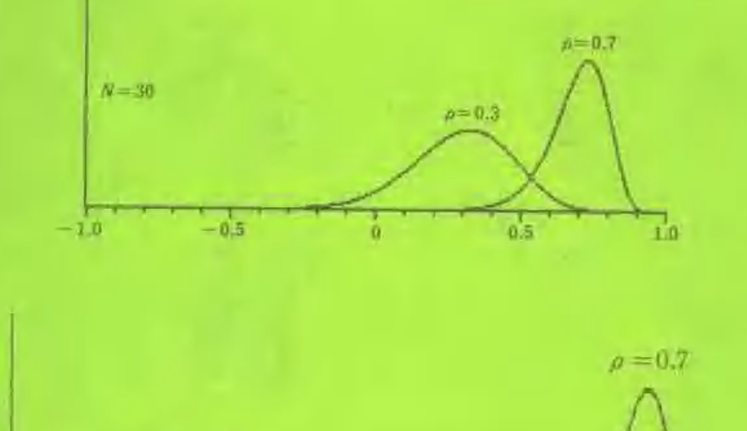
パソコンによる統計解析

杉山高一
牛沢賢二 著



朝倉書店

パソコンによる統計解析

杉山高一
牛沢賢二　著

朝倉書店

■著者略歴

杉山高一

1940年　東京都に生まれる
1963年　東京理科大学理学部数学科卒業
現　在　中央大学(理工学部数学科)教授
　　　　理学博士

牛沢賢二

1952年　山形県に生まれる
1976年　東京理科大学理学部応用数学科卒業
現　在　産業能率大学(地域科学研究所)研究員

パソコンによる **統計解析**　　　　定価 2900 円

1984年2月20日　初版第1刷
1984年8月5日　　　第2刷

著　者　杉　山　高　一
　　　　牛　沢　賢　二

発行者　朝　倉　邦　造

発行所　株式会社　朝　倉　書　店
東京都新宿区新小川町6-29
郵　便　番　号　162
電話　03（260）0141（代）
振替口座　東京6-8673番

〈検印省略〉

新日本印刷・渡辺製本

まえがき

統計的な考え方の基本を把握するのにパソコンをどう活用するか，またどう生かすかは興味深いテーマである．この本では統計の理論の理解を助ける目的で，パソコンによる統計シミュレーション実験や実際問題に応用できるプログラムを準備した．

この本の前半は，統計解析の基本的な考え方がわかるようにと意図して書いた．本文の中に出てくる事柄で，何故そうなるのかを知りたい，あるいはもう少し詳しく知りたいという読者のために，付録をのせておいた．

後半は応用編として，Basic 言語によるプログラムとそのフローチャートを 32 本のせてある．この他にも入れておきたいプログラムはあったが，頁数の制約もあってここにのせたものにしぼった．

本書の初稿の段階で，中央大学助教授青木一芳氏からいくつかの有益なコメントをいただいた．時系列をテーマとした第 10 章について，東京工業大学教授藤井光昭氏から有益なコメントをいただいた．両氏に心からの御礼を申し上げる．原稿に注意深く目を通し，たいへん行きとどいた編集をされた朝倉書店の編集部の方々に心からの謝意を表するしだいである．

1984 年初春

著　者

目　　次

応　用　編

1. 確率と乱数

§1. 1ケタの乱数
0, 1, 2,…, 9
の作り方など

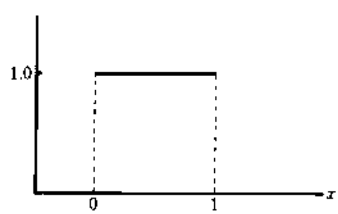
§2. ☆起こることが同様に確からしい標本点の集まり
☆標本点の集まりである事象
A, B, …
☆事象の確率の定義

§3. 確率の性質
☆和事象の確率 $P(A\cup B)$
☆積事象の確率 $P(A\cap B)$等
☆事象の独立とは

§4. ☆事象Aが起こる，起こらないのいずれか一方が生起する試行の繰り返し(二項分布)
☆二項分布のシミュレーション実験

§5. ☆一様分布
例 $f(x)=\begin{cases}1 & 0<x<1\\ 0 & \text{その他}\end{cases}$

☆一様乱数の生成法



§6. シミュレーション実験の応用
☆銀座地区の避難シミュレーション実験

1.1 乱　数　列

数字 0,1,2,3,4,5,6,7,8,9 をランダムに発生させる方法をのべる．ランダムとは，10個のどの数字も同じ割合で出現し，その現われ方に規則性のないことを意味する．この等出現性（等確率性）と無規則性（独立性）の2つの性質を有する数の列を乱数列という．

乱数を生成する簡単な方法をのべる．関数式

$$x_{n+1}=1899\,x_n$$

において，初期値 x_0 を適当に決める．いま，$x_0=0.234567$ とおくと，$1899x_0=445.442733$ になる．この小数第1位の数字4を乱数としてとる．小数部分を $x_1=0.442733$ とおくと，$1899\,x_1=840.749967$ になる．この小数第1位の数字7を乱数とする．同様のことを1000回繰り返して得られた乱数を次に記す．

```
4717106495 2676511927 2104554518 5068137822 7762416456
8342942715 3138166648 2865440030 2621241129 0603207117
6677928531 5792251086 3588057963 7535431349 2150564630
7247087721 2535715474 1852556286 4530030420 8623243557
3910670302 3611982204 8580179190 3485983942 8326298994

0343907606 8687881307 3992906657 9766324824 0954968728
9211693322 5542856151 4596270820 0654727834 6059668192
3821205768 3677814985 6863046283 9852096833 3455878400
3894794494 2767113371 3949891344 9404872931 9167419338
1993186849 2072119219 6943935974 8885539567 8874731280

1116128325 6147378116 5807514581 4369600830 9523852393
8757097174 8162709535 0420634224 4386692104 0284416174
2180038127 6718398311 7506521743 4035607034 2448319046
5015520137 4039421197 6624770059 2820650535 9660374505
9203301227 8578094106 6013989759 5845647299 8862614860

3787538027 2092135924 2948568712 5203072883 0511319336
7460886894 1293681279 8714326513 8830007052 2555984674
1672130847 3797524046 3820622020 6390549970 6466361083
8507764466 9006511140 3126709959 7974912528 5363557486
8101947322 7733711007 6777435816 2773396577 4989909057
```

各数字の1000回中の頻度は次のようである．

数　字	0	1	2	3	4	5	6	7	8	9	計
出現度数	102	100	103	103	97	94	101	105	101	94	1000
相対頻度	.102	.100	.103	.103	.097	.094	.101	.105	.101	.094	1

これを見ると相対頻度は0.10前後であり，この1000回に関しては等確率性は満たされているようである．どのような乱数の系列を得るかは，初期値に依存する．初期値をかえると別の乱数列を得ることは明らかであろう．

ここでは $x_{n+1}=1899x_n$ を用いたが，1899 のかわりに 147 を用いてもよいし，163, 173, 187, 197 等の素数を用いてもよい．このような生成法の欠点は，同じ数字の系列が再び現われるという周期性をもつことである．独立性も満たされないことが多いが，くわしくは1.5節でのべることにする．

1.2　確　率

前節でのべた乱数発生の方法は，$0, 1, 2, \cdots, 9$ のそれぞれの数字が等しい確率で現われるように作られている．この節では，この確率について説明する．

同じ条件のもとで繰り返し行うことができ，その結果が偶然に左右される実験や観測などを試行とよぶことにする．例として，数字 $0, 1, 2, 3, 4, 5, 6, 7, 8, 9$ のいずれかが起こる試行を考える．この 10 通りの結果を $\omega_0, \omega_1, \omega_2, \cdots, \omega_9$ で表わし，標本点と名づけよう．標本点全体の集まり

$$\Omega=\{\omega_0, \omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6, \omega_7, \omega_8, \omega_9\}$$

を標本空間という．つまり，標本空間はある試行において起こりうる場合全体の集まりである．試行の結果のある集まりを事象という．たとえば，奇数が起こるという事象Aは

$$A=\{\omega|\omega \text{ は奇数}\}=\{\omega_1, \omega_3, \omega_5, \omega_7, \omega_9\}$$

と表わされる．事象Aは標本空間 Ω の部分集合になっている．

試行において，**どの標本点が起こることも同様に確からしい**とする．標本空間の標本点の総数を M で，ある事象 A に含まれる標本点の個数を m で表わす．このとき，事象Aの起こる確率を

$$\mathbf{P}(A)=\frac{m}{M}$$

で定義する．いま，数字 $0, 1, 2, \cdots, 9$ のいずれか1つが，同様の確からしさで起こるとする．このとき奇数が起こるという事象は5通りあり，起こりうるす

べての場合の数は 10 通りあることから

$$\mathrm{P}(A)=\frac{5}{10}=\frac{1}{2}$$

となる.

ある試行に関する2つの事象AとBを考える. 標本空間を長方形で, その部分集合である事象AとBをそれぞれ円で表わすと, **和事象**, **積事象**, **余事象**, **排反事象**は図1.1のように表わされる.

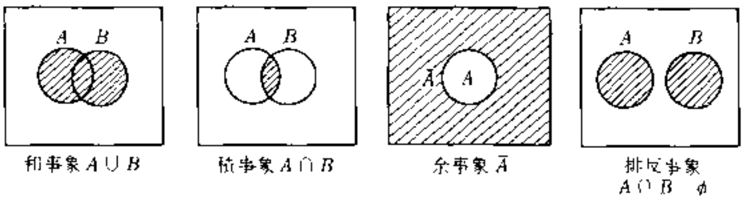


図 1.1　和事象, 積事象, 余事象, 排反事象

$A \cup B$はAかBの少なくとも一方が起こること, $A \cap B$はAとBの両方が共に起こること, $\bar{A}$はAが起こらないことを意味している. **$A \cap B=\phi$ は, AとBとは同時には起こらないこと**, つまりAとBは共有の標本点をもたないことを意味している. このとき, AとBとは互いに排反であるという.

確率の数学的な定義をのべよう. 集合体$\boldsymbol{B}$は次の条件を満たすとする.

(1)　標本空間 Ω は$\boldsymbol{B}$に属する.

(2)　ある集合Aが$\boldsymbol{B}$に属するならば, Aの補集合$\bar{A}$も$\boldsymbol{B}$に属する.

(3)　ある集合の列 $A_1, A_2, A_3, \cdots$ のそれぞれが$\boldsymbol{B}$に属するならば, $A_1 \cup A_2 \cup A_3 \cup \cdots$ も$\boldsymbol{B}$に属する.

この$\boldsymbol{B}$をボレル集合体という. $\boldsymbol{B}$に属する任意の事象Aに対して, 次の条件を満たす実数 $\mathrm{P}(A)$ を確率という.

(1)　$0 \leqq \mathrm{P}(A) \leqq 1$

(2)　$\mathrm{P}(\Omega)=1$

(3)　どの2つをとっても互いに排反である事象 $A_1, A_2, \cdots$ に対して

$$\mathrm{P}(A_1 \cup A_2 \cup \cdots)=\mathrm{P}(A_1)+\mathrm{P}(A_2)+\cdots$$

ボレル集合体$\boldsymbol{B}$は, Ω と ϕ を要素にもち, 補集合演算, 可算合併演算に関して

閉じている．標本空間Ωに，$\boldsymbol{B}$と集合関数としての確率測度Pの3つを一組にした（$\Omega, \boldsymbol{B}$, P）を定めて，確率を決めることになる．

前節で説明した乱数は，数字 0,1,2,…,9 が等確率で出現することが条件である．実際に生成して，ヒストグラムをかいたのが図1.2である．試行回数が2000回，10000回と多くなると，それぞれの数字の出現確率が安定してくるのが観察される．

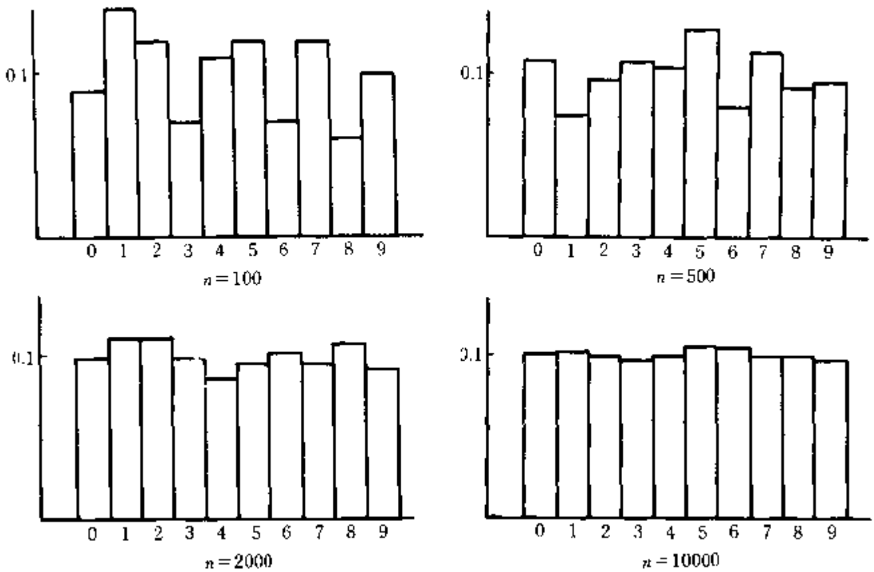


図 1.2　乱数の相対頻度

自然現象や社会現象で，データだけからある現象の起こる確率P(A)を求めなければならないことがある．このような場合は，実験または観測などを多数回繰り返し行い，N回の試行で事象Aの生起した回数N_Aより，P(A)$\fallingdotseq N_A/N$とする．つまり，試行回数Nを十分大きくし，相対度数 N_A/N がほぼ一定の値pをとるとき，P(A)$=p$ とするのである．これを**統計的確率**とよび，前の定義による確率を**数学的確率**と区別してよぶことがある．数学的確率pがわかっている場合，試行回数Nを大きくしさえすればいくらでも，相対度数 N_A/N は数学的確率pに近づくことが示される（付録 1.1，p.122 参照）．

1.3 確率の性質

標本空間 Ω のなかの任意の事象 A と B について，**加法法則**

$$\mathbf{P}(\boldsymbol{A}\cup\boldsymbol{B})=\mathbf{P}(\boldsymbol{A})+\mathbf{P}(\boldsymbol{B})-\mathbf{P}(\boldsymbol{A}\cap\boldsymbol{B}) \tag{1.1}$$

が成り立つ．A と B が排反事象であれば，$\mathrm{P}(A\cap B)=\mathrm{P}(\phi)=0$ であるから

$$\mathrm{P}(A\cup B)=\mathrm{P}(A)+\mathrm{P}(B)$$

となる．A の余事象を $\bar{A}$ で表わすと，A と $\bar{A}$ とは排反事象であるから

$$\mathrm{P}(A)+\mathrm{P}(\bar{A})=\mathrm{P}(A\cup\bar{A})=\mathrm{P}(\Omega)=1$$

が成り立ち，これから**余事象の法則**

$$\mathbf{P}(\boldsymbol{A})=\mathbf{1}-\mathbf{P}(\bar{\boldsymbol{A}}) \tag{1.2}$$

を得る．

以下の例で，数字 0, 1, 2, …, 9 は等しい確率で起こるとする．

(i)	標本空間 Ω	$\{0\ 1\ 2\ 3\ 4\ 5\ 6\ 7\ 8\ 9\}$
(ii)	6未満の起こる事象 A	$\{0\ 1\ 2\ 3\ 4\ 5\}$
(iii)	奇数の起こる事象 B	$\{1\ 3\ 5\ 7\ 9\}$

とすれば，和事象，積事象は次のようになる．

(iv)	和事象 $A\cup B$	$\{0\ 1\ 2\ 3\ 4\ 5\ 7\ 9\}$
(v)	積事象 $A\cap B$	$\{1\ 3\ 5\}$

これより，和事象の確率と積事象の確率は，それぞれ

$$\mathrm{P}(A\cup B)=\frac{8}{10},\quad \mathrm{P}(A\cap B)=\frac{3}{10}$$

である．式 (1.1) の右辺は，(ii), (iii), (v) から

$$\mathrm{P}(A)+\mathrm{P}(B)-\mathrm{P}(A\cap B)=\frac{6}{10}+\frac{5}{10}-\frac{3}{10}=\frac{8}{10}=\frac{4}{5}$$

であり，加法法則 (1.1) になっている．6未満の起こる事象 A の余事象 $\bar{A}$ は

(vi) 6以上の起こる事象 $\bar{A}$ $\{6\ 7\ 8\ 9\}$

であるから，A の確率と，1から $\bar{A}$ の確率を引いたものはそれぞれ

$$\mathrm{P}(A)=\frac{6}{10}=\frac{3}{5}$$

$$1-\mathrm{P}(\bar{A})=1-\frac{4}{10}=\frac{3}{5}$$

であり，余事象の法則(1.2)になっている．

事象Aが起こったという条件のもとで，事象Bの起こる確率を$\mathrm{P}(B|A)$で表わす．$\mathrm{P}(A)>0$のとき，**条件付確率**を

$$\mathbf{P}(\boldsymbol{B}|\boldsymbol{A})=\frac{\mathbf{P}(\boldsymbol{A}\cap\boldsymbol{B})}{\mathbf{P}(\boldsymbol{A})} \tag{1.3}$$

で定義する．これより，確率の**乗法法則**

$$\mathbf{P}(\boldsymbol{A}\cap\boldsymbol{B})=\mathbf{P}(\boldsymbol{B}|\boldsymbol{A})\mathbf{P}(\boldsymbol{A}) \tag{1.4}$$

を得る．$\mathrm{P}(B)>0$であれば

$$\mathrm{P}(A\cap B)=\mathrm{P}(A|B)\mathrm{P}(B) \tag{1.5}$$

と書くこともできる．事象$A=\{0,1,2,3,4,5\}$が起こったという条件のもとで$B=\{1,3,5,7,9\}$が起こるという事象$(B|A)$は

$$\{1\qquad 3\qquad 5\}$$

である．このときは事象Aを標本空間Ωと考えて，Bの確率を求めればよいから

$$\mathrm{P}(B|A)=\frac{3}{6}=\frac{1}{2}$$

である．また，$A\cap B=\{1,3,5\}$から

$$\frac{\mathrm{P}(A\cap B)}{\mathrm{P}(A)}=\frac{3/10}{6/10}=\frac{3}{6}=\frac{1}{2}$$

であり，条件付確率(1.3)つまり乗法法則(1.4)になっている．

事象A_1とA_2について

$$\mathbf{P}(\boldsymbol{A}_1\cap\boldsymbol{A}_2)=\mathbf{P}(\boldsymbol{A}_1)\mathbf{P}(\boldsymbol{A}_2) \tag{1.6}$$

が成り立つとき，**事象A_1とA_2は互いに独立**であるという．式(1.6)が成り立てば，乗法法則(1.4)より

$$\mathrm{P}(A_1|A_2)=\mathrm{P}(A_1) \tag{1.7}$$

であり，また式(1.7)から式(1.6)が成り立つことを示すことができる．式(1.7)はA_2が起こったということが，A_1の起こる確率に何の影響も与えていないことを意味している．

1.1節で数字$0,1,2,\cdots,9$をランダムに発生させる方法をのべた．**ランダムとは，これまでにどのような数字が出現したかということとは独立に**$0,1,2,$

…,9 **の数字が等確率で出現する**ことの意味である．乱数は，独立性と等確率性の2つの性質を満たしていることが必要である．独立性の条件式(1.7)を用いて，1.1 節で発生させた乱数を調べてみる．表 1.1 の i 行 j 列の数字は，$k-1$ 回目に数字 i が出現し，次の k 回目に数字 j が出現した回数である．どの組合せ (i,j) が表われる確率も 0.01 であることが要求される．10000 回であれば，任意の組合せ (i,j) の出現が期待される回数は 100 回ということになる．

表 1.1　10000 回のシミュレーションによる (i,j) の出現頻度

i＼j	0	1	2	3	4	5	6	7	8	9	計
0	94	94	98	115	92	113	91	106	93	89	985
1	90	117	117	83	94	98	99	88	126	94	1006
2	119	100	107	94	100	101	95	93	99	104	1012
3	101	82	106	90	117	94	107	93	105	111	1006
4	102	104	93	99	108	97	113	89	92	90	987
5	106	103	97	102	87	107	103	105	94	109	1013
6	91	121	106	90	105	107	111	99	88	102	1020
7	85	88	104	116	86	80	114	101	96	101	971
8	101	103	90	92	86	107	81	104	93	120	977
9	97	94	93	125	112	109	106	93	91	103	1023
計	986	1006	1011	1006	987	1013	1020	971	977	1023	10000

同様に3つの数字の組合せの例として数字(7,5)が出現したとき，続いて数字 0,1,2,…,9 が等確率で出現しているか否か等，いろいろな場合について調べなければならない．実際には，等確率性と独立性を調べることは大仕事である．これについては 1.5 節で再度のべることになる．

1.4　二項分布とシミュレーション実験

次の条件を満たすような試行を何回も繰り返し行う．

（1）　各回の試行で事象 A か $\bar{A}$ か必ず一方のみが起こる．

（2） 各回の試行は独立である．

（3） 各回の試行で $\mathrm{P}(A)=p$ は一定である．

この条件を満たす試行列を**ベルヌーイ試行列**という．いま，ある試行について成功か失敗かの場合を考え，成功の確率が p であるようなベルヌーイ試行列を考える．N 回の独立試行における成功の回数を表わす X は $0, 1, 2, \cdots, N$ のいずれかの値をとり，確率変数とよばれる．確率変数 X の各値に対して確率

$$\mathbf{P}(\boldsymbol{X}=\boldsymbol{k})={}_N\mathrm{C}_k\boldsymbol{p}^k\boldsymbol{q}^{N-k} \qquad (k=0, 1, 2, \cdots, N) \tag{1.8}$$

が対応している．これを X の確率分布という．式 (1.8) は**二項分布**である．

ここで

$$p+q=1$$

$${}_N\mathrm{C}_k=\frac{N!}{k!(N-k)!} \qquad (m!=m(m-1)(m-2)\cdots2\cdot1)$$

である．コンビネーション ${}_N\mathrm{C}_k$ は，互いに異なる N 個のなかから k 個取る組

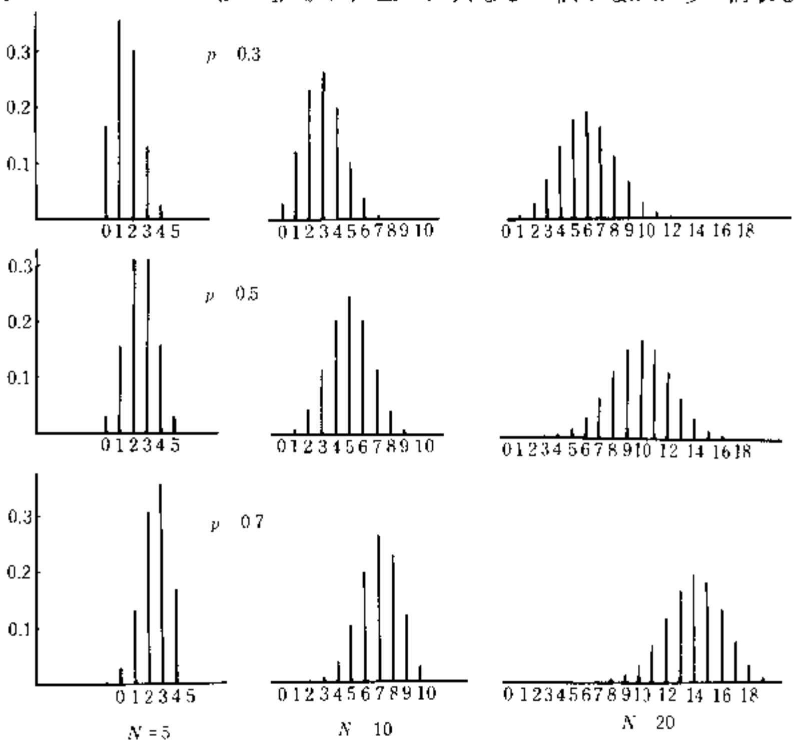


図 1.3 $p=0.3,\ 0.5,\ 0.7$ のときの二項分布