　→　94
$17\times94+19=1617$　→　17
……

この方法は係数がまずかったり，初期値が小さいと乱数列と見なせなくなることがあります。aは素数か5の奇数乗，Lは10の乗数，初期値は大きくするとよいことがわかっています。

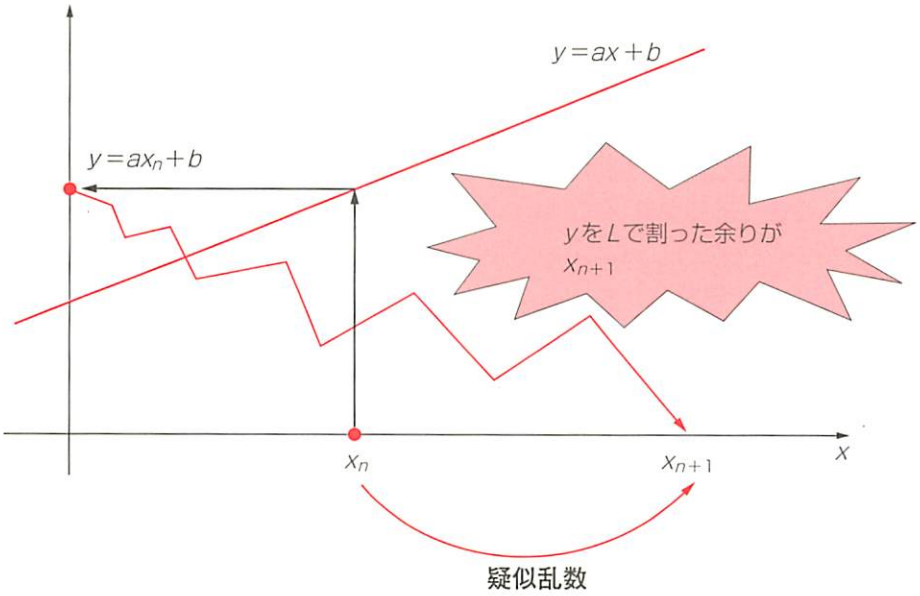
疑似乱数を作るには平均採中法，混合合同法などいろいろ

1 平均採中法（平方して真ん中をとる）



2 混合合同法

$x_{n+1}=(ax_n+b)\pmod L$を用いて変換!!



§50 一様乱数からいろいろな乱数

ここが要点！ 通常「乱数」というと一様乱数のことを指しますが，この乱数から正規乱数や指数乱数などいろいろな分布に従う乱数を作り出すことができます。

やさしい解説

乱数列で（0，1）上の一様分布に従うものを**一様乱数**（Uniform random number）といいます。

（平均=0.5，分散=$\frac{1}{\sqrt{12}}$）

この一様乱数をもとに，いろいろな分布に従う乱数を作成できます。

(1) 中心極限定理を利用する方法

中心極限定理（§69）を利用すれば，正規分布に従う乱数を簡単に作成できます。つまり，x_1，x_2，x_3，……，x_n を（0，1）上の一様乱数とすれば，

$$x=\frac{x_1+x_2+x_3+\cdots\cdots+x_n}{n}$$

は，平均が $\frac{1}{2}$，分散が $\frac{1}{\sqrt{12n}}$ の正規分布に従います。

(2) 逆関数を利用する方法

一様乱数を u とし，分布関数 $F(x)$ の逆関数を $G(x)$ とすれば，$G(u)$ が分布関数 $F(x)$ に従う乱数を生成します。

ここで，$y=F(x)$ の逆関数 $y=G(x)$ は次のように求めます。

① $y=F(x)$ を x について解いた式を $x=G(y)$ とします。

② x と y を交換し，$y=G(x)$ とします。この $y=G(x)$ が $y=F(x)$ の逆関数になります。

(例) 指数分布は確率密度関数が，

$$f(x)=\lambda e^{-\lambda x} \quad (x\geqq 0)$$

となります。この逆関数 $g(x)$ は，

$$g(x)=\frac{1}{\lambda}\log_e\frac{\lambda}{x} \quad (0<x<\lambda)$$

したがって，（0，λ）上の一様乱数 u を元に作成した

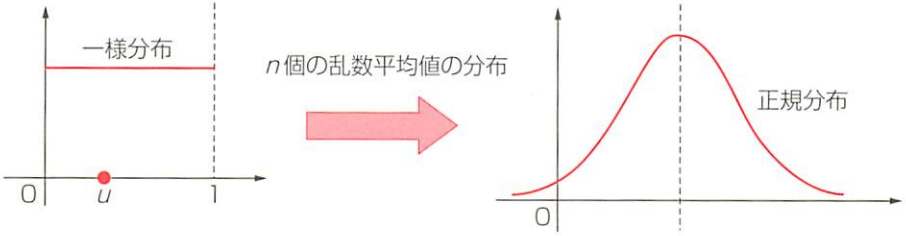
$$g(u)=\frac{1}{\lambda}\log_e\frac{\lambda}{u}$$

が指数分布に従います。

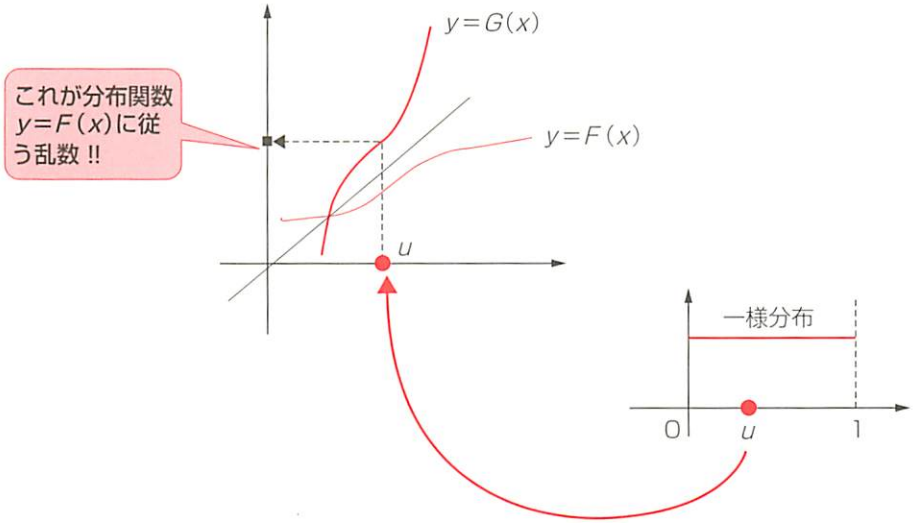
参考 $y=3x-2$ の逆関数は，まず，x について解いて，$x=\frac{y+2}{3}$ を求め，これを x について解いて，$y=\frac{x+2}{3}$ となります。

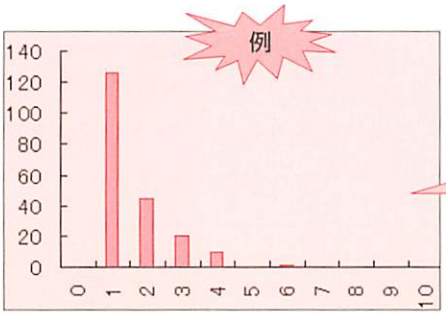
一様乱数と逆関数でいろいろな乱数生成

1 中心極限定理を利用



2 分布関数 $y = F(x)$ の逆関数 $y = G(x)$ を利用





確率密度関数が

$y = \lambda e^{-\lambda x}$ ……①

の指数分布に従う乱数は式①の逆関数

$$y = \frac{1}{\lambda} \log_e \frac{\lambda}{u}$$

で発生します。ただし，u は一様分布に従う乱数。

左図のグラフは $\lambda = 1$ として200個作成し，度数分布を調べた例である。

§51 確率の歴史

ここが要点！ 確率論の始まりは，17世紀にパスカルとフェルマーが賭事に関することで議論したことが始まりだといわれています。その後，いろいろな人々によって議論が深められ，今日の公理的確率論や確率過程論に至っています。

やさしい解説

フランスのパスカル（1623～1662）とフェルマー（1601～1665）が，サイコロゲームについて往復書簡で議論する中で確率論が始まったといわれています。そこでは，現在，高等学校で学ぶ「順列」，「組合せ」の考え方や二項分布の理論も展開されていました。その後，彼らの考えはベルヌーイ（1654～1705）やラグランジェ（1736～1813）らを経て，ラプラス（1749～1827）によって数学的（古典的）確率論としてまとめられました。それが「同様に確からしい」という前提のもとに，

$$\text{事柄}A\text{の起こる確率}=\frac{\text{事象}A\text{の起こる場合の数}}{\text{起こり得るすべての場合の数}}$$

と定義された確率の考え方です。しかし，前提条件である「同程度の確からしさ」を論理的に矛盾なく説明することは困難でした。

そこで考えられたのが統計的（先験的）確率といわれるものです。つまり，「Nが十分大きいとき，事柄Aの起こる相対度数r/Nがほぼ一定の値pに近づくならば，このpを事柄Aの起こる統計的確率という」というものです。

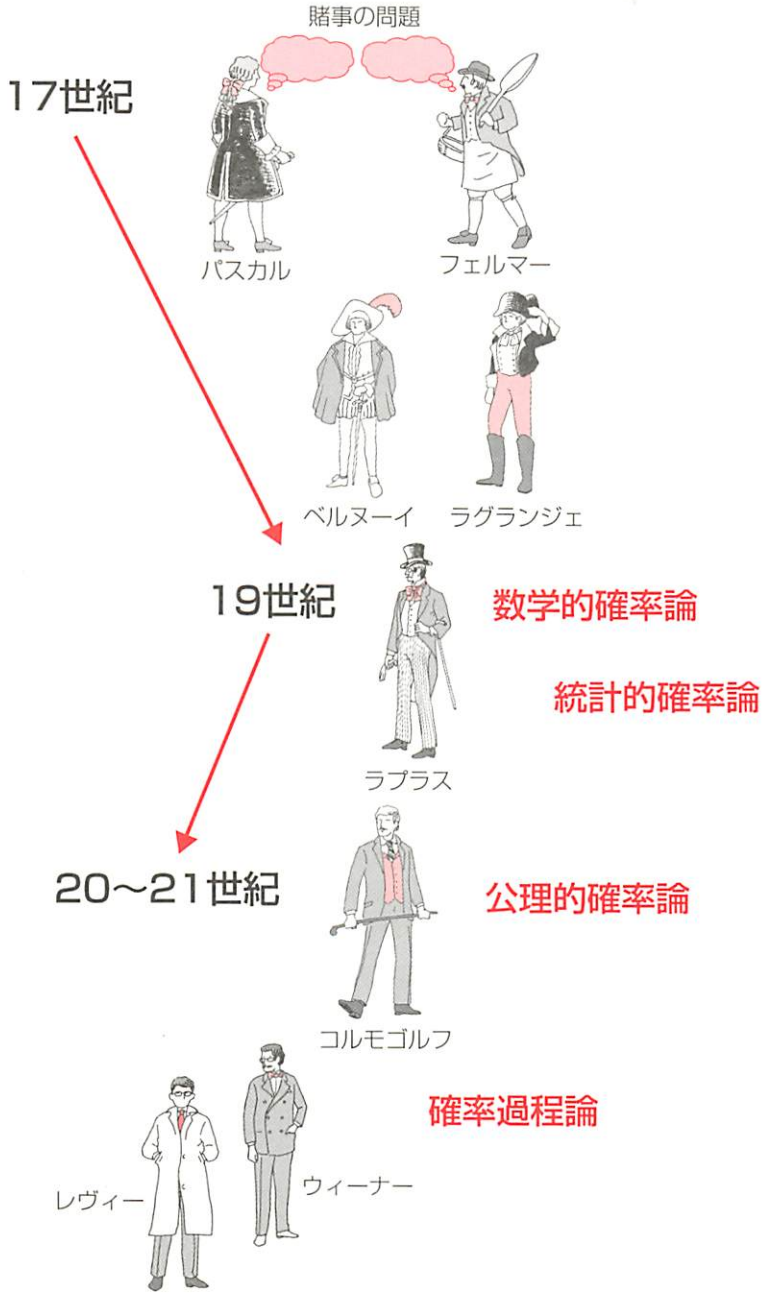
この統計的確率論と先の数学的確率論を理論的に結びつけるのが**大数の法則**（§18参照）なのです。

20世紀になってからは，ロシアの数学者コルモゴルフ（1903～）によって，確率論が測度論に基づいて公理的に構成されるようになりました。

測度論とは，長さや面積などを数学的に厳密にした理論です。公理的に構成するということは，いくつかの公理を決め，これをもとに演繹（えんえき）的に理論の大系を構築することです。ここでは，数学的確率の「同程度の確からしさ」の困難は解決されています。

現代の確率論の対象は「確率過程論」になっています。これは時間で変化する偶然現象を扱うもので，予測と制御がその中心になっています。

確率論はそんなに古くはありません



§52 データの山は宝の山

ここが要点！ データの山は，ただそれだけではゴミの山です。しかし，それらを整理しグラフ化するだけでも本質が見えてきます。また，一部のデータしかないときでも，それらから全体を推測することができます。つまり，標本調査です。異種類のデータについては，それらの関係を多変量解析で解き明かすことができます。このように，統計学の手法を駆使すれば，データの山を宝の山に変身させることができます。

やさしい解説

データの山は多くの人にとっては単なるゴミの山，見るのもうんざりという感じです。しかし，この山に統計処理を施すと，それは光り輝く宝の山に変身します。

もし，データの山が何らかの方法でコンピュータに取り込まれれば，それらのデータは，人間の指示通りに一瞬のうちに整理され加工されます。

たとえば，データが階級ごとに整理され，度数分布表や相対度数分布表が作成されるだけで，いろいろな情報が見えてきます。また作成した表をもとに，度数分布グラフや相対度数分布グラフを作成すれば，そこはもう幾何学の世界。データの特質が面積で表現されます。ここでは，微分積分という強力な道具を縦横に駆使できる世界でもあります。

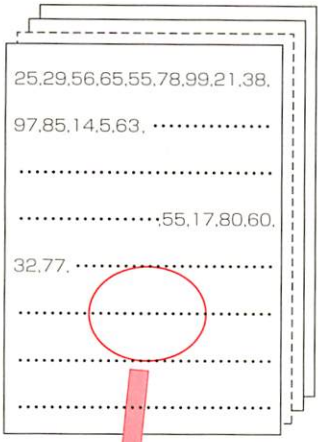
なお，データが多すぎて収集するのに困ったときは，全部を調べなくても済む方法があります。それが**標本調査**です。

つまり，全体からほんの一部を抽出し，それ（標本）を調べるだけで全体の特質を見抜いてしまう神業です。時間や経費が格安になるだけではありません。標本調査でないと調べられないこともあるのです。たとえば，2005年1月1日の20歳の人の平均体重などは，標本調査でしか調べられないことです。

また，データが複数の種類から構成されていれば多変量解析を利用し，これらの関係を解き明かすことができます。

たとえば，学科試験の成績と面接試験の成績，それに入社後の勤務成績があれば，多変量解析の手法でこれら3つの関係を解き明かすことができます。その結果，面接を重視すべきだとかの判定ができます。

データの山は最高の材料



多変量解析

関係が弱いな

学科試験	面接試験	勤務成績
65	70	60
50	90	80
…	…	…
…	…	…
…	…	…
…	…	…
…	…	…
80	40	30
67	70	80
80	95	95

関係が強いな

データの整理

階級	度数
0 ～ 10	63210
11 ～ 20	180250
21 ～ 30	378911
31 ～ 40	689778
41 ～ 50	765418
51 ～ 60	598740
61 ～ 70	358740
71 ～ 80	158874
81 ～ 90	15100
91 ～ 100	5219

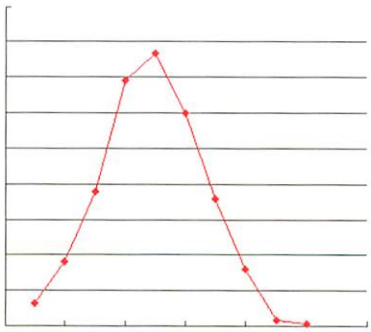
標本

統計的推測

検定

推定

グラフ化



§53 いくつに分類したらよいのか

ここが要点！ 資料を整理して度数分布表を作成する際，まずは階級数を決めることになります。階級数は，資料の背後にある母集団の特質がよくわかるように決めるのが基本です。階級数を決めるとき，よく使われるのがスタージェスの公式です。

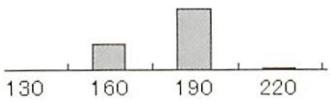
やさしい解説

資料を度数分布表にまとめる目的の1つは，ここでの資料を標本と見て，その背後にある母集団の形を想定することにあります。このためにはあまり細分すると，個々の値の変動が目立って，全体の形がわからなくなる恐れがあります。また，あまり大雑把(おおざっぱ)に分けると，個々の値の変動が埋没しすぎる恐れがあります。個々の変動を消して全体の姿が現れるような分類が好ましいのです。

以下は，実際の1000人分の身長データ（平均値165，標準偏差10）を階級数を4，9として，度数分布表と度数分布グラフを描いた例です。

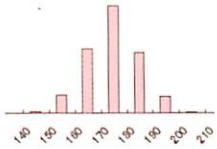
＜階級数４＞

データ区間	頻度
130	0
160	301
190	689
220	10



＜階級数９＞

データ区間	頻度
130	0
140	5
150	68
160	239
170	397
180	226
190	62
200	3
210	0
次の級	0



どうでしょうか，階級の数が4よりも9のほうが母集団の特質がはっきり表れている感じがします。

階級数を決めるのはなかなか難しいのですが，次のスタージェスの公式がよく使われています。

これは，Nをデータ数，Rをデータの範囲とするとき，階級幅ωと階級数nを次式で求めようというものです。

$$\omega = \frac{R}{1+\log_2 N}$$

$$n = 1+\log_2 N$$

階級数はスタージェスの公式で

1 階級幅をスタージェスの公式で求める

Nをデータ数，Rをデータの範囲とするとき，階級幅ωは次のスタージェスの公式より求めるとよい。

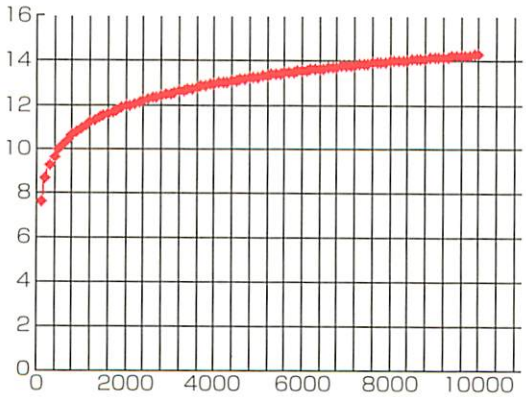
$$\omega = \frac{R}{1+\log_2 N}$$

階級			度数
0	～	10未満	25
10	～	20未満	40
20	～	30未満	50
30	～	40未満	100
40	～	50未満	180
50	～	60未満	120
60	～	70未満	70
70	～	80未満	40
80	～	90未満	25
90	～	100	20

階級数は，$1+\log_2 N$

階級幅 ω

2 スタージェスの公式によるデータ数と階級数の関係



注）スタージェスの公式は，データの数が多くなっても階級数はあまり増えないという欠点があるように思える。

§54 「一を聞いて十を知る」技

ここが要点！ 「全数調査」は，母集団の要素をすべて調べつくして，平均や分散などの母集団の特性を表す数（母数）を求めます。これに対して「標本調査」は，母集団のほんの一部を調べて母数を推定するものです。複雑で込み入った現代では，膨大な資料の特質を知りたいときには欠かせない手法です。

やさしい解説

日本に住んでいる人の身長や体重の平均，日本の労働者の所得の平均などを全部調べて，その平均を求めることは大変なことです。これらのデータを収集するだけでも莫大な費用がかかります。それに，調査時間も長期におよび，その間に個々のデータの値も変化してしまい，資料価値を失う可能性があります。

こんなときには，標本調査がその威力を発揮します。この方法は，「母集団からランダムに抽出した標本を調べることにより，母集団の平均（母平均）や比率などを推定する」ものです。

全国民（1億2千6百万人）を調べて，平均身長が170.5cmということを得るのが**全数調査**です。3000人の国民を調べて，「95％の正しさで国民の身長は169.55cm以上171.15cm以下である」と知るのが**標本調査**です。たった3000人を調べるだけで1億2千6百万人の特性を推定できるのですから，たいした手法です。

標本調査で一番注意しなければならないのが標本の抽出方法です。母集団全体からランダムに標本を抽出することが前提になっているからです。

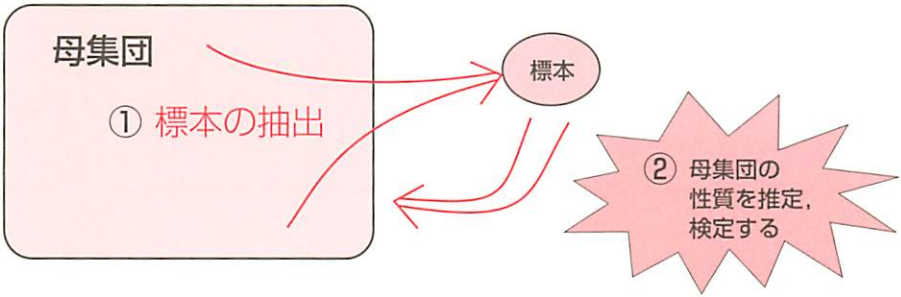
そのためには乱数表や乱数賽（サイコロ），コンピュータの発生する疑似乱数などを利用します。

また状況によっては，地域やデータの特性などを考慮して抽出しなければならないこともあります。そのための抽出方法もいろいろと考えられています。

典型的な方法には**比例抽出法**，**層化抽出法**，**多段抽出法**と呼ばれるものや，これらを組み合わせた**層化多段抽出法**，**層化比例抽出法**などがあります。いずれも母集団の特質が効率よく標本に反映されるように工夫されたものです。

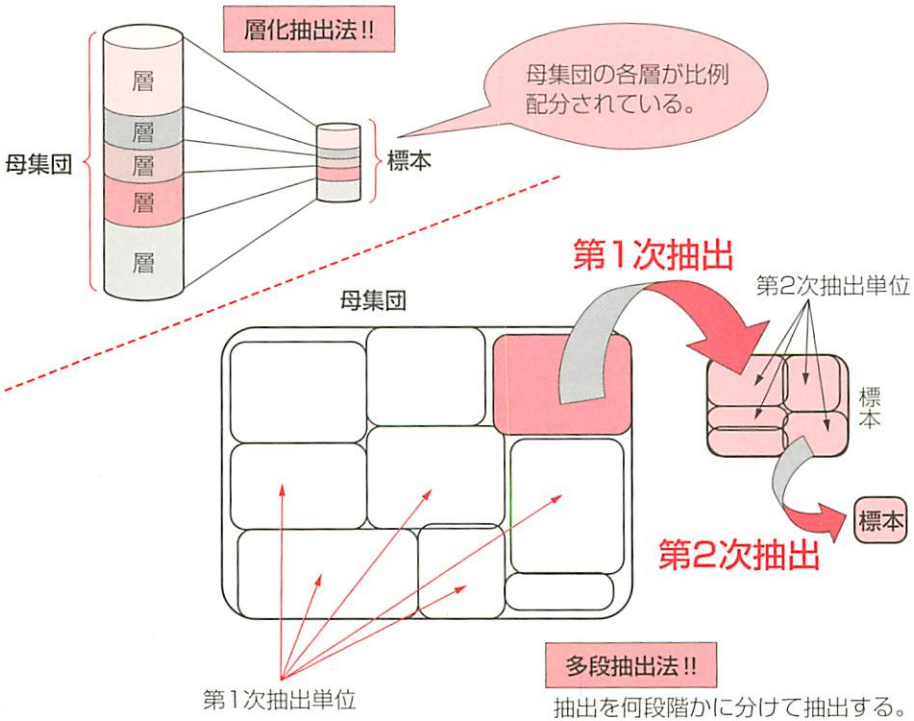
標本調査なくして世の中は語れない

1 標本調査は強力な知的戦略の武器



2 標本抽出は母集団全体からの無作為抽出が基本

標本抽出は母集団全体からの無作為抽出が基本だが，標本調査を効率よく行うためにいろいろな抽出法が考えられている。



§55 どのくらい取り出せばよいのか

ここが要点！ 標本調査では，どのくらいの大きさの標本を抽出すればよいか迷うところです。統計学に不慣れな人は，母集団の５％で十分だとか，１％では少なすぎるなどと考えがちです。ところが，母集団がある程度大きければ，抽出する標本の適正な大きさは，母集団の大きさによらずに決まってしまいます。

やさしい解説

標本調査における適正な標本の大きさは，母集団の何について調べるのかによって変化します。たとえば，大きさnの標本平均$\overline{X}$から，大きさNの母集団の平均mを推定する場合について調べてみましょう。

推定（§80）の考え方によると，「母平均mが危険率５％で$\overline{X}-\varepsilon$以上，$\overline{X}+\varepsilon$以下」になります。つまり，

$P(\overline{X}-\varepsilon \leqq m \leqq \overline{X}+\varepsilon) \geqq 0.95$ …①

といえる条件は，

$$n \geqq \frac{1}{\frac{\varepsilon^2}{4\sigma^2}\left(1-\frac{1}{N}\right)+\frac{1}{N}} \quad \cdots\cdots②$$

となります。ここで，σ^2は母分散です。母集団の大きさNが十分大きくて，$1/N \fallingdotseq 0$と見なせるのであれば式②は，

$$n \geqq \frac{4\sigma^2}{\varepsilon^2}$$

となり，Nに無関係であることがわかります。

次に，母比率Pの推定の場合について調べてみましょう。「母比率Pが危険率５％で$P-\varepsilon$以上，$P+\varepsilon$以下である」，つまり，

$P(p-\varepsilon \leqq P \leqq p+\varepsilon) \geqq 0.95$ …③

といえる条件は，

$$n \geqq \left(\frac{1.96}{2\varepsilon}\right)^2 \quad \cdots\cdots④$$

となります。ただし，pは標本比率。

この場合でも，標本の大きさは，母集団の大きさNには関係ありません。

以上のように，標本の大きさは，母集団の大きさと関係がないのです。

標本の大きさは母集団によらない

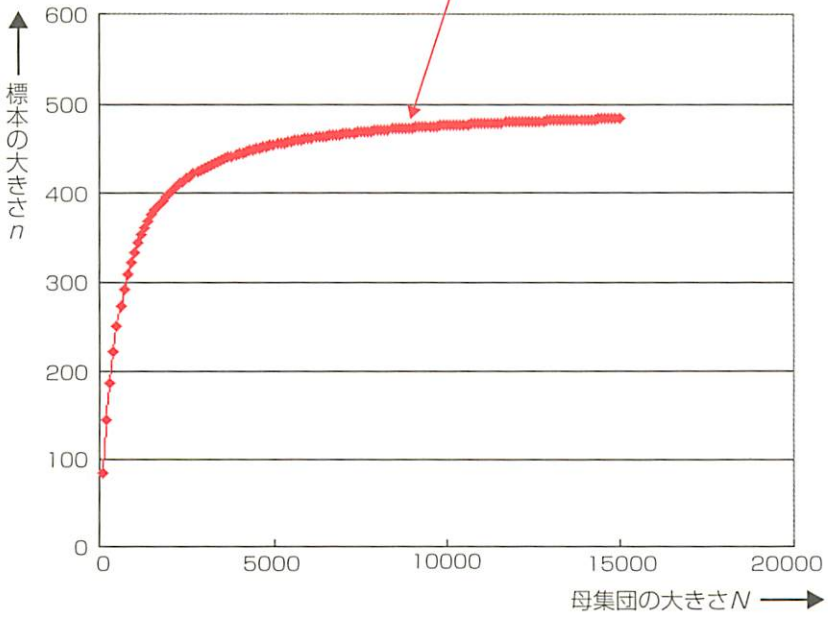
母平均 m の区間推定

$$P(\overline{X}-\varepsilon \leqq m \leqq \overline{X}+\varepsilon) \geqq 0.95$$

の場合の標本の大きさnと母集団の大きさNの関係

$$n=\frac{1}{\frac{\varepsilon^2}{4\sigma^2}\left(1-\frac{1}{N}\right)+\frac{1}{N}}$$ のグラフ

ただし，ここでは $\frac{\varepsilon^2}{4\sigma^2}=0.002$の場合

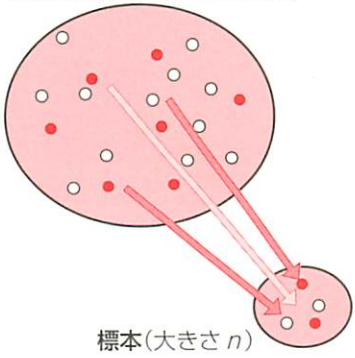


§56 もどすか，もどさないかの微妙な違い

ここが要点！ 復元抽出で求めたA候補者の支持率と，非復元抽出で求めた支持率は，当然，異なります。前者は二項分布に従い，後者は超幾何分布に従います。しかし，有権者の数，つまり母集団の大きさが大きいときには，復元でも非復元でも差を生じることはありません。

やさしい解説

大きさNの母集団から大きさnの標本を抽出し，その中で特性Aを持っている個体の個数をXとします。このとき，Xの値がkになる確率，つまり，$P(X=k)$は復元抽出か非復元抽出かによって異なります。このことを母比率をpとして調べてみることにしましょう。



なお，冒頭の例の場合「ある特性A」に相当するのが「A候補者を支持している」ということです。

(1) 復元抽出の場合

1個取り出しては元にもどす取り出し方で，n個取り出すことになる。このとき，特性Aを持ったものが何回目に取り出されるかによって${}_nC_k$通りの取り出され方があり，おのおのの確率は$p^k(1-p)^{n-k}$となります。

よって，

$$P(X=k)={}_nC_kp^k(1-p)^{n-k}$$

これは二項分布です。

(2) 非復元抽出の場合

母集団の中で特性Aを持っている個体数はNp個，持っていない個体数は$N-Np$個です。したがって，大きさnの標本に，特性Aを持っている個体数がk個である確率は次のようになります。

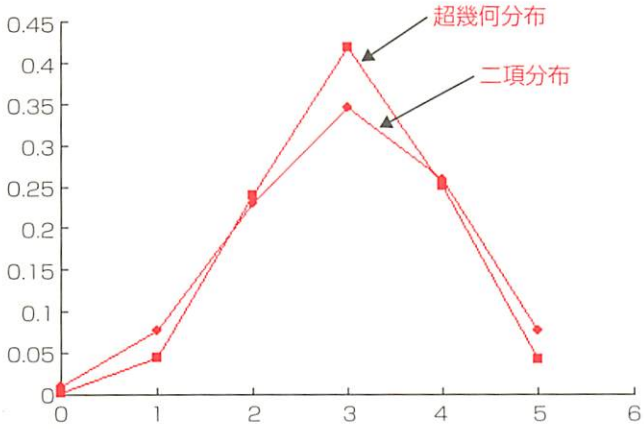
$$P(X=k)=\frac{{}_{Np}C_k\times{}_{N-Np}C_{n-k}}{{}_NC_n}$$

これは，**超幾何分布**といわれるものです。

復元抽出は二項分布，非復元抽出は超幾何分布

1 母集団の大きさが小さいと，復元抽出と非復元抽出がきわだつ

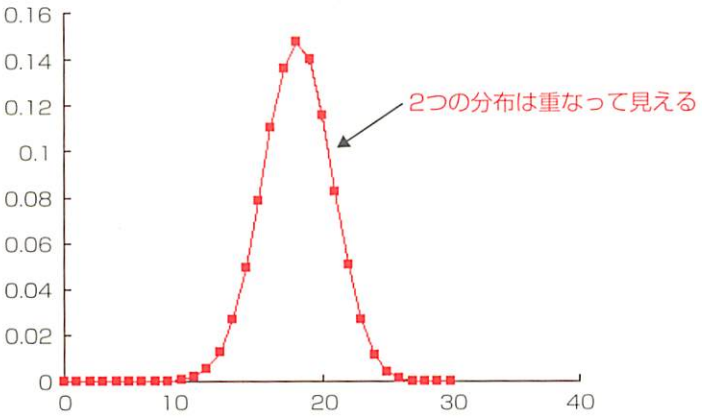
（例） $N=14$，$n=5$，$p=0.6$



2 母集団の大きさが大きいと，復元抽出と非復元抽出の違いは少ない

母集団の大きさNが大きいとき，
復元抽出と非復元抽出の違いはほとんどなし。

（例） $N=10000$，$n=30$，$p=0.6$



§57 一口に平均というけれど

ここが要点！ 平均といえば，平均点とか平均体重などを思い浮かべる人が少なくないでしょう。しかし，確率・統計の分野では，このほかにもいろいろな「平均」の考え方があります。加重平均，相乗平均，調和平均，移動平均，切り落とし平均などたくさんあります。

やさしい解説

n個の変量をx_1，x_2，x_3，…，x_nと表現するとき，次のようになります。

(1) 相加平均

n個の変量の総和をnで割ったもの。

$$\frac{x_1+x_2+x_3+\cdots+x_n}{n}$$

注)「平均」というと，通常はこの相加平均を意味します。

この相加平均は，単に平均とか平均値と呼ばれ，確率・統計ではすごく大事な役割を演じます。

(2) 加重平均

各変量の重みを考慮した平均を加重平均といいます。x_1，x_2，x_3，…，x_nのウェイトをw_1，w_2，w_3，…，w_nとすると，加重平均は次のように書けます。

$$w_1x_1+w_2x_2+w_3x_3+\cdots+w_nx_n$$

ただし，$w_1+w_2+w_3+\cdots+w_n=1$

ここで，

$$w_i=\frac{1}{n} \qquad (i=1,\ 2,\ 3,\ \cdots,\ n)$$

とすれば，加重平均と相加平均は等しくなります。

(3) 相乗平均（幾何平均）

n個の変量を乗じた積のn乗根。

$$\sqrt[n]{x_1x_2x_3\cdots x_n}$$

これは，増加率など比率が問題となっているときに用いられることが多い。

(4) 調和平均

$$\frac{n}{\frac{1}{x_1}+\frac{1}{x_2}+\frac{1}{x_3}+\cdots+\frac{1}{x_n}}$$

(5) 移動平均

データが激しく変動して，全体的な動きを捉えにくい場合，細かい変動をならし，傾向を求める考え方が移動平均です。3期移動平均値は次のようになります。

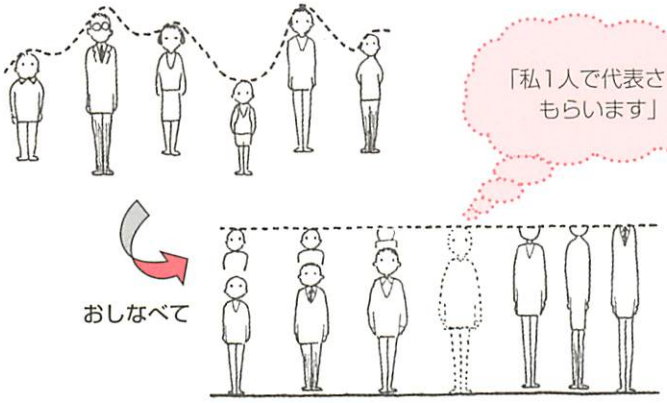
$$\frac{x_{i-1}+x_i+x_{i+1}}{3}$$

(6) 切り落とし平均

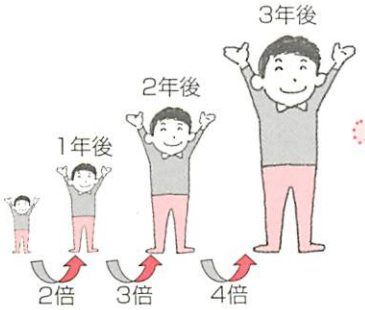
最小値や最大値などの分布の両端にある部分や，異常値を除いた相加平均のことです。

一口に平均といってもいろいろある

1 相加平均



2 相乗加平均



3年間の平均成長率は？

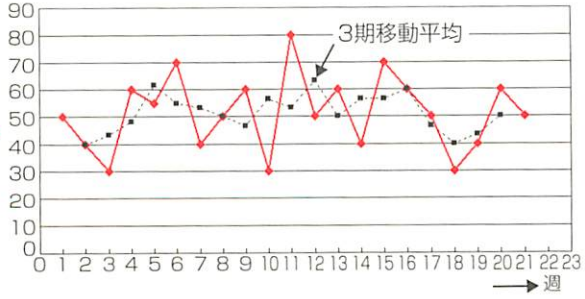
$\frac{2+3+4}{3}=3$ ……×

$\sqrt[3]{2\times3\times4}=2.8845\cdots$ ○

注）$(2.8845)^3=2\times3\times4$

3 移動平均

このグラフは週ごとの売り上げの折れ線グラフ(実線)と，3期移動平均のグラフ(破線)である。



§58 分散は世界の多様性のあかし

ここが要点！ 統計学では，資料情報は平均と分散に要約されます。このため，平均と分散をもとに統計理論は成り立っているともいえます。ここで，平均はどんな数値でもそれなりに意味があります。しかし，分散が0または0に近い世界はいけません。その世界には，統計理論が入り込む余地はないのです。

やさしい解説

n個の変量$\{x_1, x_2, \cdots, x_n\}$に対して平均$\overline{x}$はよく知られているように，次式で定義されます。

$$\overline{x}=\frac{x_1+x_2+\cdots+x_n}{n} \quad \cdots\cdots①$$

分散s^2は平均$\overline{x}$を用いて，次式で定義されます。

$$s^2=\frac{(x_1-\overline{x})^2+(x_2-\overline{x})^2+\cdots+(x_n-\overline{x})^2}{n-1} \quad \cdots\cdots②$$

これは，平均からの散らばり具合を数値で表現するために「変量と平均の差を2乗し，違いを拡大したものの平均」を求めたものです。

注）ここでは，分散の分母を$n-1$としました。他の本によってはnとしている場合もあります。これは「母集団の標本を調べているのか」，「それとも母集団そのものを調べている」のかの違いです。つまり，母集団の標本を調べているときには$n-1$で割り，母集団を調べているときにはnで割ります。

身長や体重の平均は，あきらかに正の数です。測定値と真の長さとの差の平均は0になることがあります。また，南極の平均気温はあきらかに負の数です。このように，平均はいろいろな世界で，いろいろな値をとることができます。このとき，平均が0であることは特別な意味を持ちません。

ところが，分散が0ということは，これはきわめて特別なことになります。分散が0ということは式②の分子が0，つまり，式②の分子の（　）内がすべて0ということです。したがって，

$$x_1=x_2=\cdots=x_n=\overline{x}$$

となり，n個の変量$\{x_1, x_2, \cdots, x_n\}$が互いにすべて等しいということになってしまいます。

つまり，みな同じということです。身長の分散が0とは，みな同じ身長なのです。何とも味気のない世界ではないでしょうか。残念ながらこのような世界では確率・統計は意味を失います。

分散０は個性が否定された画一的な世界

分散が大きい資料には豊富な情報量

No	x
1	50
2	50
3	50
4	50
5	50
6	50
7	50
8	50
9	50
10	50
平均	50
分散	0

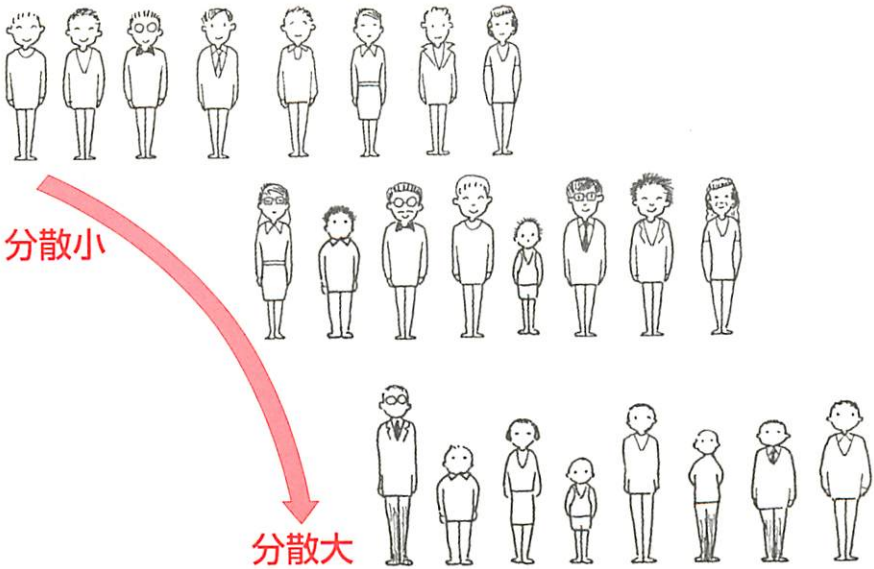
「分散0」
つまらない

No	x
1	60
2	45
3	50
4	30
5	60
6	50
7	40
8	45
9	70
10	50
平均	50
分散	128

「分散中」
個性ほどほど

No	x
1	20
2	45
3	50
4	95
5	80
6	55
7	5
8	15
9	50
10	85
平均	50
分散	928

「分散大」
個性豊か



§59 分布の特徴は平均と分散に集約

ここが要点！ 統計学で最も重要な数値といわれたら，「平均」と「分散」の２つがあげられます。なぜならば，分布が統計学の基本であり，分布の特徴は平均と分散に集約されるからです。推定，検定，多変量解析，いずれの場合でも平均と分散をもとに語られています。

注）標準偏差は，分布の正の平方根ですから分散と兄弟です。

やさしい解説

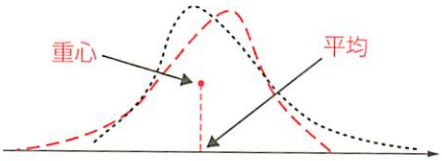
分布の特徴を表す数値には，いろいろなものがあります。たとえば，代表値（1つの数値で分布の中心を表す値）としてレンジ，メジアン，モード，平均，……があります。また，散布度（分布の広がり具合を表す値）としてレンジ，分散，標準偏差，四分偏差，平均偏差，……があります。

いずれも，それなりに意義がありますが，統計学で頻繁に利用し，かつ重要なのは平均と分散です。

n個の変量をx_1，x_2，x_3，…，x_nと表現するとき，

$$\overline{x}=\frac{x_1+x_2+x_3+\cdots+x_n}{n}$$

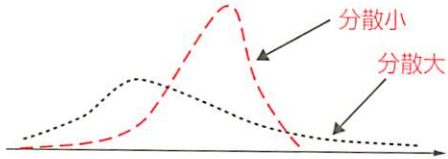
はこれらの変量の平均でした。これは，分布の重心を表します。



また，分散は次の値でした。

$$\sigma^2=\frac{(x_1-\overline{x})^2+(x_2-\overline{x})^2+\cdots+(x_n-\overline{x})^2}{n}$$

これは，個々の変量の値と平均との差を2乗し，違いを大きくしてから平均をとったもので，分布の広がり具合を表します。

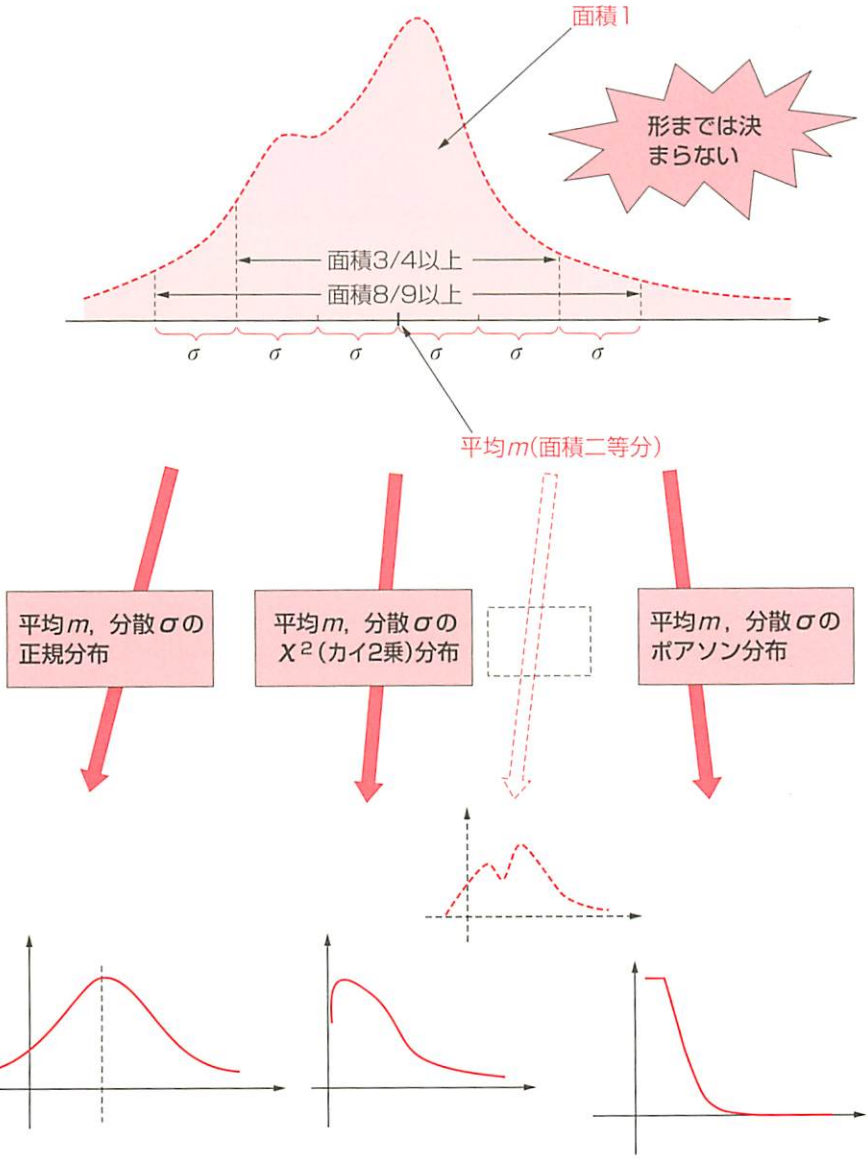


注）標準偏差も非常に大事な数値ですが，これは分散の正の平方根ですから，分散からすぐに求めることができる。

平均と分散がわかると，分布のようすがおぼろげながら見えてきます。その際，チェビシェフの定理なども参考になります。

注）チェビシェフの定理とは，「どんな分布でも，平均mを中心として，$\pm k\sigma$の範囲に全度数の$1-\frac{1}{k^2}$以上が含まれる」という定理である。詳しくは§74参照。

平均が m，分散が σ といわれると，分布がボンヤリ見えてくる



注）母集団の分散を σ^2 と表します。

§60 3つの代表値の面白い関係

ここが要点！ 分布の特徴を1つの数値でいい当てるのが代表値です。統計学で主として用いられる代表値には，「平均値」「メジアン」「モード」があります。これらは分布グラフで見ると，それぞれ特徴ある位置を占めています。

やさしい解説

n個の観測値x_1，x_2，x_3，…，x_nが与えられたとき，

$$\overline{x}=\frac{x_1+x_2+x_3+\cdots+x_n}{n} \quad \cdots①$$

を**平均値**といいます。平均身長などはまさに平均値のことです。この値は「確率分布のグラフで見ると，そのグラフの重心の横座標」となります。

注）連続的確率変数Xの確率密度関数（§62参照）を$f(X)$とするとき，式①は，

$$\overline{x}=\int_a^b x f(x)\,dx$$

と積分で計算されます。

メジアンは，**中央値**とか**中位数**などと訳されているように，資料を大きさの順に並べたとき，中央の順位にくる値のことです。

資料の個数が奇数であれば，中央の順位にくる値は必ず存在しますが，偶数のときには中央の順位が存在しません。このときには，中央の順位にくる相隣り合う2つの値の平均をメジアンとします。メジアンは**50パーセンタイル**とも呼ばれています。メジアンは記号M_eで表されることが多く，これを「確率分布のグラフで見ると，グラフの面積を左右に二分する直線の横座標」ということになります。

モードは**最頻値**とか**並数**などと訳されているように，資料のうちもっともしばしば現れる数値のことです。モードは記号M_0で表されることが多く，「確率分布のグラフで見ると，グラフの頂上部分の資料の値」になります。

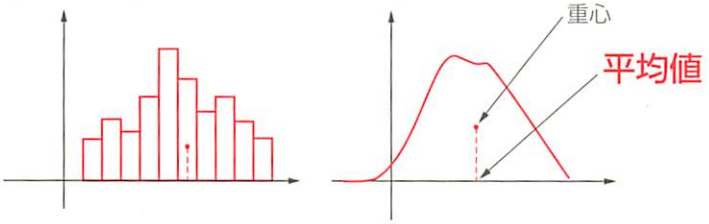
統計学では，平均値，メジアン，モードの中では平均値が一番よく使われます。しかし，メジアン，モードも欠かすことのできない代表値です。

たとえば，貯蓄額などは平均値よりもメジアンのほうが実態にあった数値になるようです。

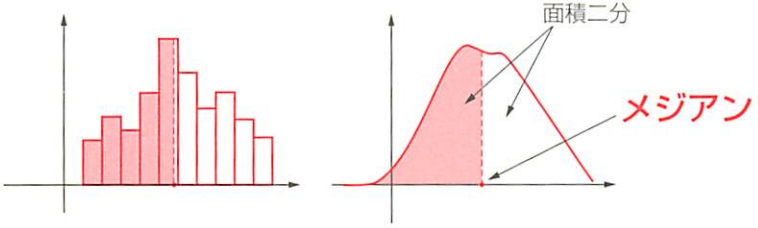
また，洋服のメーカーは，平均値よりもモードに合わせて売れる商品を開発しているようです。平均値のサイズの服だと，誰にも合わなかったということが起こり得るのです。

「平均値」「メジアン」「モード」を確率分布で見ると

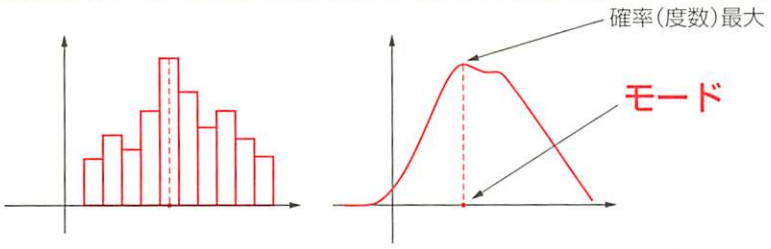
1 平均値は重心



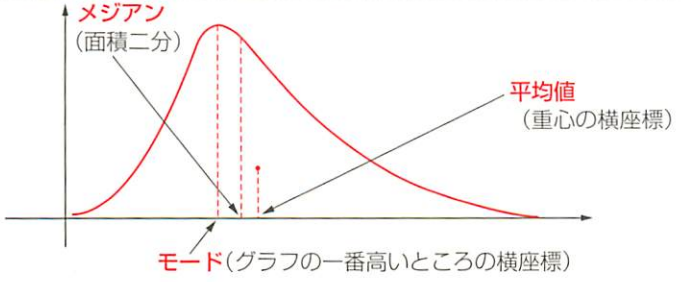
2 メジアンは面積二分



3 モードは一番高いところ



4 平均値とメジアンとモードを1つのグラフで見ると



§61 分散はn割，それとも$n-1$割

ここが要点！ 母集団の分散を調べているときには，分散は母集団の大きさNで割りました。ところが，標本の分散を調べているときには標本の大きさnでは割らずに$n-1$で割りました。この理由は$n-1$で割らないと，すべての標本についての分散の期待値が母集団の分散に一致しないからです。

やさしい解説

大きさNの母集団$\{a_1, a_2, \cdots, a_N\}$についての分散$\sigma^2$は，次の計算で求めます。ただし，$m$は母平均。

$$\sigma^2 = \frac{(a_1-m)^2+(a_2-m)^2+\cdots+(a_N-m)^2}{N} \quad \cdots\cdots①$$

つまり，分母は母集団の大きさNです。ところが，この母集団から抽出した大きさnの標本$\{x_1, x_2, \cdots, x_n\}$についての分散$s^2$は，次の計算で求めたものを利用します。ここで，$\overline{x}$は標本平均です。

$$s^2 = \frac{(x_1-\overline{x})^2+(x_2-\overline{x})^2+\cdots+(x_N-\overline{x})^2}{n-1} \quad \cdots\cdots②$$

つまり，分母は標本の大きさnではなく$n-1$なのです。

この理由は「$n-1$で割ったものを利用しないと，すべての可能な標本の分散の平均値（期待値）が母分散σ^2に一致しない」からです。

（例）母集団$\{1, 2, 3, 4, 5\}$の平均は3で，この分散は式①より2となります。この母集団から復元抽出で大きさ2の標本（$5\times5=25$通り）をとり，式②（分母は1）より求めた各標本分散は下表の通りです。

（2，3）という標本の式②で求めた分散

1個目＼2個目	1	2	3	4	5
1	0	0.5	2	4.5	8
2	0.5	0	0.5	2	4.5
3	2	0.5	0	0.5	2
4	4.5	2	0.5	0	0.5
5	8	4.5	2	0.5	0

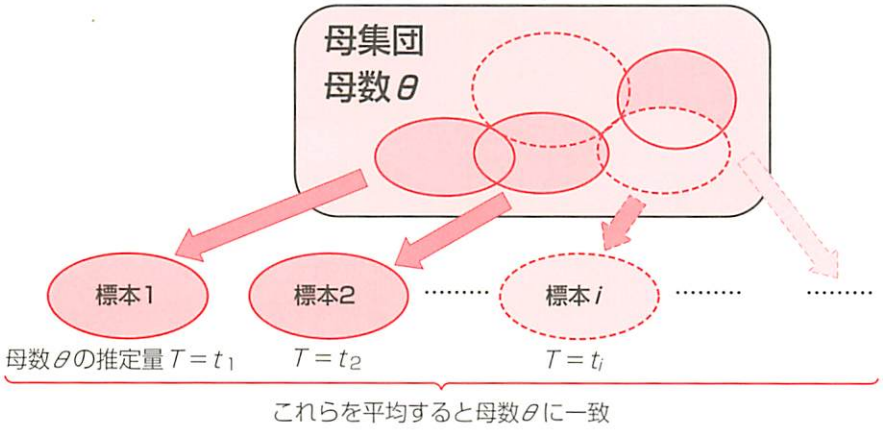
この25通りの標本分散の平均値は母集団の平均値2と一致します。式②の分母をnとすると一致しません。

このように，「標本すべてについてのある推定量の平均値が母集団の母数（母平均や母分散など）に一致する」とき，その推定量を**不偏推定量**といいます。式②で定義される標本分散は，まさしく母分散の不偏推定量です。そのため，式②を**不偏分散**といいます。

$n-1$ で割った標本の分散は不偏推定量

1 不偏推定量

標本から求めた推定量 T の期待値が，母集団の母数（母集団の特徴を表す数）と一致するとき T を不偏推定量という。

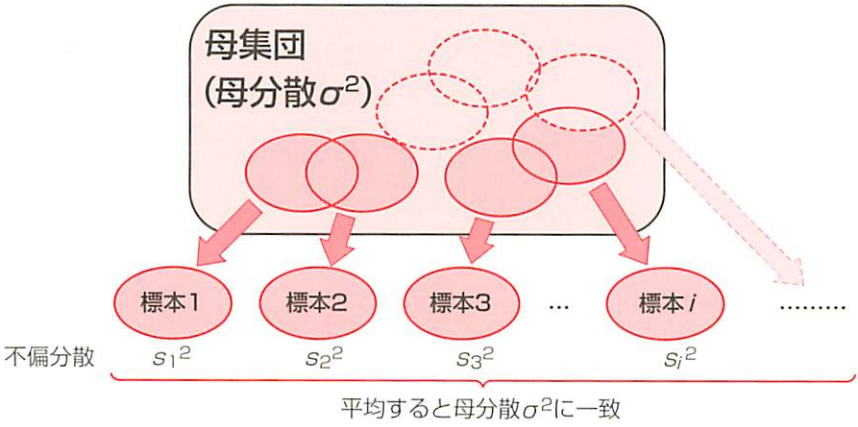


2 不偏分散

$n-1$ で割った標本の分散

$$s^2=\frac{(x_1-\bar{x})^2+(x_1-\bar{x})^2+\cdots+(x_N-\bar{x})^2}{n-1}$$

は母分散の不偏推定量。したがって，これを不偏分散という。



§62 人口密度があるように確率密度あり

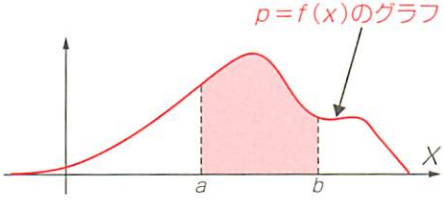
ここが要点！ 人口密度とは，「単位面積当たりに居住する人の数」のことです。これによって，人口の多寡が簡単に比較できます。確率にも同様なものがあります。それは「確率密度関数」と呼ばれるもので，連続型の確率変数がどんな値をとりやすいかどうかを示したものです。確率・統計を学ぶ上で重要な用語です。

やさしい解説

連続的な確率変数 X の関数として，

$p=f(X)$

があり，$a \leqq X \leqq b$ の値をとる確率 $P(a \leqq X \leqq b)$ が下図のうす色部分の面積で表されるものとします。



このとき，関数 $f(X)$ を確率変数 X の**確率密度関数**といいます。

確率密度関数 $f(X)$ は，次の条件を満たします。

(1) 定義域内の任意の x に対して，

$f(x) \geqq 0$

(2) 定義域内で $y=f(x)$ のグラフと x 軸で囲まれた図形の面積は1。

逆に(1)と(2)を満たす関数は，すべて確率密度関数になり得ます。

なお，上図におけるうす色部分の面積は積分を使って，

$$\int_a^b f(x)\,dx$$

と表せますから，次のようにまとめることができます。

連続的な確率変数 X が $a \leqq X \leqq b$ の値をとる確率 $P(a \leqq X \leqq b)$ が，

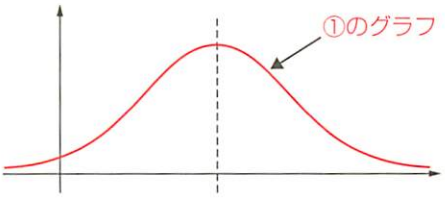
$$P(a \leqq X \leqq b) = \int_a^b f(x)\,dx$$

で表されるとき，関数 $f(x)$ を確率変数 X の確率密度関数という。

有名な確率密度関数に，

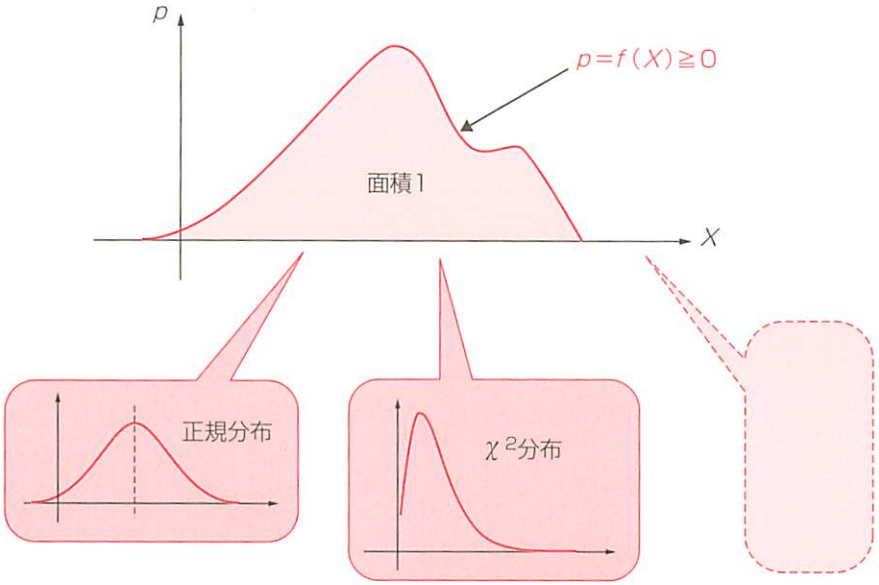
$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}\,e^{-\frac{(x-m)^2}{2\sigma^2}} \quad \cdots\cdots\cdots\cdots ①$$

があります。この式が表す確率分布は**正規分布**（**ガウス分布**）と呼ばれ，統計学では非常に重要なものです。

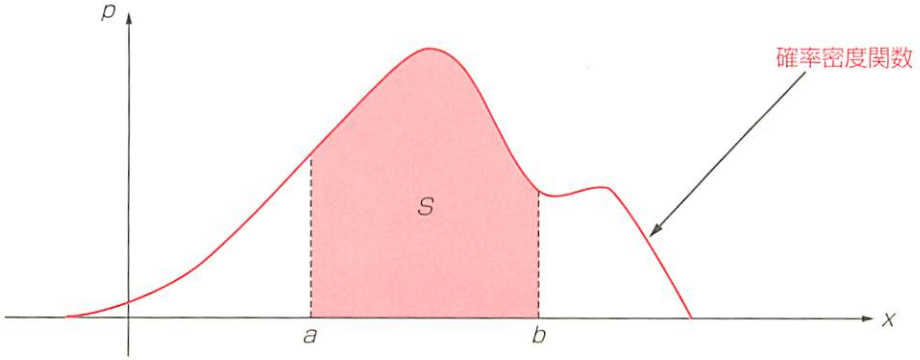


確率分布のグラフは確率密度関数のグラフ

1 確率密度関数は非負，面積 1



2 確率変数 X が a 以上 b 以下をとる確率は下図の面積 S



注）離散的な確率変数 X がとる確率を表した関数を確率関数という。

§63 確率分布グラフで一目瞭然

ここが要点！

たとえば，サイコロを10回投げたとき1の目がX回出る確率P_Xは，独立試行の定理から$P_X={}_{10}C_X\left(\frac{1}{6}\right)^X\left(\frac{5}{6}\right)^{10-X}$となります。しかし，この式を見ただけでは，Xの値の変化にともなうP_Xの変化のようすが定かではありません。こんなとき，各Xの値に対してP_Xの値を表にしておけばわかりやすくなります。さらに，XとP_Xの関係をグラフにすれば一目瞭然になります。

やさしい解説

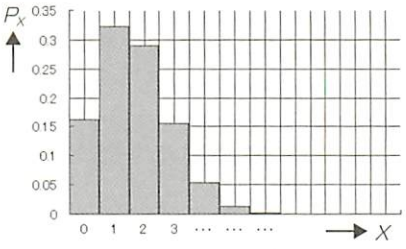
サイコロを10回投げたとき，1の目がx回出る確率P_xを，

$$P_X={}_{10}C_X\left(\frac{1}{6}\right)^X\left(\frac{5}{6}\right)^{10-X}$$

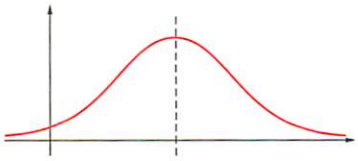
を利用して求めた表を下に示します。

X	P_X
0	0.161506
1	0.323011
2	0.29071
3	0.155045
4	0.054266
5	0.013024
6	0.002171
7	0.000248
8	1.86E-05
9	8.27E-07
10	1.65E-08

これは，「総量1の確率が確率変数Xの各値にどのように分配されているかを示す」もので，**確率分布表**と呼ばれています。この表を見ると，XとP_Xの関係がよくわかります。この表をもとに，XとP_Xの関係をグラフにしたのが下図です。



すると，XとP_Xの関係はもはや一目瞭然です。このグラフを**確率分布グラフ**といいます。確率変数Xが身長のように連続変量であれば，確率分布グラフは下図のように曲線になります。

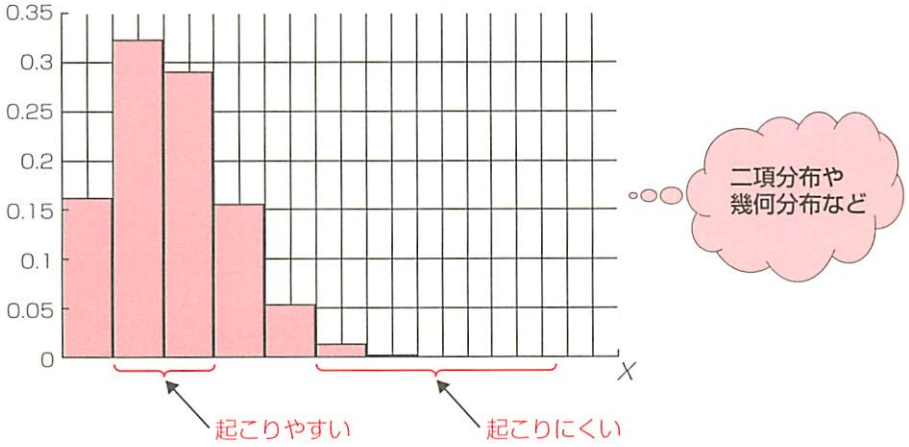


統計学では，さまざまな確率分布のグラフが利用されます。

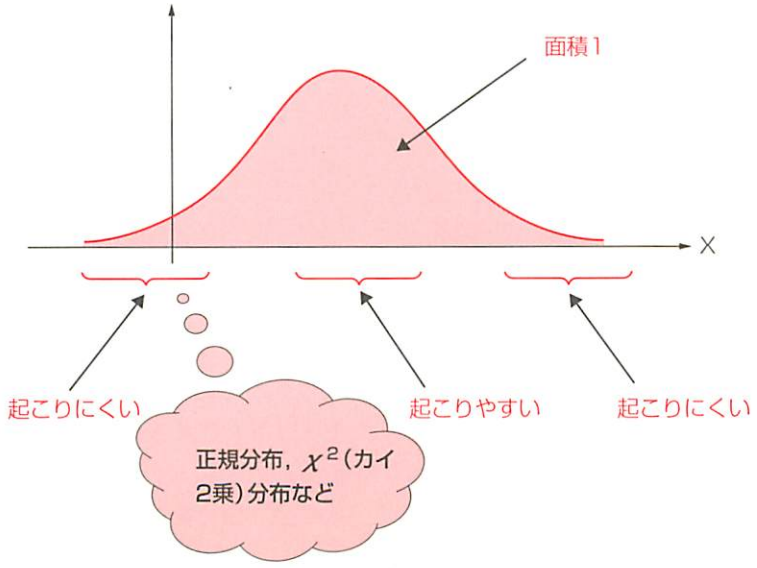
確率分布のグラフで確率は一目瞭然

面積はどのグラフでも1になる。

1 確率変数 X が離散型（とびとびの値をとる）の場合



2 確率変数 X が連続型（連続した値をとる）の場合

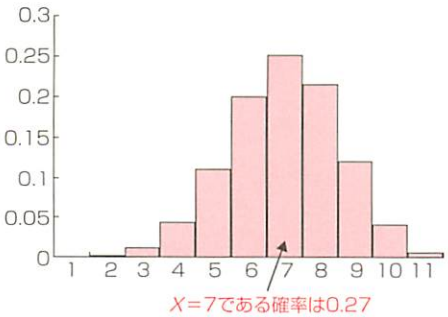


§64 ジャスト165cmである確率は0

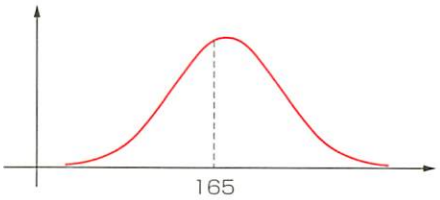
ここが要点！ 世界の人々の身長の分布は，連続型の分布といっていいでしょう。すると，奇妙なことが起こります。つまり，私の身長は165cmですが，その確率は0なのです。連続型の確率分布ではどんな分布でも，確率変数が1つの値をとる確率は0なのです。

やさしい解説

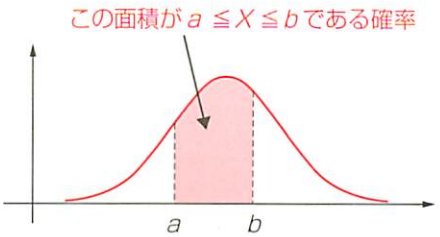
サイコロを振ったときの目の数をXとする確率変数は，とびとびの値しかとりません。コインを投げたとき表なら1，裏なら2とする確率変数Xもそうです。このように，とびとびの値しかとらない確率変数を**離散型の確率変数**といいます。離散型の確率変数の場合，$X=k$となる確率が0でないkの値が必ず存在します。



ところが，身長の値をXとするような確率変数のように，連続型の確率変数の場合は事情が異なります。kとしてどんな値を選んでも，$X=k$となる確率は0なのです。実際に身長が165cmの人がいたとしても，その確率は0なのです。ちょうど，物差しに点をデタラメに打つとき，目盛りの上に点が落ちる確率が0となるようにです。



連続型の確率変数Xの場合，Xが「a以上b以下」である確率は，下図のうす色部分の面積として定義されています。



したがってXがa，つまりXが「a以上a以下」である確率は，横幅が0の長方形の面積なので0となります。よって，$X=a$である確率は0になります。