

数理統計学

猪股 清二・鈴木 皓之・関谷 昭・堀田 和久

開成出版

まえがき

情報処理の技術の発達とともに、情報を適確に統計的に処理することの必要性が一段と増している。

現在、統計学は生物学、農学、医学をはじめ、物理学、経済学、社会学、心理学、教育学等の学問研究には欠かすことのできない有用な道具として使われている。更に企業においては品質管理、生産管理、市場調査から人事、労務に至るまで科学的な経営管理等に有用な方法を与えている。

本書は、初めて数理統計学を学ぶ者および理工系の学生を対象として書かれた教科書である。このため取り扱う内容や配列に注意して、統計学の基本的な考え方を正しく理解できるよう努めた。また、理解を助けるための例題の挿入や解答にも注意をはらい、定理の証明における計算過程なども、できるかぎり省略しないこととした。

なお予備知識として微分積分学の初等的な知識を必要とするが、それもごく初等的なものであるから初めて学ぶ者にも十分理解してもらえるものと思う。

第1章では、資料を整理して、その特性を見い出す方法を取り扱う。

第2章では、確率変数の性質を取り扱う。

第3章では、二項分布、ポアソン分布、正規分布を取り扱う。

第4章では、 χ^2 -分布、 t -分布、 F -分布を取り扱う。

第5章では、標本による母集団の特性値の推定を取り扱う。

第6章では、標本による母集団の特性値についての仮説の検定を取り扱う。

最後に、本書の企画の段階から一貫して格別のご尽力をいただき、出版の機会を与えて下さった開成出版株式会社の方々に深く感謝いたします。

1982年2月

著者らしるす

目 次

ま え が き

第1章 序 説	1
---------------	---

§1.1 統計学概要	1
§1.2 1変量データの整理	2
§1.3 2変量データの整理	8

第2章 確率変数と分布関数	15
---------------------	----

§2.1 確率と確率変数	15
§2.2 確率変数の性質	31
§2.3 モーメント母関数	45

第3章 基本分布関数	49
------------------	----

§3.1 二項分布	49
§3.2 ポアソン分布	52
§3.3 正規分布	55
§3.4 母集団と標本	59
§3.5 中心極限定理	64

第4章 その他の分布関数	71
--------------------	----

§4.1 ガンマ関数とベータ関数	71
§4.2 χ^2 - 分布	73
§4.3 t - 分布	75
§4.4 F - 分布	78
§4.5 χ^2 -分布, t -分布, F -分布をする標本分布	80

第5章 母数の推定	91
-----------------	----

§5.1 点推定と区間推定	91
§5.2 母平均の推定	93
§5.3 比率の推定	98

§ 5.4 母平均差の推定 102

§ 5.5 比率差の推定 106

§ 5.6 母分散の推定 107

第6章 仮説の検定.....109

§ 6.1 仮説の検定 109

§ 6.2 母平均の検定 110

§ 6.3 比率の検定 114

§ 6.4 母分散の検定 116

§ 6.5 等平均の検定 117

§ 6.6 等分散の検定 120

§ 6.7 $p_A = p_B$ の検定 122

§ 6.8 適合度の検定 124

§ 6.9 $h \times h$ 分割表の独立性の検定 127

[付表 1] ポアソン分布表 (1) $P_k = \frac{m^k}{k!} e^{-m}$ 133

[付表 2] ポアソン分布表 (2) 134

[付表 3] 正規分布表 $\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$ 135

[付表 4] χ^2 -分布表 $\chi^2_{n, \alpha}$ 136

[付表 5] t -分布表 $t_{n, \alpha}$ 137

[付表 6] F -分布表 $F_{n, 0.05}^m$ 138

[付表 7] F -分布表 $F_{n, 0.025}^m$ 139

[付表 8] F -分布表 $F_{n, 0.01}^m$ 140

解 答 141

索 引 145



第1章 序 説

§1.1 統計学概要

日本人の成人男子の身長の平均値を求める場合を例にとろう。この場合、成人男子全員の身長計測値(データ)の集合を母集団といい、平均値などを母集団の特性値という。

母集団がこの例のように非常に大きな場合にはすべてのデータを使用して特性値を求めることは不可能である。また、それ程大きな母集団でなくても、データを得るために破壊をとまなうものや、多額の費用がかかるものなどでは、すべてのデータを得ることは不可能である。このような場合には母集団から無作為にいくつかのデータを抽出し(個々のデータを標本という)、これらのデータを用いて母集団の特性値を推定する。

統計学は元来、多量のデータの整理や要約の必要から生まれた学問である。したがってデータの整理の方法の開発、種々の特性値の導入、データや標本の性質の研究、特性値を正しく求めるための標本抽出の仕方や標本の個数の調査などを研究の主題として発展してきた。しかし1930年ごろからは、立場をかえて、得られた標本から母集団の特性値を推定する方法の研究に研究の主題が移り、学問体系の整備と応用範囲の拡大はともに目覚しい。1930年ごろを境として、それ以前に主として発展してきた統計学の体系を記述統計学、以後のそれを推測統計学として区別する場合がある。

本書は一貫して推測統計学の立場で書かれている。まず1章ではデータの整理の方法や種々の特性値の説明がなされる。2章では母集団のデータの分布を数学的に取り扱う方法が準備される。3章では統計学上最も基本的な分布をする母集団の性質が記述され、つづいて4章では3章で取り扱ったものよりやや複雑な分布が記述されている。この3章と4章で取り扱った分布を用いて、5章では標本より母集団の特性値を推定し、6章では母集団の特性値についての仮説の検定を取り扱う。

§1.2 1 変量データの整理

観測対象の調査・実験により得られる測定値をデータという。集められたデータをただ雑然と記録するだけでは、この対象のもっている性質を把握することはできない。この節ではデータを整理する方法とデータの性質を明らかにするための指標を考えてゆく。

1.2.1 度数分布表

ある工場で1日に生産される製品の不良率を200日間調べて表1-1を得た。観測データに関する特性を明らかにし、かつ種々の指標を計算する上で便利な形に整理するため、よく度数分布表が用いられる。これは変量(測定値)の変域、すなわち最大値と最小値を含む範囲をいくつかの区間に分け、その区間に入るデータの数(度数)を求めて表にしたものである。表1-2は表1-1のデータを度数分布表にまとめたものである。この表において、いくつか区切った区間を

表1-1

1.93	1.97	1.94	1.92	2.02	2.11	2.04	2.10	1.94	1.99
2.09	2.02	2.07	2.00	1.98	2.04	1.96	2.13	2.01	1.89
2.04	1.98	2.12	2.00	2.09	2.03	2.00	1.95	2.03	2.02
1.95	2.14	1.99	2.04	1.96	1.99	1.92	2.01	1.99	2.06
1.89	2.06	1.95	2.01	1.99	1.97	2.02	2.05	1.93	2.09
2.02	2.05	2.09	2.01	2.03	2.10	2.08	2.04	2.07	1.96
1.94	2.00	2.08	1.93	1.97	2.07	1.98	2.07	2.09	2.04
1.98	2.04	2.05	2.08	1.90	2.01	1.94	1.99	2.07	1.98
2.13	2.02	1.92	2.02	2.06	2.05	1.95	2.02	1.97	2.11
1.96	2.01	2.06	1.99	2.05	1.87	2.08	2.06	2.05	2.01
2.05	2.10	2.03	2.09	2.07	2.06	1.98	2.09	2.08	2.03
1.99	2.08	1.98	1.96	2.13	2.03	2.02	2.00	1.91	2.07
2.09	2.03	1.89	1.97	2.08	1.99	2.01	1.96	1.95	1.97
2.00	1.96	2.01	2.07	2.02	2.00	1.94	2.07	2.03	2.12
1.97	1.95	2.02	2.10	2.01	2.09	2.06	2.04	2.08	2.05
1.90	2.03	2.04	2.06	1.99	1.92	1.97	1.98	2.04	2.02
2.06	2.12	1.94	2.03	2.04	2.08	2.03	2.14	2.00	2.00
2.03	1.99	2.00	1.98	1.93	1.96	2.00	2.05	1.96	1.95
2.07	2.05	1.98	1.95	2.06	1.94	1.99	2.03	1.92	2.10
2.11	1.99	2.03	1.91	1.97	2.02	2.04	1.98	2.01	2.03

階級といい、その区間の中間値を級心、ある階級に入るデータの個数(度数)とデータの総数の比を**相対度数**という。度数分布表を直観的に見やすくした図1-1を**度数柱図表**または**ヒストグラム**という。一般に測定値と測定値の現れる度数との関係を**度数分布**といっている。

データの特性を正しく知るためには、度数分布表を

作成する段階で、合理的な階級に分ける事が必要である。階級の数が多すぎても少なすぎてもいけないのであって、一般に、表1-3の程度にする。

以上述べたことをまとめて、度数分布表を作る手順を示せば、

- (1) データの数 N を調べる ($N > 100$ であることが望ましい)。
- (2) 表1-3を参照して階級の数 k をきめる。
- (3) 階級の幅をきめる。

データの中から最大値 L と最小値 S を見だし、 L と S の差を k で割り、そ

表1-2

階 級	級心	度数	累積度数	相対度数
1.855~1.885	1.87	1	1	1/200
1.885~1.915	1.90	7	8	7/200
1.915~1.945	1.93	16	24	16/200
1.945~1.975	1.96	26	50	26/200
1.975~2.005	1.99	35	85	35/200
2.005~2.035	2.02	39	124	39/200
2.035~2.065	2.05	32	156	32/200
2.065~2.095	2.08	28	184	28/200
2.095~2.125	2.11	11	195	11/200
2.125~2.155	2.14	5	200	5/200

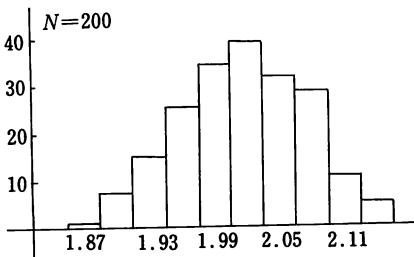


図1-1

表1-3

データの数	適当な階級の数
50~100	6~10
100~250	7~12
250 以上	10~20

の商の値に近く、かつデータの最小桁を単位として、その整数倍の都合のよい値を階級の幅とする(例えば表 1-1 の例で、 $L=2.14$ 、 $S=1.87$ であり、 $k=10$ と選ぶと $(L-S)/k=0.027$ となる。このデータは小数以下 2 桁を最小測定桁としているので 0.01 の整数倍で 0.027 に近い値 0.03 を階級の幅と定める)。

逆に階級の幅を適当に定めたとき、階級の数をも求め、これが表 1-3 と比較して適当であるならば、この幅を採用してよい。

(4) 階級の限界値をきめる。この場合 1 つのデータが 2 つの階級に入る事のないように注意する。限界値の数値は最小測定桁より 1 桁下げておくことよい。

(5) 度数分布表の作成。

(6) ヒストグラムの作成。

以上の手順を経るのであるが、データの中に異常にかけ離れた値をもつものがあれば、そのデータは取り除く方がよい。

(例) 表 1-4 は、ある学校の学生 100 人の数学の成績である。これより度数分布表、ヒストグラムを作れ。

表 1-4

76	66	74	76	68	82	57	75	35	34
65	63	85	87	79	77	76	74	76	78
75	60	96	74	73	87	52	98	88	64
76	69	60	74	72	76	57	64	67	58
72	80	72	56	73	82	78	45	75	56
75	86	66	86	50	78	66	79	68	60
80	83	87	78	80	77	81	92	57	52
58	82	73	95	66	60	84	80	79	63
80	88	58	84	96	87	72	65	79	80
86	68	76	41	80	40	63	90	83	94

(解) (1) $N=100$

(2) 階級の数 は 7~10 がよい。

(3) 最大値=98, 最小値=34.

$98-34=64$ であるから 5 点きざみにすると階級の数 は 13 となって多すぎる。10 点きざみにすると階級の数 は 7 となるから、10 点きざみにきめる。

(4) 階級の限界値は 29.5~39.5, 39.5~49.5……とする。

以上から表 1-5 と図 1-2 が得られる。

ヒストグラムの柱の上端の中心を順次結んでできる折線を度数多角形といい、データの数が多くなれば、階級の数も増加し、度数多角形は次第になめらかな曲線に近づく。この曲線を度数分布曲線という。

ヒストグラムの形によって、分布を対称型、L 字型、U 字型などと呼ぶ場合がある。

表 1-5

階 級	級心	度数
29.5~39.5	34.5	2
39.5~49.5	44.5	3
49.5~59.5	54.5	11
59.5~69.5	64.5	20
69.5~79.5	74.5	32
79.5~89.5	84.5	25
89.5~99.5	94.5	7

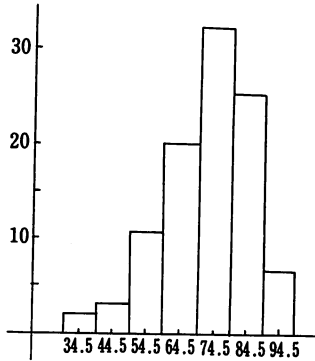


図 1-2

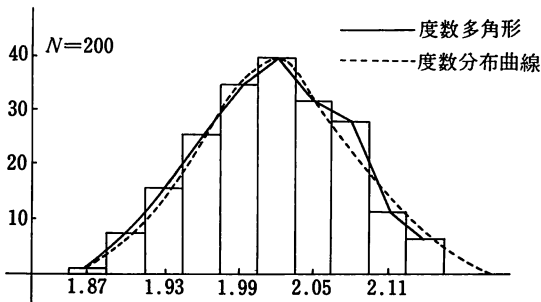


図 1-3

1.2.2 平均と分散

データの特性を表す指標としてよく用いられるものに平均、分散および標準偏差がある。

平均 μ とは n 個のデータ $x_1, x_2, \dots, x_n$ の算術平均をいい、

$$\mu = \frac{1}{n} \sum_{i=1}^n x_i$$

である。また

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2$$

で定義される σ^2 を分散といい、この正の平方根 σ を標準偏差という。つまり

$$\sigma = \sqrt{\sigma^2}$$

である。ここで

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \mu^2$$

なる恒等式が成り立つことは明らかであろう。計算をする場合には後者の式の方が便利ことが多い。

μ, σ^2 がそれぞれもっている性質から、平均はデータの代表値として用いられ、分散や標準偏差はデータのばらつきを示す尺度となっている。

表 1-6

級心	x_1	x_2	x_3	……	x_n	計
度数	f_1	f_2	f_3	……	f_n	N

度数分布表(表 1-6)から平均、分散を求

めるときには、級心を階級の代表値として
考え

$$\mu = \frac{1}{N} \sum_{i=1}^n f_i x_i$$

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^n f_i (x_i - \mu)^2 = \frac{1}{N} \sum_{i=1}^n f_i x_i^2 - \mu^2$$

とする。

なお、変数変換を施して、計算を簡便化する方法もよく用いられる。

1° 表 1-6 において

$$x_i = au_i + b \quad (i=1, 2, \dots, n; a, b: \text{定数})$$

とする時、 x_i, u_i の平均を μ_x, μ_u 、分散を σ_x^2, σ_u^2 とすれば

$$\mu_x = a\mu_u + b, \quad \sigma_x^2 = a^2\sigma_u^2$$

である。

$$\begin{aligned} \therefore \mu_x &= \frac{1}{N} \sum_{i=1}^n f_i x_i = \frac{1}{N} \sum_{i=1}^n f_i (au_i + b) \\ &= a \frac{1}{N} \sum_{i=1}^n f_i u_i + b \frac{1}{N} \sum_{i=1}^n f_i \\ &= a\mu_u + b \end{aligned}$$

$$\sigma_x^2 = \frac{1}{N} \sum_{i=1}^n f_i (x_i - \mu_x)^2 = \frac{1}{N} \sum_{i=1}^n f_i \{(au_i + b) - (a\mu_u + b)\}^2$$

$$= a^2 \frac{1}{N} \sum_{i=1}^n f_i (u_i - \mu_u)^2 = a^2 \sigma_u^2$$

この関係を用いて度数分布表から平均と分散を求めてみよう。

(例) つぎの度数分布表(表1-7)から平均と分散を求めよ。

(解) $x_i = 0.05u_i + 1.65$ とおき, x_i, u_i の平均, 分散をそれぞれ $\mu_x, \mu_u, \sigma_x^2, \sigma_u^2$ とおく。

x_i	u_i	f_i	$f_i u_i$	$f_i u_i^2$
1.45	-4	3	-12	48
1.50	-3	10	-30	90
1.55	-2	35	-70	140
1.60	-1	41	-41	41
1.65	0	64	0	0
1.70	1	52	52	52
1.75	2	12	24	48
1.80	3	3	9	27
計		220	-68	446

表 1-7	
身長(x_i)	度数
1.45	3
1.50	10
1.55	35
1.60	41
1.65	64
1.70	52
1.75	12
1.80	3

$$\mu_u = -68 \div 220 \doteq -0.31$$

$$\sigma_u^2 = 446 \div 220 - (0.31)^2 \doteq 1.93$$

ゆえに求める平均 μ_x , 分散 σ_x^2 は

$$\mu_x = 0.05 \times (-0.31) + 1.65 = 1.634$$

$$\sigma_x^2 = (0.05)^2 \times 1.93 = 0.0048$$

分散がデータのばらつきを表すことはすでに述べたが, この性質をより定量的にとらえてみよう。

2° チェビシェフの不等式

表1-6において, 平均 μ , 標準偏差を σ とおけば

$$\sum_{\{i: |x_i - \mu| \leq \lambda \sigma\}} f_i > N \left(1 - \frac{1}{\lambda^2} \right)$$

が成り立つ。

なぜならば $\sum_{\{i: |x_i - \mu| > \lambda \sigma\}} f_i = n$ とおくと

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^n f_i (x_i - \mu)^2 > \frac{\lambda^2}{N} \sigma^2 n$$

$$\therefore n < \frac{N}{\lambda^2}$$

$$\therefore \sum_{\{i: |x_i - \mu| \leq \lambda\sigma\}} f_i = N - n > N \left(1 - \frac{1}{\lambda^2}\right)$$

いま、この不等式において、 $\lambda=3$ とすれば、

$$\sum_{\{i: |x_i - \mu| \leq 3\sigma\}} f_i > \frac{8}{9}n$$

となり、 $\mu-3\sigma$ から $\mu+3\sigma$ の間にある x_i の個数は全体の約 89% 以上であることを示している。この性質は σ の大小には無関係であるので、 σ が小さければ小さい程、 $\mu-3\sigma$ と $\mu+3\sigma$ の間隔は狭くなり、やはりその中に全体の 89% 以上がはいってしまう。この性質を **3 σ の定理**と呼んでいる。

問1 100 人の生徒の身長について、つぎの分布表を得た。これより、平均と標準偏差を求めよ。

身長(cm)	140	143	146	149	152	155	158	161	164	167	170
度数	1	3	6	14	25	26	11	6	5	2	1

問2 5 個の貨幣を 100 回投げて、各回に出た表の数の度数分布表を作れ。これより平均と分散を求めよ。

問3 数学のテストの結果、平均点 53 点、標準偏差 8 点であった。1 クラス 45 人の学級で成績が 41 点から 65 点までの者が何人以上いるか。

§1.3 2変量データの整理

同一の観測対象から、2 つの異なるデータを得た場合、2 変量間の関係を定量的かつ直観的に把握するための指標やデータの整理方法は、1 変量だけの場合と異なってくる。

1.3.1 相 関 表

x, y の 2 変量を測定して表 1-8 を得た。

このままの形では 2 変量の間にはどのような関係があるのか判断しにくい。そこで表 1-9 の相関表を作成する。これは x, y の値を階級に分け、両方の階級を満足するデータの個数を各マス目に記入するのである。

表 1-8

x	71	73	71	71	73	75	77	71	71	75
y	64	70	61	64	67	76	76	70	67	76

表 1-9

	70~72 71	73~75 74	76~78 77	79~81 80
60~63 61.5	1			
64~67 65.5	3	1		
68~71 69.5	1	1		
72~75 73.5				
76~79 77.5		2	1	

表 1-10

	x_1	x_2	x_3	...	x_i	...	x_n	計
y_1	f_{11}	f_{21}	f_{31}	...	f_{i1}	...	f_{n1}	$f_{\cdot 1}$
y_2	f_{12}	f_{22}	f_{32}	...	f_{i2}	...	f_{n2}	$f_{\cdot 2}$
...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...
y_j	f_{1j}	f_{2j}	f_{3j}	...	f_{ij}	...	f_{nj}	$f_{\cdot j}$
...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...
y_m	f_{1m}	f_{2m}	f_{3m}	...	f_{im}	...	f_{nm}	$f_{\cdot m}$
計	$f_{1\cdot}$	$f_{2\cdot}$	$f_{3\cdot}$	...	$f_{i\cdot}$	...	$f_{n\cdot}$	N

2変量 x と y の間の相関表を一般的に書くと表 1-10 のようになる。

この表で f_{ij} は x については x_i , y については y_j を級心とする階級に含まれるデータの度数である。また一番下の行の $f_{i\cdot}$ は y を考えない場合の x_i の度数を示し、一番右の列の $f_{\cdot j}$ は x を考えない場合の y_j の度数を示している。こ

れらを周辺分布と呼んでいる。 N はデータの総数であって、

$$N = \sum_{i=1}^n f_{i.} = \sum_{j=1}^m f_{.j} = \sum_{i=1}^n \sum_{j=1}^m f_{ij}$$

である。

	71	74	77	80
61.5	•			
65.5	• • •	•		
69.5	•	•		
73.5				
77.5		•	•	

図 1-4

相関表の各欄に度数を記入する代わりに、度数と同数の点を記入した図表を相関図と呼ぶ。この図表によりデータの分布を直観的に把握することができる。図 1-4 は表 1-9 から得られた相関図である。

1.3.2 共分散と相関係数

2 変量 x, y に関する測定データが

$$(x_1, y_1), (x_2, y_2), (x_3, y_3), \dots, (x_n, y_n)$$

で与えられているとする。 x に関する平均と分散をそれぞれ μ_x, σ_x^2 とし、 y に関する平均と分散をそれぞれ μ_y, σ_y^2 とする。すなわち

$$\mu_x = \frac{1}{n} \sum_{i=1}^n x_i \quad \sigma_x^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu_x)^2$$

$$\mu_y = \frac{1}{n} \sum_{i=1}^n y_i \quad \sigma_y^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \mu_y)^2$$

とするとき、共分散 σ_{xy} と相関係数 ρ をつぎのように定義する。

$$\sigma_{xy} = \frac{1}{n} \sum_{i=1}^n (x_i - \mu_x)(y_i - \mu_y) = \frac{1}{n} \sum_{i=1}^n x_i y_i - \mu_x \mu_y$$

$$\rho = \frac{\sigma_{xy}}{\sigma_x \cdot \sigma_y}$$

相関係数 ρ の性質を調べてみよう。

I. $-1 \leq \rho \leq 1$

(証明) シュワルツの不等式より

$$\left(\sum_{i=1}^n (x_i - \mu_x)(y_i - \mu_y) \right)^2 \leq \left(\sum_{i=1}^n (x_i - \mu_x)^2 \right) \cdot \left(\sum_{i=1}^n (y_i - \mu_y)^2 \right)$$

ゆえに

$$\sigma_{xy}^2 \leq \sigma_x^2 \cdot \sigma_y^2$$

が成り立つ。ゆえに

$$\rho^2 = \frac{\sigma_{xy}^2}{\sigma_x^2 \cdot \sigma_y^2} \leq 1$$

$$\therefore -1 \leq \rho \leq 1$$

II. $y_i = ax_i + b$ のとき

$$\rho = 1 \quad (a > 0) \quad \text{または} \quad \rho = -1 \quad (a < 0)$$

(証明) 省略

III. 相関図と ρ の値の関係は図 1-5 に示す。これらの図より、2 変量間に 1 次関係が強ければ強いほど相関係数は ± 1 に近づくことが知れよう。

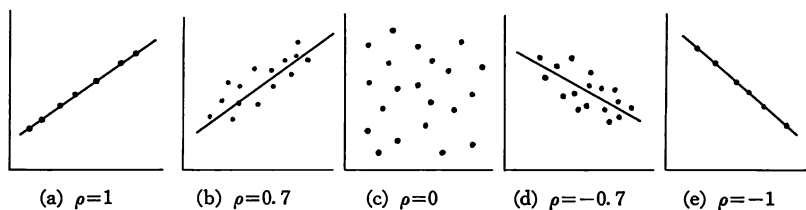


図 1-5

2 次元においても計算の簡便化を考えよう。まず表 1-10 より x_i, y_j の平均、分散、共分散、相関係数はそれぞれ

$$\mu_x = \frac{1}{N} \sum_{i=1}^n f_i \cdot x_i, \quad \mu_y = \frac{1}{N} \sum_{j=1}^m f_j \cdot y_j$$

$$\sigma_x^2 = \frac{1}{N} \sum_{i=1}^n f_i \cdot (x_i - \mu_x)^2 = \frac{1}{N} \sum_{i=1}^n f_i \cdot x_i^2 - \mu_x^2$$

$$\sigma_y^2 = \frac{1}{N} \sum_{j=1}^m f_j \cdot (y_j - \mu_y)^2 = \frac{1}{N} \sum_{j=1}^m f_j \cdot y_j^2 - \mu_y^2$$

$$\sigma_{xy} = \frac{1}{N} \sum_{i=1}^n \sum_{j=1}^m f_{ij} (x_i - \mu_x)(y_j - \mu_y) = \frac{1}{N} \sum_{i=1}^n \sum_{j=1}^m f_{ij} x_i y_j - \mu_x \mu_y$$

$$\rho = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

である。ここで

$$x_i = au_i + b, \quad y_j = cv_j + d \quad (i=1, 2, \dots, n; j=1, 2, \dots, m; ac \neq 0)$$

とおき、 u_i, v_j の共分散、相関係数を σ_{uv}, R とすれば、つぎの等式が成り立つ。

$$\sigma_{xy} = ac\sigma_{uv}$$

$$\rho = \pm R \quad (ac > 0 \text{ のとき} +, ac < 0 \text{ のとき} -)$$

これはつぎの式から容易に導かれる.

$$\mu_x = a\mu_u + b \quad \sigma_x^2 = a^2\sigma_u^2 \quad \sigma_x = |a|\sigma_u$$

$$\mu_y = c\mu_v + d \quad \sigma_y^2 = c^2\sigma_v^2 \quad \sigma_y = |c|\sigma_v$$

$$(x_i - \mu_x)(y_j - \mu_y) = ac(u_i - \mu_u)(v_j - \mu_v)$$

相関表(表 1-10)から, 各変量 x, y の平均値と分散および 2 変量間の共分散を計算するときには, 系統的に計算する意味で, 表 1-11 のような形で下記の諸量をあらかじめ計算しておくといふ.

$$f_i u_i, f_j v_j, f_i u_i^2, f_j v_j^2, F_i = \sum_{j=1}^m f_{ij} v_j, G_j = \sum_{i=1}^n f_{ij} u_i, F_i u_i, G_j v_j$$

このとき

$$\sigma_{uv} = \frac{1}{N} \sum_{i=1}^n F_i u_i - \mu_u \mu_v = \frac{1}{N} \sum_{j=1}^m G_j v_j - \mu_u \mu_v$$

表 1-11

		x_1	x_2	$\cdots$	x_i	$\cdots$	x_n					
		u_1	u_2	$\cdots$	u_i	$\cdots$	u_n	$f_{.j}$	$f_{.j}v_j$	$f_{.j}v_j^2$	G_j	G_jv_j
y_1	v_1	f_{11}	f_{21}	$\cdots$	f_{i1}	$\cdots$	f_{n1}	$f_{.1}$	$f_{.1}v_1$	$f_{.1}v_1^2$	G_1	G_1v_1
y_2	v_2	f_{12}	f_{22}	$\cdots$	f_{i2}	$\cdots$	f_{n2}	$f_{.2}$	$f_{.2}v_2$	$f_{.2}v_2^2$	G_2	G_2v_2
$\vdots$	$\vdots$			$\cdots$		$\cdots$	$\vdots$					
y_j	v_j	f_{1j}	f_{2j}	$\cdots$	f_{ij}	$\cdots$	f_{nj}	$f_{.j}$	$f_{.j}v_j$	$f_{.j}v_j^2$	G_j	G_jv_j
$\vdots$					$\vdots$							
y_m	v_m	f_{1m}	f_{2m}	$\cdots$	f_{im}	$\cdots$	f_{nm}	$f_{.m}$	$f_{.m}v_m$	$f_{.m}v_m^2$	G_m	G_mv_m
	$f_{i.}$	$f_{1.}$	$f_{2.}$	$\cdots$	$f_{i.}$	$\cdots$	$f_{n.}$	N	$\sum_{j=1}^m f_{.j}v_j$	$\sum_{j=1}^m f_{.j}v_j^2$	$\sum_{j=1}^m G_j$	$\sum_{j=1}^m G_jv_j$
	f_iu_i	f_1u_1	f_2u_2	$\cdots$	f_iu_i	$\cdots$	f_nu_n	$\sum_{i=1}^n f_iu_i$				
	$f_iu_i^2$	$f_2u_1^2$	$f_2u_2^2$	$\cdots$	$f_iu_i^2$	$\cdots$	$f_nu_n^2$	$\sum_{i=1}^n f_iu_i^2$				
	F_i	F_1	F_2	$\cdots$	F_i	$\cdots$	F_n	$\sum_{i=1}^n F_i$				
	F_iu_i	F_1u_1	F_2u_2	$\cdots$	F_iu_i	$\cdots$	F_nu_n	$\sum_{i=1}^n F_iu_i$				

問 表 1-9 の相関表から 2 変量 x, y の平均, 分散および相関係数を求めよ.

1.3.3 回 帰 直 線

2 変量 x, y が $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ なる値を取り, 相関係数 $\rho = \pm 1$ であるときには, $y_i = ax_i + b$ なる 1 次関係が成立していることはすでに述べた (図 1-5(a), (e)).

一般に $0 < |\rho| < 1$ の場合にも, たとえば図 1-5 の (b) や (d) のように, 全体としてのデータの傾向が 1 次関係をみたしているとしてさしつかえない場合には, 2 変量 x と y の間に $y = ax + b$ において, $x = x_i$ のときの y の値 Y_i とデータ y_i の値の差の二乗の和, $\sum_{i=1}^n (y_i - Y_i)^2$ が最小になるように係数 a, b を定め, 直線 $y = ax + b$ を y の x への回帰直線と呼んでいる. また a を y の x への回帰係数と呼ぶ. 係数 a, b を決定しよう. いま

$$F(a, b) = \sum_{i=1}^n (y_i - Y_i)^2 = \sum_{i=1}^n (y_i - ax_i - b)^2$$

とおき, $\sum_{i=1}^n (y_i - Y_i)^2$ を a, b の関数と考えると, $F(a, b)$ が最小になる点では

$$\frac{\partial F}{\partial a} = \frac{\partial F}{\partial b} = 0$$

であるから,

$$\frac{\partial F}{\partial a} = \sum_{i=1}^n \{-2x_i(y_i - ax_i - b)\} = 0 \quad (1.3.1)$$

$$\frac{\partial F}{\partial b} = \sum_{i=1}^n \{-2(y_i - ax_i - b)\} = 0 \quad (1.3.2)$$

(1.3.2) より

$$\begin{aligned} b &= \frac{1}{n} \sum_{i=1}^n (y_i - ax_i) \\ &= \mu_y - a\mu_x \end{aligned} \quad (1.3.3)$$

(1.3.3) を (1.3.1) に代入すると

$$\begin{aligned} 0 &= \sum_{i=1}^n x_i (y_i - ax_i - \mu_y + a\mu_x) \\ &= \sum_{i=1}^n x_i y_i - a \sum_{i=1}^n x_i^2 - \mu_y \sum_{i=1}^n x_i + a\mu_x \sum_{i=1}^n x_i \end{aligned}$$

$$\begin{aligned}
&= n \left(\frac{1}{n} \sum_{i=1}^n x_i y_i - \mu_x \mu_y \right) - a n \left(\frac{1}{n} \sum_{i=1}^n x_i^2 - \mu_x^2 \right) \\
\therefore a &= \frac{\left(\frac{1}{n} \sum_{i=1}^n x_i y_i - \mu_x \mu_y \right)}{\left(\frac{1}{n} \sum_{i=1}^n x_i^2 - \mu_x^2 \right)} = \frac{\sigma_{xy}}{\sigma_x^2} = \frac{\sigma_y}{\sigma_x} \rho \\
\therefore b &= \mu_y - \frac{\sigma_y}{\sigma_x} \rho \mu_x
\end{aligned}$$

したがって y の x への回帰直線は

$$y = \frac{\sigma_y}{\sigma_x} \rho (x - \mu_x) + \mu_y$$

となる。同様に x の y への回帰直線も求めることができ

$$x = \frac{\sigma_x}{\sigma_y} \rho (y - \mu_y) + \mu_x$$

となる。

いま $a = \frac{\sigma_y}{\sigma_x} \rho, b = \mu_y - \frac{\sigma_y}{\sigma_x} \rho \mu_x$ として $F(a, b)$ に代入してみると

$$\begin{aligned}
F(a, b) &= \sum_{i=1}^n (y_i - a x_i - b)^2 \\
&= \sum_{i=1}^n \left(y_i - \frac{\sigma_y}{\sigma_x} \rho x_i - \mu_y + \frac{\sigma_y}{\sigma_x} \rho \mu_x \right)^2 \\
&= \sum_{i=1}^n (y_i - \mu_y)^2 + \frac{\sigma_y^2}{\sigma_x^2} \rho^2 \sum_{i=1}^n (x_i - \mu_x)^2 - 2 \frac{\sigma_y}{\sigma_x} \rho \sum_{i=1}^n (x_i - \mu_x)(y_i - \mu_y) \\
&= n \sigma_y^2 + n \sigma_y^2 \rho^2 - 2 n \sigma_y^2 \rho^2 \\
&= n \sigma_y^2 (1 - \rho^2)
\end{aligned}$$

この式から、相関係数 ρ の絶対値が大きければ大きいほど $F(a, b)$, つまり直線からのずれが小さくなることがわかる。したがって相関係数は、1次関係の強さを示すとした1.3.2のⅢの説明も納得がいく。

第2章 確率変数と分布関数

§2.1 確率と確率変数

2.1.1 標本空間と事象

1つのデータを得るための試みを試行といい、試行の結果として起こりうるすべての場合を要素とする集合を標本空間という。いま標本空間を S とし、試行により起こりうるすべての結果を $E_1, E_2, \dots, E_n$ とすると

$$S = \{E_1, E_2, \dots, E_n\}$$

と書くことができる。

(例1) サイコロを振るという試行で

$$E_i = i \text{ の目が出ること } (i=1, 2, 3, 4, 5, 6)$$

と定義すると

$$S = \{E_1, E_2, E_3, \dots, E_6\}$$

である。

(例2) 3枚の貨幣を投げるとする試行において

$$E_i = \text{表が } i \text{ 枚出ること } (i=0, 1, 2, 3)$$

と定義すると

$$S = \{E_0, E_1, E_2, E_3\}$$

となる。

S の部分集合を事象という。上の例1で、“偶数の目が出る”という事象は $\{E_2, E_4, E_6\}$ であり、例2で“1枚も表が出ない”という事象は $\{E_0\}$ である。

S の1つの要素からなる事象を単純事象という。

1つの試行による標本空間は、ただ1通りに定まるものではない。たとえば、例1で、

$$Fodd = \text{奇数の目が出ること}; \quad Feven = \text{偶数の目が出ること}$$

と定義すれば、

$$S = \{Fodd, Feven\}$$

となる。

事象と事象の間には、和、積、否定という3つの演算が定義される。事象の間に、これらの演算が行われた結果は、また1つの事象であり、それぞれ、和事象、積事象、余事象という。

A, B を各々事象とするとき

和事象: A と B のどちらかが起こる事象。記号で $A \cup B$ と書く。

積事象: A と B の両方が同時に起こる事象。記号で $A \cap B$ と書く。

余事象: A 以外が起こる事象。記号で $\bar{A}$ と書く。

また、必ず起こる事象を全事象といい、 I で表し、決して起こらない事象を空事象といい、 ϕ で表す。たとえば、サイコロを振るという試行において、1, 2, 3, 4, 5, 6 の目のいずれかが出る事象は全事象であり、偶数の目と奇数の目が同時に出る事象は空事象である。事象の間の関係は図 2-1 に示すような図を用いると理解しやすい。

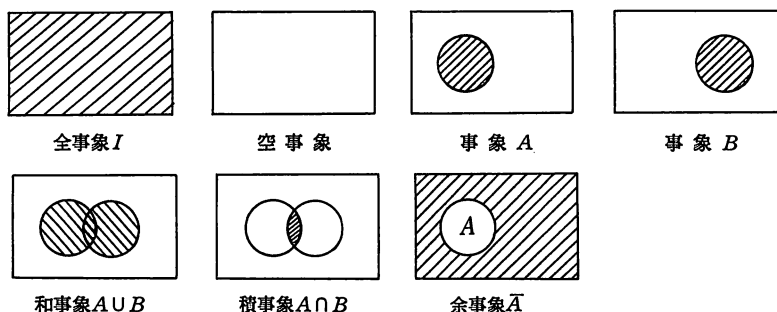
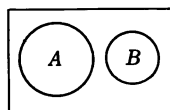


図 2-1

$A \cap B = \phi$ である事象、つまり決して同時には起こりえない事象 A, B を互いに排反事象という。

サイコロを1回投げる試行により起こりうるすべての場合は、サイコロの出



排反事象 $A \cap B = \phi$

図 2-2

る目に注目すれば6通りであり、出る目が偶数か奇数かに注目すれば2通りである。前者の場合に標本空間 $S_1 = \{E_i | E_i = \text{目 } i \text{ が出る事象}\}$ が得られ、後者の場合に標本空間 $S_2 = \{\text{Fodd}, \text{Feven}\}$ が得られることはすでに述べた。 S_2 の事

象 $\{Fodd\}$ と $\{Feven\}$ は

$$\{Fodd\} = \{E_1\} \cup \{E_3\} \cup \{E_5\}, \quad \{Feven\} = \{E_2\} \cup \{E_4\} \cup \{E_6\}$$

である。しかし事象 $\{E_1\}, \{E_2\}, \dots, \{E_6\}$ は他の空でないいかなる事象の和事象としても表現することはできない。このような事象を基本事象という。

問 袋の中に赤玉が3個、白玉が2個入っている。袋の中から2つ取り出すときの標本空間の要素を書きあげよ。

2.1.2 確 率

「標本空間を $S = \{E_1, E_2, E_3, \dots, E_n\}$ とし、 n 個の要素のどれもが、同様の確からしさで起こると考えられる場合、 S の m 個 ($1 \leq m \leq n$) の要素からなる事象 A の起こる確率 $P(A)$ は

$$P(A) = \frac{m}{n}$$

である」。

これをラプラスの定義といい、このようにして定義された確率を数学的確率という。

(例1) 大小2つのサイコロを投げて、2つとも1の目が出る確率を求めよ。

(解) サイコロを投げるとき、どの目も同様の確からしさで出ると考えられる。標本空間 S の要素を E_{ij} とし、

E_{ij} = 大きいサイコロは i の目、小さいサイコロは j の目が出ること

$$(i, j = 1, 2, 3, 4, 5, 6)$$

と考えると、 S の要素の数は36。このうち、 E_{11} は S の一つの要素であるから、両方とも1の目が出る確率は $\frac{1}{36}$ である。

(例2) 袋の中に6個の赤球と4個の白球がある。3個を取り出して、2個が赤、1個が白である確率を求めよ。

(解) 10個の中から3個選出す組み合わせの数は ${}_{10}C_3$ 通りで、考えている場合の数は ${}_6C_2 \times {}_4C_1$ であるから、求める確率は

$$\frac{{}_6C_2 \times {}_4C_1}{{}_{10}C_3} = \frac{1}{2}$$

である。

問 1 5 個のサイコロを投げて、4 個が同じ目を出す確率を求めよ。

問 2 4 人を 1 列に並べるとき、特定の 2 人 A と B が、 AB の順に隣合う確率を求めよ。

問 3 3 枚の貨幣を投げたとき、表の出る枚数がそれぞれ 0, 1, 2, 3 である確率を求めよ。

問 4 袋の中に 6 個の赤球と 4 個の白球と 3 個の黒球がある。3 個を取り出して赤、白、黒がそれぞれ 1 個ずつである確率を求めよ。

問 5 0, 1, 2, …, 9 が書かれた 10 枚の札から 3 枚を抜き出したとき、この 3 枚が (1, 2, 3), (7, 5, 6) 等の続き番号である確率を求めよ。

問 6 前問で 3 枚の札を 1 枚ずつ順番に抜いたとしたとき、順に 1 つずつ大きな数を引き当てる確率を求めよ。

問 7 52 枚のトランプのカードをよくきり、その中から 5 枚ぬいて 5 枚とも全部同種のカードである確率を求めよ。

ラプラスの定義は、“試行によって起こる結果は、すべて同様の確からしさを持つ”ということを前提としているため、実際問題に適用できない場合も多い。たとえば、歪んでいたり、材質が一樣でないサイコロを振った場合、各々の目が出る確率は $\frac{1}{6}$ とはいえない。したがって例 1 のような計算は意味を持たない。このような場合に、ある事象の起こる確率を求めるには、実際に何度も試行して、その結果、 n 回の試行のうち r 回その事象が起こったとすれば、近似的にその事象の起こる確率を $\frac{r}{n}$ と考える。 n をしだいに大きくすれば $\frac{r}{n}$ は一定の値に近づくことが知られている。これを大数の法則と呼び、基本的な定理である (後で証明される)。 n を十分に大きくしたときの $\frac{r}{n}$ を経験的確率という。経験的確率はラプラスの定義による数学的確率を包含している。なぜなら、“試行の結果はすべて同様の確からしさを持つ”ことが先験的に明白な場合には、実際に試行を多数回行えば、経験的確率は数学的確率に近づくからである。

以上、数学的確率と経験的確率の定義を通して、確率の概念を明確にしてきた。しかし、統計学の基本概念として確率を導入するためには、確率の概念を十分に抽象化し、その中に包まれる“あいまいさ”や非本質的部分を取り除く必要がある。

抽象化された概念を公理として設定し、これを出発点として、統計学の諸概

念を組みたててゆくことにしよう。

任意の事象 A に対して、実数 $P(A)$ を結びつけ、この $P(A)$ がつぎの3つの性質をみたすとき、これを事象 A の確率と定義する。

公理 I $0 \leq P(A) \leq 1$

公理 II $P(I) = 1, P(\phi) = 0$

公理 III A と B が排反事象であるとき $P(A \cup B) = P(A) + P(B)$

これらの公理より、つぎの基本的な定理を導くことができる。

〔定理 2.1〕 $P(\bar{A}) = 1 - P(A)$

(証明) $\bar{A} \cup A = I$ であるから

$$P(\bar{A} \cup A) = 1 \quad (\text{公理 II})$$

また $\bar{A} \cap A = \phi$ であるから

$$P(\bar{A} \cup A) = P(\bar{A}) + P(A) \quad (\text{公理 III})$$

$$\therefore P(\bar{A}) = 1 - P(A)$$

〔定理 2.2〕 $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ (加法定理)

(証明) $P(A \cup B) = P\{(A \cap \bar{B}) \cup B\}$

$$= P(A \cap \bar{B}) + P(B) \quad (\text{公理 III})$$

他方 $P(A) = P\{(A \cap \bar{B}) \cup (A \cap B)\}$

$$= P(A \cap \bar{B}) + P(A \cap B) \quad (\text{公理 III})$$

$$\therefore P(A \cap \bar{B}) = P(A) - P(A \cap B)$$

$$\therefore P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

〔定理 2.3〕 2つの事象 A, B が $A \subset B$ であるとき

$$P(A) \leq P(B)$$

である。

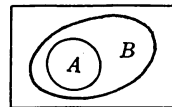
(証明) $P(B) = P\{A \cup (\bar{A} \cap B)\}$

$$= P(A) + P(\bar{A} \cap B) \quad (\text{公理 III})$$

$$\geq P(A) \quad (\text{公理 I})$$

$$\therefore P(B) \geq P(A)$$

つぎに2つの事象 A, B があって、



$A \subset B$

図 2-3

$A = (A \cap \bar{B}) \cup (A \cap B)$ と書ける
互いに排反

$$A \cap B = \{E_{11}, E_{12}, \dots, E_{1r}\}$$

$$A \cap \bar{B} = \{E_{21}, E_{22}, \dots, E_{2a}\}$$

$$\bar{A} \cap B = \{E_{31}, E_{32}, \dots, E_{3b}\}$$

$$\bar{A} \cap \bar{B} = \{E_{41}, E_{42}, \dots, E_{4c}\}$$

$$n = r + a + b + c$$

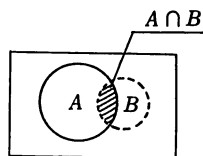


図 2-4

であるとし、どの E_{ij} も同様の確からしきで起こるものとしたとき、ラプラスの定義から

$$\frac{r}{r+a} = \frac{r/n}{(r+a)/n} = \frac{P(A \cap B)}{P(A)}$$

である。この式の左辺は、 A が起こったと仮定したときの B の起こる確率と考えられるから、一般に、

$$\frac{P(A \cap B)}{P(A)}$$

を A のもとでの B の条件つき確率といい、 $P(B/A)$ と書く。すなわち、

$$P(B/A) = \frac{P(A \cap B)}{P(A)}$$

と定義する。

この関係を用いて、直ちにつぎの定理 2.4 を得ることができる。

〔定理 2.4〕 $P(A \cap B) = P(A) \cdot P(B/A)$ (乗法定理)

〔系〕 $P(A \cap B \cap C) = P(A) \cdot P(B/A) \cdot P(C/A \cap B)$

(証明) $P(A \cap B \cap C) = P[(A \cap B) \cap C]$

$$= P(A \cap B) \cdot P(C/A \cap B)$$

$$= P(A) \cdot P(B/A) \cdot P(C/A \cap B)$$

いま $P(B/A) = P(B)$ が成り立つとき、事象 A と事象 B は互いに独立事象であるという。この定義から直ちにつぎの定理 2.5 を得る。

〔定理 2.5〕 事象 A と B が互いに独立事象であるならば

$$P(A \cap B) = P(A) \cdot P(B)$$

(証明) 定理 2.4 と独立事象の定義より明らかである。

〔定理 2.6〕 $P(B/A) = P(B)$ ならば $P(A/B) = P(A)$ である。

$$(証明) \quad P(A/B) = \frac{P(B \cap A)}{P(B)} = \frac{P(B \cap A)}{P(B/A)} = \frac{P(B \cap A)}{P(A \cap B)/P(A)} = P(A)$$

(例3) A, B 2人の学生が数学の問題を解く確率が、それぞれ $3/4, 2/3$ であるとき、ある数学の問題を

- (1) 2人とも解く確率 (3) 少なくとも1人は解く確率
(2) 2人とも解けない確率 (4) B は解けて A は解けない確率

をそれぞれ求めよ。

(解) A がその数学の問題を解く事象を E_1 , B が解く事象を E_2 とすれば, E_1, E_2 は互いに独立事象と考えられるから、

$$(1) \quad P(E_1 \cap E_2) = P(E_1) \cdot P(E_2) = \frac{3}{4} \cdot \frac{2}{3} = \frac{1}{2}$$

$$(2) \quad P(\bar{E}_1 \cap \bar{E}_2) = P(\bar{E}_1) \cdot P(\bar{E}_2) = \left(1 - \frac{3}{4}\right) \left(1 - \frac{2}{3}\right) = \frac{1}{12}$$

- (3) 少なくとも1人が解く事象を E とおけば

$$\bar{E} = \bar{E}_1 \cap \bar{E}_2$$

$$\therefore P(E) = 1 - P(\bar{E}) = 1 - P(\bar{E}_1 \cap \bar{E}_2) = 1 - \frac{1}{12} = \frac{11}{12}$$

$$(4) \quad P(\bar{E}_1 \cap E_2) = P(\bar{E}_1) \cdot P(E_2) = \frac{1}{4} \cdot \frac{2}{3} = \frac{1}{6}$$

(例4) 1つのサイコロを5回投げて、2回1の目が出る確率を求めよ。

(解) 1つのサイコロを投げて1の目が出たとき○印で、出ないときは×印で表すことにすれば、5回のうち2回1の目が出る事象は、つぎの ${}_5C_2=10$ 通りの事象の和事象である。

$$E_1: \text{○○} \times \times \times$$

$$E_2: \text{○} \times \text{○○} \times$$

$$E_3: \times \text{○○} \times \times$$

.....

.....

$$E_{10}: \times \times \times \text{○○}$$

$E_1, E_2, \dots, E_{10}$ は互いに排反であるから

$$P(E_1 \cup E_2 \cup \dots \cup E_{10}) = P(E_1) + P(E_2) + \dots + P(E_{10})$$

サイコロを投げる1回1回は独立であるから

$$P(E_i) = \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)^8 \quad (i=1, 2, \dots, 10)$$

ゆえに、求める確率は

$${}_5C_2 \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)^3$$

である。

(例5) 1回の試行で、ある事象 A の起こる確率が p ならば、これを n 回試行して r 回事象 A の起こる確率は

$${}_nC_rp^rq^{n-r} \quad (q=1-p)$$

である。

問 年賀ハガキのお年玉の抽選に用いられる回転円板の標的に矢を射て、矢のささった位置の数字を読み取る。矢を2回射たとき、つぎの確率を求めよ。

- (1) 2数の和が10でない。
- (2) どちらの数も5以上。
- (3) 数の差が2以下。
- ④ いずれかの数字が3以下。

ただし、0から9までのどの数字に当たる確率も等しいとする。

〔定理2.7〕 n 個の事象 $A_1, A_2, A_3, \dots, A_n$ が互いに排反事象で、 E を任意の事象とする。

$$A_1 \cup A_2 \cup A_3 \cup \dots \cup A_n = I$$

ならば

$$P(A_i/E) = \frac{P(A_i) \cdot P(E/A_i)}{\sum_{j=1}^n P(A_j) \cdot P(E/A_j)}$$

である。(ベイズの定理)

$$(\text{証明}) \quad P(A_i/E) = \frac{P(E \cap A_i)}{P(E)}$$

しかるに

$$\begin{aligned} P(E \cap A_i) &= P(A_i \cap E) \\ &= P(A_i) \cdot P(E/A_i) \end{aligned}$$

かつ

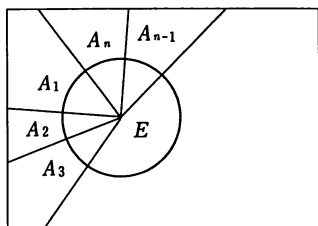


図2-5

$$E = (A_1 \cap E) \cup (A_2 \cap E) \cup \dots \cup (A_n \cap E)$$

であるから

$$\begin{aligned}
 P(E) &= \sum_{j=1}^n P(A_j \cap E) && \text{(公理Ⅲ)} \\
 &= \sum_{j=1}^n P(A_j) \cdot P(E/A_j)
 \end{aligned}$$

したがって

$$P(A_i/E) = \frac{P(A_i) \cdot P(E/A_i)}{\sum_{j=1}^n P(A_j) \cdot P(E/A_j)}$$

(例6) (ポリアの壺の問題) 壺の中に a 個の黒球と b 個の白球がはいっている。その壺の中から1個取り出して、壺の中へ返す。その際その球と同色の球を c 個壺の中に加えることにする。2回目に取り出した球が黒であるとき、最初に取り出した球が黒球である確率を求めよ。

(解) 最初に取り出した球が黒である事象を A_1 、白球である事象を A_2 とし、2回目に取り出した球が黒球である事象を E とすれば、求める確率は $P(A_1/E)$ と書ける。

$$A_1 \cup A_2 = I$$

であるから、ベイズの定理を用いて

$$\begin{aligned}
 P(A_1/E) &= \frac{P(A_1) \cdot P(E/A_1)}{P(A_1) \cdot P(E/A_1) + P(A_2) \cdot P(E/A_2)} \\
 &= \frac{\frac{a}{a+b} \cdot \frac{a+c}{a+b+c}}{\frac{a}{a+b} \cdot \frac{a+c}{a+b+c} + \frac{b}{a+b} \cdot \frac{a}{a+b+c}} \\
 &= \frac{a+c}{a+b+c}
 \end{aligned}$$

(例7) ある製造会社で、使用する1つの部品を a_1, a_2, a_3 の3社から購入している。購入比率はそれぞれ25%, 45%, 30%で、各社の部品中の不良品混入率はそれぞれ10%, 15%, 5%である。

(1) 部品の山から勝手に1つ取り出したときに、それが良品である確率を求めよ。

(2) (1)で良品であることが判明した場合、それが a_1 で作られたものである確率を求めよ。

(解) 1つの部品を取り出した場合に、それが a_1, a_2, a_3 社のものであるという事象をそれぞれ $A_1, A_2,$

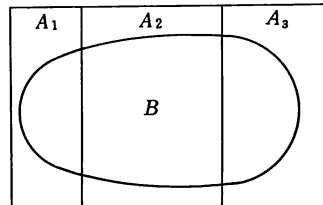


図2-6

A_3 と書き、それが良品であるという事象を B と書く。

$$\begin{aligned}
 (1) \quad P(B) &= P((A_1 \cap B) \cup (A_2 \cap B) \cup (A_3 \cap B)) \\
 &= P(A_1 \cap B) + P(A_2 \cap B) + P(A_3 \cap B) \\
 &= P(A_1) \cdot P(B/A_1) + P(A_2) \cdot P(B/A_2) + P(A_3) \cdot P(B/A_3) \\
 &= 0.25 \times 0.90 + 0.45 \times 0.85 + 0.30 \times 0.95 \\
 &\doteq 0.89
 \end{aligned}$$

(2) ベイズの定理を用いて

$$\begin{aligned}
 P(A_1/B) &= \frac{P(A_1)P(B/A_1)}{P(A_1)P(B/A_1) + P(A_2)P(B/A_2) + P(A_3)P(B/A_3)} = \frac{0.25 \times 0.90}{0.8925} \\
 &\doteq 0.25
 \end{aligned}$$

問 1 100 個の製品の中に 10 個の不良品が含まれている。この中から 10 個の製品を取り出してつぎの場合の確率をそれぞれ求めよ。

- (1) どれも不良品でない確率
- (2) 2 個だけ不良品である確率

問 2 10 本のくじの中に 3 本の当りくじがある。この中から 3 本引いて、少なくとも 1 本当たる確率を求めよ。

問 3 甲、乙、丙がこの順で貨幣を投げて、はじめに表を出したものを勝ちとする。各人の勝つ確率を求めよ。ただし勝負が決まるまでくり返すものとする。

無限個の事象 $A_1, A_2, \dots, A_n, \dots$ について、どの 2 つも排反ならば

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n)$$

これを用いよ。ここで

$$\bigcup_{n=1}^{\infty} A_n = A_1 \cup A_2 \cup \dots \cup A_n \cup \dots$$

問 4 街道に沿って a, b, c の 3 か所に交通信号があり、これが青になる時間は 1 分間にそれぞれ 30 秒、40 秒、50 秒である。この街道を通る自動車についてつぎの場合の確率を求めよ。

- (1) 3 か所とも止まらずに通過できる確率
- (2) 3 か所とも止められる確率
- (3) 1 か所だけ止められる確率

問 5 5 枚の 10 円玉を投げて、表が出る枚数を X 枚とするとき

$$X = 0, 1, 2, 3, 4, 5$$

である確率をそれぞれ求めよ。

問 6 2 つのサイコロを投げて、その目の和を X とするとき、つぎの場合の確率を求

めよ.

- (1) $X = 10$, (2) $X \leq 5$, (3) $X > 9$

問7 ある電球の不良率が2%であるとき, この中から任意に10個抽出して, 不良品が2個以下である確率を求めよ.

問8 袋の中に均質等大の球, 赤が7個, 白が5個はいつている. 1個ずつ取り出して2回目に白球である確率を求めよ(ただし取り出した球はもどさないものとする).

2.1.3 確率変数と分布関数

1つの試行において, 注目している偶然変動量にある変数の実現値と考えたとき, その変数を確率変数といい, X, Y, Z などで表す. 確率変数は, (1)それにより事象の表現が簡単になる. (2)いくつかの試行を組み合わせると新しい試行とみなすとき, 確率変数を用いると構造が明確になる.

(例1) サイコロを振るという試行で, 出る目を表す確率変数を X とすると
事象「1の目が出る」は $(X=1)$,

$$\begin{aligned}\text{事象「5以上の目が出る」は } (X \geq 5) &= (X=5 \text{ or } X=6) \\ &= (X=5) \cup (X=6)\end{aligned}$$

この事象の確率を $P(X=1)$, $P(X \geq 5)$ で表す. したがってサイコロが理想的ならば,

$$\begin{aligned}P(X=1) &= \frac{1}{6} \\ P(X \geq 5) &= P(X=5 \text{ or } X=6) = P\{(X=5) \cup (X=6)\} \\ &= P(X=5) + P(X=6) = \frac{2}{6}\end{aligned}$$

である.

一般に変数 X が離散的な値

$$X = x_1, x_2, x_3, \dots, x_n$$

をとり, かつ確率 $P(X=x_i)$ ($i=1, 2, 3, \dots, n$) があわせ考えられるとき, このような変数 X を離散的確率変数といい

$$P(X=x_i) = p_i \quad (i=1, 2, 3, \dots, n)$$

と書く. この p_i を x_i における X の確率密度, p_i ($i=1, 2, \dots, n$) を X の確

率分布という。このとき、つぎが成り立つ。

$$(1) \quad p_i \geq 0 \quad (i=1, 2, \dots, n)$$

$$(2) \quad \sum_{i=1}^n p_i = 1$$

(1)は確率の性質より、(2)はつぎのことから導かれる。

$$(X=x_1) \cup (X=x_2) \cup (X=x_3) \cup \dots \cup (X=x_n) = I$$

$$(X=x_i) \cap (X=x_j) = \phi \quad (i \neq j)$$

また

$$P(X \leq x) = F(x)$$

とおき、 $F(x)$ を X の分布関数と呼ぶ。

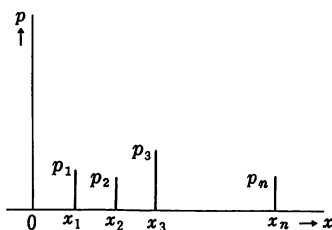


図 2-7

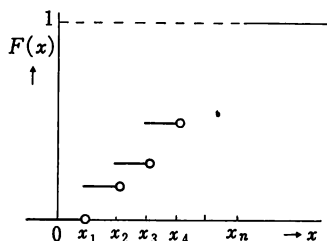


図 2-8

いま、 $x_1, x_2, x_3, \dots, x_n$ を小さい順に並び換えて、これをあらためて $x_1, x_2, x_3, \dots, x_n$ とし、これに対応して確率密度も入れ換えを行えば

$$F(x) = \begin{cases} 0 & (x < x_1) \\ \sum_{i=1}^k p_i & (x_k \leq x < x_{k+1}) \quad (k=1, 2, \dots, n-1) \\ 1 & (x \geq x_n) \end{cases}$$

が成立する。

つぎに連続的率変数を、例を通して考えてみよう。

(例 2) 図のように小球が円形の軌道上を運動している。いまこの小球を、無意識に停止させたとき、軌道上の定点Aから小球までの距離を左まわりに測ったとする。この距離を X とおけば、 X のとり得る値の範囲は

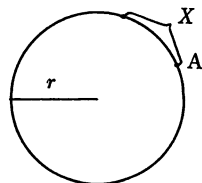


図 2-9

$$0 \leq X < 2\pi r$$

で、 0 と $2\pi r$ の間の値を連続的にとる。このとき $0 \leq x < 2\pi r$ なる x を取れば、

$$P(0 \leq X \leq x) = \frac{x}{2\pi r}$$

である(この場合 $P(X=0)$ および $P(X=x)$ は 0 と考えられるから上式の等号は付けても付けなくても同じである)。

離散的なときと同様に

$$P(X \leq x) = F(x)$$

とおき、 $F(x)$ を求めると

$$F(x) = \begin{cases} 0 & (x < 0) \\ \frac{x}{2\pi r} & (0 \leq x < 2\pi r) \\ 1 & (x \geq 2\pi r) \end{cases}$$

となる。

一般に、変数 X が連続的な値をとり、任意の実数 a, b に対して、確率

$$P(a < X \leq b)$$

があわせ考えられるとき、 X を連続的確率変数といい、

$$P(X \leq x) = F(x)$$

とおき、 $F(x)$ を X の分布関数という。もし

$$F(x) = \int_{-\infty}^x p(x) dx, \quad p(x) \geq 0$$

なる関数 $p(x)$ が存在するならば、この $p(x)$ を確率変数 X の確率密度関数という。上式より

$$p(x) = \frac{dF(x)}{dx} \quad \text{および} \quad \int_{-\infty}^{\infty} p(x) dx = 1$$

を得る。

例2において、 X の分布関数 $F(x)$ 、確率密度関数 $p(x)$ のグラフを描くと、図2-8、図2-9を得る。このような X の分布を一様分布という。

確率変数 X が離散的であっても連続的であっても、 X の分布関数 $F(x)$ はつぎの性質をもっている。

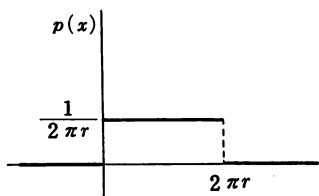


図 2-10

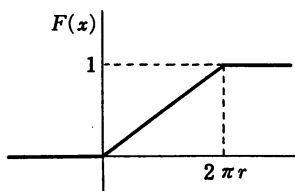


図 2-11

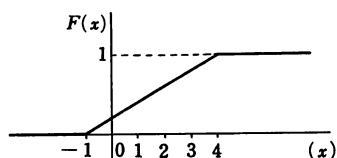
(1) $a < b$ ならば $F(a) \leq F(b)$

(2) $\lim_{x \rightarrow +\infty} F(x) = 1$

(3) $\lim_{x \rightarrow -\infty} F(x) = 0$

問 1 5 枚の貨幣を投げるとき、表が出る枚数を確率変数 X として、 X の確率密度と X の分布関数を求めよ。

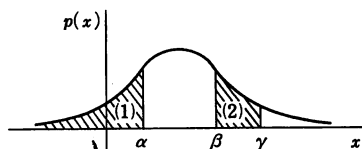
問 2 分布関数 $F(x)$ が右のグラフによって与えられるとき、



$$P(X \leq 2), P(1 < X \leq 3)$$

の値を求めよ。

問 3 X の確率密度関数 $p(x)$ が右のグラフで与えられているとき、斜線をした (1), (2) の部分の面積は何を意味するか。



2.1.4 2次元分布

(1) X, Y が離散的確率変数のとき:

X, Y のとりうる値をそれぞれ

$$x_1, x_2, x_3, \dots, x_m; y_1, y_2, y_3, \dots, y_n$$

とし、

$$P(X=x_i, Y=y_j) = p_{ij} \quad (i=1, 2, \dots, m; j=1, 2, \dots, n)$$

とおき、 p_{ij} を $X=x_i, Y=y_j$ における 2 次元確率密度という。このとき、

$$p_{ij} \geq 0, \quad \sum_{i=1}^m \sum_{j=1}^n p_{ij} = 1$$

が成り立つ。

$X=x_i$ における確率密度を p_i , $Y=y_j$ における確率密度を q_j とすれば,

$$\begin{aligned} p_i &= P(X=x_i) \\ &= P\{(X=x_i, Y=y_1) \cup (X=x_i, Y=y_2) \cup \cdots \cup (X=x_i, Y=y_n)\} \\ &= \sum_{j=1}^n P(X=x_i, Y=y_j) \\ &= \sum_{j=1}^n p_{ij} \end{aligned}$$

$$\therefore p_i = \sum_{j=1}^n p_{ij}$$

同様に

$$q_j = \sum_{i=1}^m p_{ij}$$

(2) X, Y が連続的確率変数のとき:

確率変数 X, Y の値の組を xy -平面上の座標と考え、座標点 (X, Y) が任意の領域 D に含まれる確率を $P\{(X, Y) \in D\}$ と書くとき

$$P\{(X, Y) \in D\} = \iint_D p(x, y) dx dy, \quad p(x, y) \geq 0 \quad (2.1.1)$$

なる関係が常に成り立つ2変数関数 $p(x, y)$ が存在するならば、この $p(x, y)$ を X, Y の2次元確率密度関数という(図2-12参照).

いま、領域 D を図2-13のようにとれば、(2.1.1)は

$$\begin{aligned} &P\{(X, Y) \in D\} \\ &= P(-\infty < X \leq x, -\infty < Y \leq y) \\ &= \int_{-\infty}^x \int_{-\infty}^y p(x, y) dx dy \quad (2.1.2) \end{aligned}$$

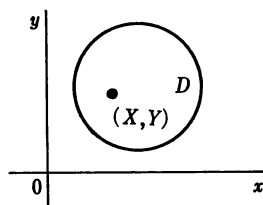


図2-12

と書けるから、今後、(2.1.2)をみたす負にならない関数 $p(x, y)$ を確率変数 X, Y の2次元確率密度関数と呼ぶことにする。

連続的確率変数 X, Y の確率密度関数をそれぞれ $f(x), g(y)$, X, Y の2次元確率密度関数を $p(x, y)$ とおけば、定義より

$$P(X \leq x) = \int_{-\infty}^x f(x) dx$$

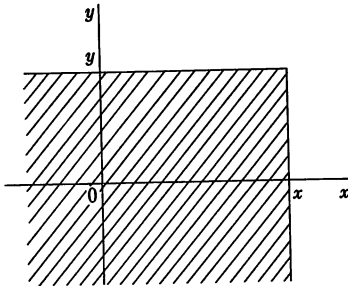


図 2-13

かつ

$$P(X \leq x) = P(X \leq x, -\infty < Y < +\infty)$$

であるから,

$$\int_{-\infty}^x f(x) dx = \int_{-\infty}^x \int_{-\infty}^{\infty} p(x, y) dx dy$$

この両辺を x で微分して,

$$f(x) = \int_{-\infty}^{+\infty} p(x, y) dy$$

同様に

$$g(y) = \int_{-\infty}^{+\infty} p(x, y) dx$$

が成り立つ.

〔定理 2.8〕 連続的確率変数 X, Y の 2 次元確率密度関数 $p(x, y)$ が存在し,

$$X = f_1(U, V), \quad Y = f_2(U, V)$$

であるとき, 連続的確率変数 U, V の 2 次元確率密度関数は

$$p(f_1(u, v), f_2(u, v)) |J|$$

である. ただし

$$J = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} \neq 0 \quad \begin{pmatrix} x = f_1(u, v) \\ y = f_2(u, v) \end{pmatrix} \quad (J \text{ はヤコビ行列式とよばれる})$$

(証明) $X = f_1(U, V), \quad Y = f_2(U, V)$ によって UV -平面上の領域 E が XY -平面上の領域 D に写像されたとする.

このとき,

$$P\{(U, V) \in E\} = P\{(X, Y) \in D\} = \iint_D p(x, y) dx dy$$

二重積分の性質により

$$\iint_D p(x, y) dx dy = \iint_E p(f_1(u, v), f_2(u, v)) |J| du dv$$

であるから,

$$P\{(U, V) \in E\} = \iint_E p(f_1(u, v), f_2(u, v)) |J| du dv$$

したがって、確率変数 U, V の2次元確率密度関数は

$$p(f_1(u, v), f_2(u, v)) |J|.$$

§ 2.2 確率変数の性質

2.2.1 確率変数の独立性

X, Y が離散的確率変数で、それぞれ

$$x_1, x_2, \dots, x_m; y_1, y_2, \dots, y_n$$

の値をとるとき

$$P(X=x_i, Y=y_j) = P(X=x_i) \cdot P(Y=y_j)$$

$$(i=1, 2, \dots, m; j=1, 2, \dots, n)$$

なる関係が常に成立するとき、確率変数 X と Y は互いに独立であるという。

$X=x_i, Y=y_j$ のときの確率密度をそれぞれ f_i, g_j とし、またこのときの2次元確率密度を p_{ij} とすると

$$P(X=x_i) = f_i, P(Y=y_j) = g_j, P(X=x_i, Y=y_j) = p_{ij}$$

であるから、 X, Y が互いに独立のときは、定義式より直ちに

$$p_{ij} = f_i \cdot g_j$$

を得る。

X, Y が連続的確率変数であるときは、任意の実数 a, b, c, d に対し

$$P\{a < X \leq b, c < Y \leq d\} = P(a < X \leq b) \cdot P(c < Y \leq d)$$

なる関係が常に成立するとき、確率変数 X と Y は互いに独立であるという。

X, Y の確率密度関数をそれぞれ $f(x), g(y)$, X, Y の2次元確率密度関数を $p(x, y)$ とおけば、定義より

$$P(a < X \leq b) = \int_a^b f(x) dx, P(c < Y \leq d) = \int_c^d g(y) dy$$

$$P(a < X \leq b, c < Y \leq d) = \int_a^b \int_c^d p(x, y) dx dy$$

であるから、 X, Y が互いに独立のときは

$$\begin{aligned}
 \int_a^b \int_c^d p(x, y) dx dy &= P(a < X \leq b, c < Y \leq d) = P(a < X \leq b) \cdot P(c < Y \leq d) \\
 &= \int_a^b f(x) dx \cdot \int_c^d g(y) dy = \int_a^b \int_c^d f(x) g(y) dx dy \\
 \therefore p(x, y) &= f(x) g(y)
 \end{aligned}$$

定義より明らかなように、確率変数 X, Y の独立性と、事象 A, B の独立性とは同義である。

2.2.2 確率変数の演算

確率変数 X, Y に対して演算を導入する。演算の結果はまた、確率変数である。

たとえば大小2つのサイコロを投げるとき、大きいサイコロの出る目を X 、小さいサイコロの出る目を Y とおけば、 X, Y はそれぞれ離散的確率変数で、 $X+Y$ は大小2つのサイコロの目の和を表す確率変数である。したがって、 $X+Y$ のとりうる値は

$$2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12$$

一般に X, Y が離散的確率変数で、そのとりうる値がそれぞれ

$$a_1, a_2, \dots, a_m; b_1, b_2, \dots, b_n$$

のとき、確率変数 $X+Y$ は

$$a_i + b_j \quad (i=1, 2, \dots, m; j=1, 2, \dots, n)$$

なる値をとる離散的確率変数である。 $X+Y$ を X と Y の和という。同様に、とりうる値が

$$a_i - b_j, a_i \times b_j, a_i \div b_j \quad (i=1, 2, \dots, m; j=1, 2, \dots, n)$$

であるような離散的確率変数をそれぞれ、確率変数 X, Y の差、積、商といい

$$X-Y, XY, X/Y$$

で表す。

また、 X が $a_1, a_2, \dots, a_m$ の値をとる離散的確率変数であれば、 $\alpha X + \beta$ は

$$\alpha a_1 + \beta, \alpha a_2 + \beta, \dots, \alpha a_m + \beta$$

の値をとる離散的確率変数である。いま

$$Y = \alpha X + \beta$$

とおけば

$$P(Y=\alpha a_i+\beta)=P(X=a_i) \quad (i=1, 2, \dots, m)$$

である.

(例1) 1つのサイコロを2回投げて, 1の目が出る回数を X 回とおけば, X は 0, 1, 2 の値をとる離散的確率変数である. いま k 回目に 1 の目が出たときに 1, 1 以外の目が出たときに 0 の値をとる確率変数を X_k とおくと,

$$X = X_1 + X_2$$

と書ける. さらに X_1 と X_2 は独立であるから, X の分布はつぎのようになる.

$$P(X=0) = P(X_1=0, X_2=0) = P(X_1=0)P(X_2=0) = \left(\frac{5}{6}\right)^2 = \frac{25}{36}$$

$$P(X=1) = P\{(X_1=1, X_2=0) \cup (X_1=0, X_2=1)\}$$

$$= P(X_1=1)P(X_2=0) + P(X_1=0)P(X_2=1) = \frac{10}{36}$$

$$P(X=2) = P(X_1=1, X_2=1) = P(X_1=1)P(X_2=1) = \frac{1}{36}$$

以上, 離散的確率変数 X, Y について, 和 $X+Y$, 差 $X-Y$, 積 XY , 商 X/Y および $\alpha X + \beta$ を考えたが, 同様のことが連続的確率変数 X, Y についても定義できる.

〔定理 2.9〕 確率変数 X, Y の分布関数をそれぞれ $F(x), G(y)$ とし,

$$Y = \alpha X + \beta \quad (\alpha > 0)$$

なる関係があれば

$$G(y) = F\left(\frac{y-\beta}{\alpha}\right) \quad (2.2.1)$$

特に, X, Y が連続的確率変数のとき X, Y の確率密度関数を $f(x), g(y)$ とおけば,

$$g(y) = \frac{1}{\alpha} f\left(\frac{y-\beta}{\alpha}\right) \quad (2.2.2)$$

(証明) $Y \leq y$ ならば $X \leq (y-\beta)/\alpha$ であるから

$$G(y) = P(Y \leq y) = P\left(X \leq \frac{y-\beta}{\alpha}\right) = F\left(\frac{y-\beta}{\alpha}\right)$$

特に, X, Y が連続的確率変数であれば, (2.2.1) の両辺を y で微分して,

$$g(y) = \frac{1}{\alpha} f\left(\frac{y-\beta}{\alpha}\right)$$

(例 2) 連続的確率変数 X の確率密度関数 $f(x)$ が

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

ならば, $Y = \sigma X + \mu$ とおいたとき Y の確率密度関数 $g(y)$ は, 定理 2.9 により

$$g(y) = \frac{1}{\sigma} f\left(\frac{y-\mu}{\sigma}\right) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(y-\mu)^2/2\sigma^2}$$

となる.

後で説明するが (§ 3.3), 確率密度関数が $f(x)$ の式によって与えられるものを $N(0, 1)$ 型の正規分布, $g(y)$ の式によって与えられるものを $N(\mu, \sigma^2)$ 型の正規分布という.

〔定理 2.10〕 X, Y の 2 次元確率密度関数が, $p(x, y)$ で与えられるとき, $U = X/Y$ の確率密度関数 $p(u)$ は

$$p(u) = \int_{-\infty}^{\infty} p(uv, v) |v| dv$$

である.

(証明) $X/Y = U, Y = V$ とおけば

$$X = UV, Y = V$$

であるから

$$J = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} v & u \\ 0 & 1 \end{vmatrix} = v$$

ゆえに定理 2.8 により, U, V の 2 次元確率密度関数は

$$p(uv, v) |v|$$

である. したがって, $U = X/Y$ の確率密度関数は

$$p(u) = \int_{-\infty}^{\infty} p(uv, v) |v| dv$$

問 確率変数 X, Y の2次元確率密度関数が $p(x, y)$ によって与えられるとき, $U = X + Y$ の確率密度関数は, $V = Y$ として

$$p(u) = \int_{-\infty}^{\infty} p(u-v, v) dv$$

であることを導け.

2.2.3 平均と分散

ある大学の学生について, 学生の身長を集まりを考える. これは有限個の実数の集まりで有限母集団という. この有限母集団の中から無作為に選り出した1つの実数を X とすれば, X のとりうる値は離散的で, かつその個数は有限個である. X がその中の任意の

表 2-1

1つの数である確率は, その数に対する相対度数であるから, X は離散的確率変数である.

X	x_1	x_2	……	x_i	……	x_{n-1}	x_n
度数	a_1	a_2	……	a_i	……	a_{n-1}	a_n
相対度数	p_1	p_2	……	p_i	……	p_{n-1}	p_n

一般に有限母集団が上の度数分布表で与えられるとき, X のとりうる値は

$$x_1, x_2, \dots, x_n$$

で

$$P(X=x_i) = p_i \quad (i=1, 2, \dots, n)$$

である. 以上のように有限母集団に対して, 1つの離散的確率変数を結びつけることができる.

つぎに日本全体の大学生について, その身長を集まりを考える. これは非常に多くの実数の集まりで無限母集団という. この中から任意に選り出した1つの実数を X とすれば, X のとりうる値は連続的と考えられ, かつ, 任意の実数 $a, b (a < b)$ に対して $P(a < X \leq b)$ が考えられるから, X は連続的確率変数である.

一般に無限母集団に対して, 連続的確率変数か, 離散的確率変数を結びつけることができる.

離散的確率変数 X が $X = x_1, x_2, \dots, x_n$ をとり, かつ

$$P(X=x_i) = p_i \quad (i=1, 2, \dots, n)$$

のとき, X の平均 $E(X)$ を

$$E(X) = \sum_{i=1}^n x_i p_i$$

と定義すれば、ある母集団に結びついた離散的確率変数 X の平均 $E(X)$ とその母集団の平均とが一致していることは明らかであろう。

以上のことから、母集団の性質を調べるためには、それに結びついた確率変数の性質を調べればよいことになる。この意味で今後、この確率変数 X を母集団を定める確率変数という。

また、一般に

$$E(f(X)) = \sum_{i=1}^n f(x_i) p_i$$

を $f(X)$ の平均という。

つぎに確率変数 X が連続的なとき、 X の確率密度関数を $p(x)$ とすれば、確率変数 X の平均 $E(X)$ を

$$E(X) = \int_{-\infty}^{\infty} x p(x) dx$$

によって定義する。この定義は離散的確率変数 X の平均の自然な拡張になっている。さらに

$$E(f(X)) = \int_{-\infty}^{\infty} f(x) p(x) dx$$

と定義し、 $E(f(X))$ を $f(X)$ の平均という。

〔定理 2.11〕 確率変数 X, Y に対し、

$$E(X+Y) = E(X) + E(Y)$$

$$E(\alpha X + \beta) = \alpha E(X) + \beta$$

が成り立つ。

(証明) (1) X, Y が離散的確率変数のとき：

X, Y のとりうる値をそれぞれ

$$x_1, x_2, \dots, x_m; y_1, y_2, \dots, y_n$$

とし

$$P(X=x_i) = f_i, P(Y=y_j) = g_j$$

かつ

$$P(X=x_i, Y=y_j) = p_{ij} \quad (i=1, 2, \dots, m; j=1, 2, \dots, n)$$

とおけば

$$\begin{aligned} E(X+Y) &= \sum_{i=1}^m \sum_{j=1}^n (x_i + y_j) p_{ij} \\ &= \sum_{i=1}^m \sum_{j=1}^n x_i p_{ij} + \sum_{i=1}^m \sum_{j=1}^n y_j p_{ij} \\ &= \sum_{i=1}^m x_i \sum_{j=1}^n p_{ij} + \sum_{j=1}^n y_j \sum_{i=1}^m p_{ij} \\ &= \sum_{i=1}^m x_i f_i + \sum_{j=1}^n y_j g_j \quad (2.1.4 \text{ 節}) \\ &= E(X) + E(Y) \end{aligned}$$

$$\begin{aligned} E(\alpha X + \beta) &= \sum_{i=1}^m (\alpha x_i + \beta) f_i = \alpha \sum_{i=1}^m x_i f_i + \beta \sum_{i=1}^m f_i \\ &= \alpha E(X) + \beta \end{aligned}$$

(2) X, Y が連続的確率変数のとき:

X, Y の確率密度関数をそれぞれ $f(x)$, $g(y)$, 2次元確率密度関数を $p(x, y)$ とおき, $E(X+Y)$ を離散的確率変数の場合の拡張として極限の概念を用いれば

$$\begin{aligned} &P(x_{i-1} < X \leq x_i, y_{j-1} < Y \leq y_j) \\ &\doteq p(x_i, y_j)(x_i - x_{i-1})(y_j - y_{j-1}) \\ &\doteq p(x_i, y_j) \Delta x_i \Delta y_j \end{aligned}$$

であるから,

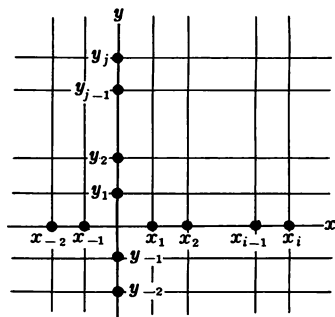


図 2-14

$$\begin{aligned} E(X+Y) &= \lim_{\substack{\Delta x_i \rightarrow 0 \\ \Delta y_j \rightarrow 0}} \sum_{i=-\infty}^{+\infty} \sum_{j=-\infty}^{+\infty} (x_i + y_j) p(x_i, y_j) \Delta x_i \Delta y_j \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x+y) p(x, y) dx dy \\ &= \int_{-\infty}^{+\infty} x \left\{ \int_{-\infty}^{+\infty} p(x, y) dy \right\} dx + \int_{-\infty}^{+\infty} y \left\{ \int_{-\infty}^{+\infty} p(x, y) dx \right\} dy \end{aligned}$$

$$= \int_{-\infty}^{+\infty} xf(x)dx + \int_{-\infty}^{+\infty} yg(y)dy \quad (2.1.4 \text{ 節})$$

$$= E(X) + E(Y)$$

$$E(\alpha X + \beta) = \int_{-\infty}^{+\infty} (\alpha x + \beta)f(x)dx$$

$$= \alpha \int_{-\infty}^{+\infty} xf(x)dx + \beta \int_{-\infty}^{+\infty} f(x)dx$$

$$= \alpha E(X) + \beta$$

〔定理 2.12〕 確率変数 X, Y が互いに独立ならば

$$E(XY) = E(X) \cdot E(Y)$$

である。

(証明) (記号は定理 2.11 の約束に従う)

(1) X, Y が離散的確率変数のとき:

X, Y は互いに独立であるから, 2.2.2 節より

$$p_{ij} = f_i \cdot g_j$$

したがって

$$\begin{aligned} E(XY) &= \sum_{i=1}^m \sum_{j=1}^n x_i y_j p_{ij} \\ &= \sum_{i=1}^m \sum_{j=1}^n x_i y_j f_i g_j \\ &= \sum_{i=1}^m x_i f_i \sum_{j=1}^n y_j g_j \\ &= E(X) \cdot E(Y) \end{aligned}$$

(2) X, Y が連続的確率変数のとき:

2.2.2 節より $p(x, y) = f(x)g(y)$ であるから

$$\begin{aligned} E(XY) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xy p(x, y) dx dy \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xy f(x) g(y) dx dy \\ &= \int_{-\infty}^{+\infty} x f(x) dx \int_{-\infty}^{+\infty} y g(y) dy \end{aligned}$$

$$= E(X)E(Y)$$

確率変数 X の分散 $\sigma^2(X)$ を定義しよう. X の平均を $E(X)=\mu$ とおくと

$$\sigma^2(X) = \begin{cases} \sum_{i=1}^n (x_i - \mu)^2 p_i & (X \text{ が離散的なとき}) \\ \int_{-\infty}^{\infty} (x - \mu)^2 p(x) dx & (X \text{ が連続的なとき}) \end{cases}$$

と定義する. ただし $p_i, p(x)$ はそれぞれ X の確率密度, 確率密度関数である.

定義より

$$\begin{aligned} \sigma^2(X) &= E\{(X - E(X))^2\} \quad (\mu = E(X)) \\ &= E(X^2 - 2\mu X + \mu^2) \\ &= E(X^2) - 2\mu E(X) + \mu^2 \\ &= E(X^2) - \{E(X)\}^2 \\ \therefore \sigma^2(X) &= E(X^2) - \{E(X)\}^2 \end{aligned}$$

〔定理 2.13〕 確率変数 X, Y が互いに独立ならば

$$\sigma^2(X+Y) = \sigma^2(X) + \sigma^2(Y)$$

である.

$$(\text{証明}) \quad E(X) = \mu_x$$

$$E(Y) = \mu_y$$

$$E(X+Y) = \mu_x + \mu_y \quad (\text{定理 2.11 より})$$

したがって

$$\begin{aligned} \sigma^2(X+Y) &= E\{(X+Y)^2\} - \{E(X+Y)\}^2 \\ &= E(X^2) + E(Y^2) + 2E(XY) - (\mu_x + \mu_y)^2 \\ &= E(X^2) + E(Y^2) + 2E(X)E(Y) - (\mu_x + \mu_y)^2 \\ &\quad (\text{定理 2.12 より}) \\ &= E(X^2) + E(Y^2) + 2\mu_x\mu_y - (\mu_x + \mu_y)^2 \\ &= \{E(X^2) - \mu_x^2\} + \{E(Y^2) - \mu_y^2\} \\ &= \sigma^2(X) + \sigma^2(Y) \end{aligned}$$

〔定理 2.14〕 確率変数 X に対して

$$\sigma^2(\alpha X + \beta) = \alpha^2 \sigma^2(X)$$

が成り立つ.

$$\begin{aligned}
 (\text{証明}) \quad \sigma^2(\alpha X + \beta) &= E\{(\alpha X + \beta)^2\} - \{E(\alpha X + \beta)\}^2 \\
 &= E(\alpha^2 X^2 + 2\alpha\beta X + \beta^2) - \{\alpha E(X) + \beta\}^2 \\
 &= \alpha^2 E(X^2) + 2\alpha\beta E(X) + \beta^2 \\
 &\quad - \alpha^2 \{E(X)\}^2 - 2\alpha\beta E(X) - \beta^2 \\
 &= \alpha^2 [E(X^2) - \{E(X)\}^2] \\
 &= \alpha^2 \sigma^2(X)
 \end{aligned}$$

(例 1) 5 枚の貨幣を投げて表の出る枚数を X とするとき, X の平均と分散を求めよ.

$$(\text{解}) \quad P(X=k) = p_k = {}_5C_k \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{5-k} = {}_5C_k \left(\frac{1}{2}\right)^5 \quad (k=0, 1, 2, \dots, 5)$$

$$\begin{aligned}
 E(X) &= \sum_{k=0}^5 k p_k = \sum_{k=0}^5 k {}_5C_k \left(\frac{1}{2}\right)^5 \\
 &= \left(\frac{1}{2}\right)^5 (1 \times 5 + 2 \times 10 + 3 \times 10 + 4 \times 5 + 5 \times 1) = \frac{80}{32} = \frac{5}{2}
 \end{aligned}$$

$$\begin{aligned}
 \sigma^2(X) &= E(X^2) - \{E(X)\}^2 = \sum_{k=0}^5 k^2 p_k - \left(\frac{5}{2}\right)^2 \\
 &= \left(\frac{1}{2}\right)^5 (1 \times 5 + 4 \times 10 + 9 \times 10 + 16 \times 5 + 25 \times 1) - \left(\frac{25}{4}\right) = \frac{240}{32} - \frac{25}{4} = \frac{40}{32} = \frac{5}{4}
 \end{aligned}$$

(例 2) 確率変数 X の確率密度関数 $p(x)$ が

$$p(x) = \begin{cases} \frac{x}{2} & (0 \leq x \leq 2) \\ 0 & (x < 0, x > 2) \end{cases}$$

で与えられているとき, X の平均, 分散を求めよ.

$$\begin{aligned}
 (\text{解}) \quad E(X) &= \int_{-\infty}^{\infty} x p(x) dx = \int_0^2 x \left(\frac{x}{2}\right) dx = \left[\frac{x^3}{6}\right]_0^2 = \frac{4}{3} \\
 \sigma^2(X) &= \int_{-\infty}^{\infty} x^2 p(x) dx - \left(\frac{4}{3}\right)^2 = \int_0^2 x^2 \cdot \frac{x}{2} dx - \left(\frac{4}{3}\right)^2 \\
 &= \left[\frac{x^4}{8}\right]_0^2 - \left(\frac{4}{3}\right)^2 = \frac{2}{9}
 \end{aligned}$$

問 1 2 つのサイコロを投げるとき, 目の和, 積を各々確率変数とするととき, その平均, 分散を求めよ.

問 2 確率変数 X の確率密度関数 $p(x)$ が

$$p(x) = \begin{cases} e^{-x} & (x \geq 0) \\ 0 & (x < 0) \end{cases}$$

のとき, $E(X), \sigma^2(X)$ を求めよ.

2.2.4 チェビシェフの不等式と大数の法則

1.2節において, n 個の数値 $x_1, x_2, \dots, x_n$ の集まりに対してチェビシェフの不等式を説明したが, この拡張として確率変数 X についてのチェビシェフの不等式を説明しよう.

〔定理 2.15〕 確率変数 X の平均を $E(X)$, 分散を $\sigma^2(X)$ とすれば, つぎの不等式

$$P\{|X - E(X)| \leq \lambda \sigma(X)\} > 1 - \frac{1}{\lambda^2} \quad (\lambda > 1)$$

が成り立つ.

(証明) (1) 確率変数 X が離散的なとき:

X が $X = x_1, x_2, \dots, x_n$ をとり,

$$P(X = x_i) = p_i \quad \text{かつ} \quad E(X) = \mu$$

とおけば

$$\sigma^2(X) = \sum_{i=1}^m (x_i - \mu)^2 p_i$$

である. いま

$$|x_i - \mu| > \lambda \sigma(X)$$

なるような i を選び出し, これを $i_1, i_2, \dots, i_r$ とすれば

$$\begin{aligned} \sigma^2(X) &\geq (x_{i_1} - \mu)^2 p_{i_1} + (x_{i_2} - \mu)^2 p_{i_2} + \dots + (x_{i_r} - \mu)^2 p_{i_r} \\ &> \lambda^2 \sigma^2(X) (p_{i_1} + p_{i_2} + \dots + p_{i_r}) \end{aligned}$$

$$\therefore p_{i_1} + p_{i_2} + \dots + p_{i_r} < \frac{1}{\lambda^2}$$

すなわち $P\{|X - \mu| > \lambda \sigma(X)\} < \frac{1}{\lambda^2}$ であるから,

$$P\{|X - \mu| \leq \lambda \sigma(X)\} = 1 - P\{|X - \mu| > \lambda \sigma(X)\} > 1 - \frac{1}{\lambda^2}$$

(2) 確率変数が連続なとき:

X の確率密度関数を $p(x)$ とおけば

$$\begin{aligned}\sigma^2(X) &= \int_{-\infty}^{\infty} (x-\mu)^2 p(x) dx \geq \int_{|x-\mu| > \lambda\sigma(X)} (x-\mu)^2 p(x) dx \\ &> \lambda^2 \sigma^2(X) \int_{|x-\mu| > \lambda\sigma(X)} p(x) dx = \lambda^2 \sigma^2(X) P\{|X-\mu| > \lambda\sigma(X)\}\end{aligned}$$

$$\therefore P\{|X-\mu| > \lambda\sigma(X)\} < \frac{1}{\lambda^2}$$

$$\therefore P\{|X-\mu| \leq \lambda\sigma(X)\} > 1 - \frac{1}{\lambda^2}$$

このチェビシェフの不等式において $\lambda=3$ を代入すれば

$$P\{|X-\mu| \leq 3\sigma(X)\} > \frac{8}{9}$$

となり、特にこの不等式を 3σ の定理という。

(例1) ある機械から作られる製品の1日の不良率の平均は6%で、標準偏差は2%であった。不良率が3%~9%の間にある確率を求めよ。

(解) 不良率を $X\%$ とすれば

$$E(X) = 6, \quad \sigma(X) = 2$$

である。 $|X-6| \leq 3$ なる確率を求めればよいから、 $\lambda\sigma(X)=3$ とおき

$$\lambda = \frac{3}{2}$$

ゆえにチェビシェフの不等式から

$$P(|X-6| \leq 3) = P\left\{|X-6| \leq \frac{3}{2} \cdot \sigma(X)\right\} > 1 - \frac{1}{(3/2)^2} = \frac{5}{9}$$

大数の法則をチェビシェフの不等式を用いて説明しよう。

いま、互いに独立な確率変数

$$X_1, X_2, \dots, X_n, \dots$$

があって、

$$E(X_i) = \mu, \quad \sigma^2(X_i) \leq k \quad (i=1, 2, \dots, n, \dots)$$

とする。

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad \text{とおけば}$$

$$\begin{aligned}
 E(\bar{X}) &= E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) \\
 &= \frac{1}{n} \sum_{i=1}^n \{E(X_i)\} \quad (\text{定理 2.11}) \\
 &= \frac{1}{n} \cdot n\mu = \mu
 \end{aligned}$$

$X_1, X_2, \dots, X_n, \dots$ が独立であるから

$$\begin{aligned}
 \sigma^2(\bar{X}) &= \sigma^2\left(\frac{1}{n} \sum_{i=1}^n X_i\right) \\
 &= \frac{1}{n^2} \sigma^2\left(\sum_{i=1}^n X_i\right) \quad (\text{定理 2.14}) \\
 &= \frac{1}{n^2} \sum_{i=1}^n \sigma^2(X_i) \quad (\text{定理 2.13}) \\
 &\leq \frac{1}{n^2} \cdot nk = \frac{k}{n}
 \end{aligned}$$

ゆえに、確率変数 $\bar{X}$ についてチェビシェフの不等式を用いれば、

$$\begin{aligned}
 P\{|\bar{X} - \mu| \leq \lambda \sigma(\bar{X})\} &> 1 - \frac{1}{\lambda^2} \\
 \therefore P\left\{|\bar{X} - \mu| \leq \lambda \frac{\sqrt{k}}{\sqrt{n}}\right\} &> 1 - \frac{1}{\lambda^2}
 \end{aligned}$$

ここで $\lambda \rightarrow \infty$ のとき $1 - (1/\lambda^2) \rightarrow 1$ となる。他方 λ がどんなに大きな定数であっても、それに比べて n をもっと大きくとれば、いかほどでも

$$\lambda \cdot \frac{\sqrt{k}}{\sqrt{n}} \rightarrow 0$$

とすることができる。したがって、どんなに小さな正数 ε をとっても、 n を大きくしさえすれば、

$$P(|\bar{X} - \mu| \leq \varepsilon)$$

を1に近づけることができる。これを大数の法則といい、経験的確率はこの性質を支えとしている。

たとえば、1個のサイコロを n 回投げるとき、 i 回目に関数 X_i が1の出れば $X_i = 1$ 、1以外の目が出れば $X_i = 0$ である確率変数 X_i ($i=1, 2, \dots, n, \dots$) をと

れば

$$P(X_i=1) = \frac{1}{6}, \quad P(X_i=0) = \frac{5}{6} \quad (i=1, 2, \dots, n, \dots)$$

であるから

$$E(X_i) = \frac{1}{6}, \quad \sigma^2(X_i) = \frac{5}{36} \quad (i=1, 2, \dots, n, \dots)$$

である。ここで、

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

とおけば、確率変数 $\bar{X}$ は、サイコロを n 回投げたとき 1 の目が出る比率を表しているから、大数の法則により、 n を大きくすれば、ほぼ 100% の確からしきで、1 の目が出る比率を $1/6$ に近づけることができる。

したがって、経験的確率は数学的確率を包含するとした 2.1.2 節の説明も、十分に理解されよう。

(例 2) 1 つのサイコロを 900 回投げたとき、1 の目が出る回数は確率 80% で何回ぐらいと考えられるか。

(解) 900 回投げるとき、 i 回めに 1 の目が出るとき $X_i=1$ 、1 以外の目が出るとき $X_i=0$ である確率変数を $X_i (i=1, 2, 3, \dots, 900)$ とすると

$$E(X_i) = \frac{1}{6}, \quad \sigma^2(X_i) = \frac{5}{36}$$

$$\bar{X} = \frac{1}{900} \sum_{i=1}^{900} X_i \quad \text{とおけば}$$

$$E(\bar{X}) = \frac{1}{6}, \quad \sigma^2(\bar{X}) = \frac{5}{36 \cdot 900}$$

であるから、チェビシェフの不等式により

$$P\left(\left|\bar{X} - \frac{1}{6}\right| \leq \lambda \cdot \frac{\sqrt{5}}{180}\right) > 1 - \frac{1}{\lambda^2}$$

題意より $1 - \frac{1}{\lambda^2} = 0.8$ とおき、これを解いて $\lambda = \sqrt{5}$

$$\therefore P\left(\left|\bar{X} - \frac{1}{6}\right| \leq \frac{5}{180}\right) > 0.8$$

すなわち

$$P\left(\frac{25}{180} \leq \bar{X} \leq \frac{35}{180}\right) > 0.8$$

$$\therefore P\left(\frac{25}{180} \times 900 \leq \bar{X} \times 900 \leq \frac{35}{180} \times 900\right) > 0.8$$

$$\therefore P(125 \leq 900\bar{X} \leq 175) > 0.8$$

ゆえに 80% 以上の確率で 125 回~175 回の間と考えられる(このような問題を推定の問題といい、第5章でくわしく調べる)。

§2.3 モーメント母関数

確率変数 X に対し $e^{\theta X}$ の平均 $E(e^{\theta X})$ を θ の関数と考えて $M(\theta)$ と書き、 X のモーメント母関数という。このとき

$$M(0) = 1, M'(0) = E(X), M''(0) = E(X^2)$$

$$\dots\dots\dots, M^{(n)}(0) = E(X^n)$$

なる関係が成り立つ。これを証明しよう。

X が離散的確率変数で、 $X = x_1, x_2, \dots, x_m$ をとり

$$P(X = x_i) = p_i \quad (i = 1, 2, \dots, m)$$

とすると、

$$M(0) = \sum_{i=1}^m p_i = 1$$

$$M'(0) = \left(\frac{d}{d\theta} \sum_{i=1}^m e^{\theta x_i} p_i \right)_{\theta=0} = \left(\sum_{i=1}^m x_i e^{\theta x_i} p_i \right)_{\theta=0} = \sum_{i=1}^m x_i p_i = E(X)$$

$$M''(0) = \left(\frac{d^2}{d\theta^2} \sum_{i=1}^m e^{\theta x_i} p_i \right)_{\theta=0} = \left(\sum_{i=1}^m x_i^2 e^{\theta x_i} p_i \right)_{\theta=0} = \sum_{i=1}^m x_i^2 p_i = E(X^2)$$

.....
.....

$$M^{(n)}(0) = \left(\frac{d^n}{d\theta^n} \sum_{i=1}^m e^{\theta x_i} p_i \right)_{\theta=0} = \left(\sum_{i=1}^m x_i^n e^{\theta x_i} p_i \right)_{\theta=0} = \sum_{i=1}^m x_i^n p_i = E(X^n)$$

X が連続的確率変数のとき、 $p(x)$ を X の確率密度関数とすると

$$M(0) = \int_{-\infty}^{\infty} p(x) dx = 1$$

$$M'(0) = \left(\frac{d}{d\theta} \int_{-\infty}^{\infty} e^{\theta x} p(x) dx \right)_{\theta=0} = \left(\int_{-\infty}^{\infty} x e^{\theta x} p(x) dx \right)_{\theta=0}$$

$$= \int_{-\infty}^{\infty} x p(x) dx = E(X)$$

$$\begin{aligned}
M''(0) &= \left(\frac{d^2}{d\theta^2} \int_{-\infty}^{\infty} e^{\theta x} p(x) dx \right)_{\theta=0} = \left(\int_{-\infty}^{\infty} x^2 e^{\theta x} p(x) dx \right)_{\theta=0} \\
&= \int_{-\infty}^{\infty} x^2 p(x) dx = E(X^2) \\
&\dots\dots\dots \\
&\dots\dots\dots \\
M^{(n)}(0) &= \left(\frac{d^n}{d\theta^n} \int_{-\infty}^{\infty} e^{\theta x} p(x) dx \right)_{\theta=0} = \left(\int_{-\infty}^{\infty} x^n e^{\theta x} p(x) dx \right)_{\theta=0} \\
&= \int_{-\infty}^{\infty} x^n p(x) dx = E(X^n)
\end{aligned}$$

$M^{(n)}(0)$ を X の原点のまわりの n 次のモーメントという。

X の平均 $E(X)$, 分散 $\sigma^2(X)$ はモーメント母関数から求めることができる。上記の性質を利用して

$$\begin{aligned}
E(X) &= M'(0) \\
\sigma^2(X) &= E(X^2) - \{E(X)\}^2 = M''(0) - \{M'(0)\}^2
\end{aligned}$$

が成り立つ。

いま定義したモーメント母関数の性質は、今後確率変数 X の分布を調べるときに重要な役割をはたす。

モーメント母関数の基本的な性質について調べよう。

〔定理 2.16〕 確率変数の分布関数は、そのモーメント母関数から唯一通りに定められる。

〔定理 2.17〕 確率変数 $X_1, X_2, \dots, X_n, \dots, X$ の分布関数およびモーメント母関数をそれぞれ

$$\begin{aligned}
&F_1(x), F_2(x), \dots, F_n(x), \dots, F(x) \\
&M_1(\theta), M_2(\theta), \dots, M_n(\theta), \dots, M(\theta)
\end{aligned}$$

とおくとき $\lim_{i \rightarrow \infty} M_i(\theta) = M(\theta)$ とすれば、 $M(\theta)$ により定まる分布関数 $F(x)$ は

$$\lim_{i \rightarrow \infty} F_i(x) = F(x)$$

によって与えられる。

定理 2.16, 定理 2.17 の証明は本書の程度を超えるので省略する。

[定理 2.18] 2つの確率変数 X, Y が互いに独立のとき, $X, Y, X+Y$ のモーメント母関数をそれぞれ $M_x(\theta), M_y(\theta), M(\theta)$ とすれば

$$M(\theta) = M_x(\theta) \cdot M_y(\theta)$$

が成り立つ.

(証明) (1) X, Y が離散的確率変数のとき:

$$X = x_1, x_2, \dots, x_m, \quad P(X=x_i) = f_i$$

$$Y = y_1, y_2, \dots, y_n, \quad P(Y=y_j) = g_j$$

とし, X, Y の2次元確率密度を p_{ij} とすれば, X, Y が互いに独立であることから

$$p_{ij} = f_i \cdot g_j$$

したがって

$$\begin{aligned} M(\theta) &= E(e^{\theta(X+Y)}) = \sum_{i=1}^m \sum_{j=1}^n e^{\theta(x_i+y_j)} p_{ij} \\ &= \sum_{i=1}^m \sum_{j=1}^n e^{\theta(x_i+y_j)} f_i g_j = \sum_{i=1}^m e^{\theta x_i} f_i \cdot \sum_{j=1}^n e^{\theta y_j} g_j \\ &= M_x(\theta) M_y(\theta) \end{aligned}$$

(2) X, Y が連続的確率変数のとき:

X, Y の確率密度関数をそれぞれ $f(x), g(y)$ とすれば, X, Y の2次元確率密度関数 $p(x, y)$ は, X, Y が互いに独立であることから,

$$p(x, y) = f(x) \cdot g(y)$$

したがって

$$\begin{aligned} M(\theta) &= E(e^{\theta(X+Y)}) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{\theta(x+y)} p(x, y) dx dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{\theta(x+y)} f(x) \cdot g(y) dx dy \\ &= \int_{-\infty}^{\infty} e^{\theta x} f(x) dx \int_{-\infty}^{\infty} e^{\theta y} g(y) dy \\ M(\theta) &= M_x(\theta) \cdot M_y(\theta) \end{aligned}$$

[定理 2.19] $X, \alpha X + \beta$ のモーメント母関数をそれぞれ $M_x(\theta), M(\theta)$ とおけば

$$M(\theta) = e^{\beta\theta} M_x(\alpha\theta)$$

(証明) (1) X が離散的確率変数のとき:

$$X = x_1, x_2, \dots, x_m, \quad P(X=x_i) = p_i \quad (i=1, 2, \dots, m)$$

とすれば

$$M(\theta) = \sum_{i=1}^m e^{\theta(a x_i + \beta)} p_i = e^{\beta\theta} \sum_{i=1}^m e^{(\alpha\theta) x_i} \cdot p_i = e^{\beta\theta} M_x(\alpha\theta)$$

(2) X が連続的確率変数のとき:

X の確率密度関数を $p(x)$ とすれば

$$M(\theta) = \int_{-\infty}^{\infty} e^{\theta(ax+\beta)} p(x) dx = e^{\beta\theta} \int_{-\infty}^{\infty} e^{(\alpha\theta)x} p(x) dx = e^{\beta\theta} M_x(\alpha\theta)$$

問1 3枚の貨幣を投げて, 表の出る枚数を X 枚とするとき, X のモーメント母関数を求めよ. またその結果を用いて平均, 分散を求めよ.

問2 連続的確率変数 X の確率密度関数 $p(x)$ が

$$p(x) = \begin{cases} 0 & (x \leq 0) \\ ae^{-ax} & (x > 0) \end{cases} \quad (a > 0)$$

で与えられるとき, モーメント母関数を求めよ. また, これを用いて確率変数 X の平均, 分散を求めよ.

第3章 基本分布関数

確率変数はその確率密度関数の型によって種々の名前が付されて分類されている。

この章では、最も基本となる二項分布、ポアソン分布、正規分布を取りあげ、これらの分布に従う確率変数についてその性質を調べ、具体例を取りあげる。

最後に取り扱う中心極限定理は、統計学において最も重要な基本定理であり、この定理によって、正規分布が重要な意味をもつことを知るであろう。

§3.1 二項分布

1つのサイコロを n 回投げて、1の目の出る回数を X 回とおけば

$$P(X=k) = {}_n C_k \left(\frac{1}{6}\right)^k \left(\frac{5}{6}\right)^{n-k}$$

である。一般に離散的確率変数 X があって、 X のとりうる値が $X=0, 1, 2, \dots, n$ で

$$P(X=k) = {}_n C_k p^k (1-p)^{n-k} \quad (k=0, 1, 2, \dots, n) \quad (3.1.1)$$

が成り立つとき、確率変数 X は二項分布に従うという。 $1-p=q$ とおけば、(3.1.1)は

$$P(X=k) = {}_n C_k p^k q^{n-k} \quad (k=0, 1, 2, \dots, n) \quad (3.1.2)$$

となる。二項定理によれば

$$(p+q)^n = \sum_{k=0}^n {}_n C_k p^k q^{n-k}$$

である。二項分布の名称はこれに由来する。

二項分布のモーメント母関数を求めると、

$$\begin{aligned} M(\theta) &= E(e^{\theta X}) = \sum_{k=0}^n e^{\theta k} {}_n C_k p^k q^{n-k} = \sum_{k=0}^n {}_n C_k (pe^{\theta})^k q^{n-k} \\ &= (pe^{\theta} + q)^n \end{aligned}$$

X の平均, 分散を求めるために $M'(\theta)$, $M''(\theta)$ を計算すると

$$M'(\theta) = npe^\theta(pe^\theta + q)^{n-1}$$

$$M''(\theta) = npe^\theta(pe^\theta + q)^{n-1} + n(n-1)p^2p^{2\theta}(pe^\theta + q)^{n-2}$$

ゆえに

$$M'(0) = np(p+q)^{n-1} = np$$

$$M''(0) = np(p+q)^{n-1} + n(n-1)p^2(p+q)^{n-2} = np + n(n-1)p^2$$

これらより X の平均 $E(X)$, 分散 $\sigma^2(X)$ が求まる.

$$E(X) = M'(0) = np$$

$$\sigma^2(X) = M''(0) - \{M'(0)\}^2 = np + n(n-1)p^2 - n^2p^2 = np(1-p) = npq$$

また, 後で,

$$P(X \geq k) = \sum_{x=k}^n {}_nC_x p^x q^{n-x} \quad \text{および} \quad P(X \leq k) = \sum_{x=0}^k {}_nC_x p^x q^{n-x}$$

の値を知る必要が生じる. この計算結果を定理として与えておこう. 後で示すように, この定理は二項分布と F -分布の関係を与える.

【定理 3.1】 確率変数 X が二項分布に従うとき, すなわち

$$P(X=k) = {}_nC_k p^k q^{n-k}$$

のとき,

$$P(X \geq k) = \frac{n!}{(k-1)!(n-k)!} \int_0^p t^{k-1}(1-t)^{n-k} dt$$

(証明) 部分積分の公式をくり返して用いて

$$\begin{aligned} \int_0^p t^{k-1}(1-t)^{n-k} dt &= \frac{p^k}{k}(1-p)^{n-k} + \frac{n-k}{k} \int_0^p t^k(1-t)^{n-k-1} dt \\ &= \frac{1}{k} p^k (1-p)^{n-k} + \frac{n-k}{k(k+1)} p^{k+1} (1-p)^{n-k-1} \\ &\quad + \frac{(n-k)(n-k-1)}{k(k+1)(k+2)} p^{k+2} (1-p)^{n-k-2} \\ &\quad + \cdots \\ &\quad + \frac{(n-k)(n-k-1) \cdots 2 \cdot 1}{k(k+1) \cdots (n-2)(n-1)} \int_0^p t^{n-1} dt \end{aligned}$$

$$= \frac{1}{k} p^k (1-p)^{n-k} + \frac{n-k}{k(k+1)} p^{k+1} (1-p)^{n-k-1} + \frac{(n-k)(n-k-1)}{k(k+1)(k+2)} p^{k+2} (1-p)^{n-k-2} \\ + \cdots + \frac{(n-k)(n-k-1) \cdots 2 \cdot 1}{k(k+1) \cdots (n-2)(n-1)n} p^n$$

両辺に $\frac{n!}{(k-1)!(n-k)!}$ をかけて

$$\frac{n!}{(k-1)!(n-k)!} \int_0^p t^{k-1} (1-t)^{n-k} dt \\ = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} + \frac{n!}{(k+1)!(n-k-1)!} p^{k+1} (1-p)^{n-k-1} \\ + \frac{n!}{(k+2)!(n-k-2)!} p^{k+2} (1-p)^{n-k-2} + \cdots + \frac{n!}{n!0!} p^n$$

したがって

$$\frac{n!}{(k-1)!(n-k)!} \int_0^p t^{k-1} (1-t)^{n-k} dt = \sum_{x=k}^n {}_n C_x p^x (1-p)^{n-x} = P(X \geq k)$$

二項分布の重要な性質として、

- (1) 二項分布で p が非常に小さく、 n が非常に大きくて np がある定まった値と考えられるとき、近似的にポアソン分布に従う。
- (2) 二項分布で n が大きいとき、近似的に

$$\text{平均} = np \quad \text{分散} = npq$$

の正規分布に従う。

がある。(1)についてはつぎの3.2のポアソン分布で、(2)については§3.5の中心極限定理の中で説明される。

(例1) 100桁の乱数を発生させたとき、数字0があらわれる回数を X とするとき、 X の平均と分散を求めよ。また X のモーメント母関数を求めよ。

(解) 二項分布において

$$n = 100, \quad p = \frac{1}{10}, \quad q = \frac{9}{10}$$

であるから

$$E(X) = np = 10, \quad \sigma^2(X) = npq = 9$$

$$M(\theta) = (pe^\theta + q)^n = \left(\frac{1}{10} e^\theta + \frac{9}{10} \right)^{100}$$

§3.2 ポアソン分布

離散的確率変数 X のとりうる値が $X=0, 1, 2, \dots, n, \dots$ で

$$P(X=k) = \frac{m^k}{k!} e^{-m} \quad (m>0)$$

のとき, X はポアソン分布に従うという. このとき, 確率密度の性質から

$$P(X \geq 0) = \sum_{k=0}^{\infty} \frac{m^k}{k!} e^{-m} = 1$$

でなければならないが, 実際に

$$\begin{aligned} \sum_{k=0}^{\infty} \frac{m^k}{k!} e^{-m} &= e^{-m} \sum_{k=0}^{\infty} \frac{m^k}{k!} \\ &= e^{-m} \cdot e^m \quad (\text{テイラー定理による}) \\ &= 1 \end{aligned}$$

である.

ポアソン分布のモーメント母関数 $M(\theta)$, 平均 $E(X)$, 分散 $\sigma^2(X)$ を求めよう.

$$\begin{aligned} M(\theta) &= E(e^{\theta X}) = \sum_{k=0}^{\infty} e^{\theta k} \cdot \frac{m^k}{k!} e^{-m} = e^{-m} \sum_{k=0}^{\infty} \frac{(e^{\theta} m)^k}{k!} \\ &= e^{-m} \cdot e^{m e^{\theta}} = e^{m(e^{\theta}-1)} \end{aligned}$$

$$\therefore M'(\theta) = e^{m(e^{\theta}-1)} \cdot m e^{\theta}$$

$$M''(\theta) = e^{m(e^{\theta}-1)} \cdot (m e^{\theta})^2 + e^{m(e^{\theta}-1)} \cdot m e^{\theta} = m e^{\theta} \cdot e^{m(e^{\theta}-1)} (m e^{\theta} + 1)$$

$$\therefore M'(0) = m, \quad M''(0) = m(m+1)$$

$$\therefore E(X) = M'(0) = m$$

$$\sigma^2(X) = M''(0) - \{M'(0)\}^2 = m(m+1) - m^2 = m$$

${}_n C_k p^k (1-p)^{n-k}$ において, $np=m$ において m を固定し, n を限りなく大きくすると

$$\begin{aligned} \lim_{n \rightarrow \infty} {}_n C_k p^k (1-p)^{n-k} &= \lim_{n \rightarrow \infty} {}_n C_k \left(\frac{m}{n}\right)^k \left(1 - \frac{m}{n}\right)^{n-k} \\ &= \lim_{n \rightarrow \infty} \frac{n(n-1)(n-2) \cdots (n-k+1)}{k!} \left(\frac{m}{n}\right)^k \left(1 - \frac{m}{n}\right)^{n-k} \\ &= \lim_{n \rightarrow \infty} \frac{m^k}{k!} \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right) \left(1 - \frac{m}{n}\right)^n \left(1 - \frac{m}{n}\right)^{-k} \end{aligned}$$

ここで, $n \rightarrow \infty$ のとき

$$\begin{aligned} \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right) &\rightarrow 1, \quad \left(1 - \frac{m}{n}\right)^{-k} \rightarrow 1 \\ \left(1 - \frac{m}{n}\right)^n &= \left\{ \left(1 + \frac{1}{(-n/m)}\right)^{(-n/m)} \right\}^{-m} \rightarrow e^{-m} \\ (\text{なぜなら } x \rightarrow \pm\infty \text{ のとき } \left(1 + \frac{1}{x}\right)^x &\rightarrow e) \end{aligned}$$

したがって

$$\lim_{n \rightarrow \infty} {}_n C_k p^k (1-p)^{n-k} = \frac{m^k}{k!} e^{-m}$$

ゆえに n が大きく, p が小さく, np がある定まった値と考えられるとき, 二項分布は

$$m = np$$

表 3-1

k	${}_n C_k p^k (1-p)^{n-k}$ ($n=20, p=0.1$)	$\frac{m^k}{k!} e^{-m}$ ($m=2$)	k	${}_n C_k p^k (1-p)^{n-k}$ ($n=30, p=0.1$)	$\frac{m^k}{k!} e^{-m}$ ($m=3$)
0	0.1216	0.1353	0	0.0424	0.0498
1	0.2702	0.2707	1	0.1413	0.1494
2	0.2802	0.2707	2	0.2277	0.2240
3	0.1901	0.1804	3	0.2361	0.2240
4	0.0898	0.0902	4	0.1771	0.1680
5	0.0319	0.0361	5	0.1023	0.1008
6	0.0089	0.0120	6	0.0474	0.0504
7	0.0020	0.0034	7	0.0180	0.0016
8	0.0004	0.0009	8	0.0058	0.0081
9	0.0001	0.0002	9	0.0016	0.0027
10	0.0000	0.0000	10	0.0004	0.0008
11			11	0.0001	0.0002
12			12	0.0000	0.0001

とおいたときのポアソン分布に近似していることがわかる。二項分布がポアソン分布で近似される様子は表 3-1 でも理解できるであろう。

(例 1) ある本が 10 ページにつき 1 字の割合で誤植があるとき、20 ページからなる章で X 字($X=0, 1, 2, 3, 4, 5$)の誤植のある確率を求めよ。

(解) 1 ページは a 字詰めとすると、10 ページにつき 1 字の割合で誤植があるということは、 $10a$ 字につき 1 字の割合で誤植があることである。すなわち 1 字活字をひろって、それが誤植である確率は $1/10a$ である。ゆえに 20 ページで X 字誤植する確率は

$${}^{20a}C_X \left(\frac{1}{10a} \right)^X \left(1 - \frac{1}{10a} \right)^{20a-X} \doteq \frac{2^X}{X!} e^{-2}, \left(n=20a, p=\frac{1}{10a}, np=2 \right)$$

である。したがって、その値はポアソン分布表より下の表が得られる。

X	0	1	2	3	4	5
確率	0.1353	0.2707	0.2707	0.1804	0.0902	0.0361

(例 2) 3 個のサイコロを 216 回投げて、3 個とも 1 の目が出る回数を X とするとき

$$X = 0, 1, 2, 3, 4, 5$$

のときの確率をそれぞれ求めよ。

(解) 3 個のサイコロを 1 回投げて、3 個とも 1 の目が出る確率は

$$\left(\frac{1}{6} \right)^3 = \frac{1}{216}$$

であるから、216 回投げて 3 個とも 1 の目が出る回数が X 回である確率は

$${}^{216}C_X \left(\frac{1}{216} \right)^X \left(1 - \frac{1}{216} \right)^{216-X} \doteq \frac{1^X}{X!} e^{-1}$$

である。したがって、ポアソン分布表より下の表が得られる。

X	0	1	2	3	4	5	6	7
確率	0.3679	0.3679	0.1839	0.0613	0.0153	0.0031	0.0005	0.0001

問 1 確率変数 X がポアソン分布に従うとき、平均 $E(X)$ 、分散 $\sigma^2(X)$ を定義式より直接求めよ。

問 2 不良率 0.2% の部品の山から、任意に 100 個取り出したとき、その中に不良品が混じる確率を求めよ。

§3.3 正 規 分 布

連続的確率変数 X の確率密度関数 $p(x)$ が

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2} \quad (3.3.1)$$

で与えられるとき, X は $N(\mu, \sigma^2)$ 型の正規分布に従うという. $p(x)$ のグラフは直線 $x=\mu$ に関して対称で, 図 3-1 のような形となる. この曲線を正規分布曲線という.

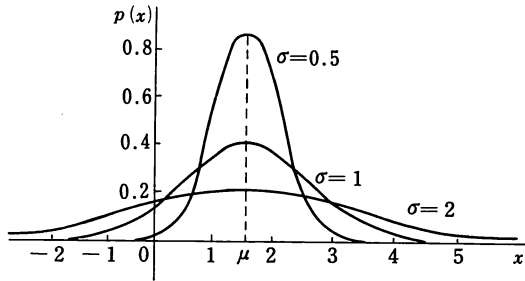


図 3-1

統計学において重要な役割をもつ正規分布の性質について調べよう.

まず, $p(x)$ が確率密度関数であるための必要十分条件を吟味する.

$$\begin{aligned} \int_{-\infty}^{\infty} p(x) dx &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-t^2/2} dt \quad \left(\frac{x-\mu}{\sigma} = t \text{ とおく} \right) \\ &= \frac{1}{\sqrt{2\pi}} \cdot \sqrt{2\pi} = 1 \end{aligned}$$

したがって, (3.3.1) で与えられる $p(x)$ は確率密度関数であってよい.

特に $\mu=0, \sigma=1$ のときの正規分布 $N(0, 1)$ を標準正規分布といい, このとき

$$p(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

である.

X が $N(0, 1)$ 型に従う確率変数であるとし,

$$Y = \sigma X + \mu$$

とおけば Y は $N(\mu, \sigma^2)$ 型の正規分布に従う. なぜなら, 確率変数 X, Y の確率密度関数をそれぞれ $p_1(x), p_2(y)$ とすれば, 定理 2.10 より

$$p_2(y) = \frac{1}{\sigma} p_1\left(\frac{y-\mu}{\sigma}\right) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(y-\mu)^2/2\sigma^2} \quad \left(p_1(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\right)$$

となるからである。逆に Y が $N(\mu, \sigma^2)$ 型の正規分布に従う確率変数であるとき

$$X = \frac{Y-\mu}{\sigma}$$

とおけば X は $N(0, 1)$ 型の正規分布に従う。

X が $N(0, 1)$ 型の正規分布に従うとき、 X のモーメント母関数 $M_x(\theta)$ は

$$\begin{aligned} M_x(\theta) &= E(e^{\theta X}) = \int_{-\infty}^{\infty} e^{\theta x} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{(\theta x - x^2/2)} dx = \frac{1}{\sqrt{2\pi}} e^{\theta^2/2} \int_{-\infty}^{\infty} e^{-(x-\theta)^2/2} dx \\ &= \frac{1}{\sqrt{2\pi}} e^{\theta^2/2} \int_{-\infty}^{\infty} e^{-t^2/2} dt \quad (x-\theta=t \text{ とおく}) \\ &= e^{\theta^2/2} \end{aligned}$$

である。ここで $Y = \sigma X + \mu$ とおけば、 Y は $N(\mu, \sigma^2)$ 型の正規分布に従うから、 Y のモーメント母関数 $M(\theta)$ は、定理 2.19 より

$$M(\theta) = e^{\mu\theta} M_x(\sigma\theta) = e^{\mu\theta} \cdot e^{\sigma^2\theta^2/2} = e^{\mu\theta + (\sigma^2\theta^2)/2}$$

となる。したがって

$$M'(\theta) = (\mu + \sigma^2\theta) e^{\mu\theta + (\sigma^2\theta^2)/2}, \quad M''(\theta) = \{\sigma^2 + (\mu + \sigma^2\theta)^2\} e^{\mu\theta + (\sigma^2\theta^2)/2}$$

これらより、平均 $E(Y)$ 、分散 $\sigma^2(Y)$ を求めると、

$$E(Y) = M'(0) = \mu$$

$$\sigma^2(Y) = M''(0) - \{M'(0)\}^2 = (\sigma^2 + \mu^2) - \mu^2 = \sigma^2$$

これらの結果は重要である。

$N(0, 1)$ 型の正規分布の分布関数はよく $\Phi(x)$ で表す。

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$$

の x の種々の正の値に対する計算値が巻末に表として与えてある。これを正規分布表という。 $N(0, 1)$ 型の確率変数の確率密度関数のグラフは $x=0$ に関して対称であるから、 x の負の値に対する分布関数の値は

$$\Phi(x) = 1 - \Phi(-x)$$

として求められる.

$N(0, 1)$ 型の確率変数 X と $N(\mu, \sigma^2)$ 型の確率変数 Y の分布関数をそれぞれ $\Phi(x), F(y)$ とすると, $Y = \sigma X + \mu$ であるから定理 2.9 より

$$F(y) = \Phi\left(\frac{y - \mu}{\sigma}\right)$$

これよりいかなる型の正規分布であっても, その分布関数の値を正規分布表から求めることができる.

(例 1) 確率変数 X が $N(23, 4^2)$ 型のとき, それぞれ次の確率を求めよ.

$$(1) X \leq 30, \quad (2) X > 22, \quad (3) |X - 23| < 3$$

(解) $N(0, 1)$ と $N(23, 4^2)$ の分布関数をそれぞれ $\Phi(x), F(x)$ とする.

$$(1) P(X \leq 30) = F(30) = \Phi\left(\frac{30 - 23}{4}\right) = \Phi(1.75) = 0.9599$$

$$\begin{aligned} (2) P(X > 22) &= 1 - P(X \leq 22) = 1 - F(22) \\ &= 1 - \Phi\left(\frac{22 - 23}{4}\right) = 1 - \Phi(-0.25) \\ &= 1 - \{1 - \Phi(0.25)\} = \Phi(0.25) = 0.5987 \end{aligned}$$

$$\begin{aligned} (3) P(|X - 23| < 3) &= P(20 < X < 26) = F(26) - F(20) \\ &= \Phi(0.75) - \Phi(-0.75) \\ &= 2\{\Phi(0.75) - \Phi(0)\} \\ &\quad (N(0, 1) \text{ の確率密度関数は } x=0 \text{ に関して対称}) \\ &= 2(0.7734 - 0.5) = 0.5468 \end{aligned}$$

[定理 3.2] 2つの確率変数 X, Y が互いに独立で X が $N(\mu_1, \sigma_1^2)$ 型, Y が $N(\mu_2, \sigma_2^2)$ 型に従うとき, $X + Y$ は $N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$ 型に従う. この性質を正規分布の再生性という.

(証明) X, Y のモーメント母関数をそれぞれ $M_1(\theta), M_2(\theta)$ とすれば

$$M_1(\theta) = e^{\mu_1\theta + (\sigma_1^2\theta^2)/2}$$

$$M_2(\theta) = e^{\mu_2\theta + (\sigma_2^2\theta^2)/2}$$

であり, X, Y が独立であることから定理 2.18 により, $X + Y$ のモーメント母関数 $M(\theta)$ は

$$M(\theta) = M_1(\theta) \cdot M_2(\theta) = e^{(\mu_1 + \mu_2)\theta + [(\sigma_1^2 + \sigma_2^2)\theta^2]/2}$$

である。これは確率変数 $X+Y$ が正規分布 $N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$ 型であることを示している。

定理 3.2 を一般化して定理 3.3 を得る。

〔定理 3.3〕 n 個の確率変数 $X_1, X_2, \dots, X_n$ が互いに独立で、 X_i が $N(\mu_i, \sigma_i^2)$ 型 ($i=1, 2, \dots, n$) に従うならば、

$$X = X_1 + X_2 + \dots + X_n$$

は $N\left(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2\right)$ 型に従う。

(証明) 定理 3.2 をくり返し使うことにより証明される。

〔定理 3.4〕 n 個の確率変数 $X_1, X_2, \dots, X_n$ が互いに独立で、すべて $N(\mu, \sigma^2)$ 型に従うならば

$$\bar{X} = \frac{1}{n}(X_1 + X_2 + \dots + X_n)$$

は $N\left(\mu, \frac{\sigma^2}{n}\right)$ 型に従う。

(証明) X_i ($i=1, 2, \dots, n$) が $N(\mu, \sigma^2)$ 型に従うから、 $\frac{1}{n}X_i$ は $N\left(\frac{\mu}{n}, \frac{\sigma^2}{n^2}\right)$ 型に従う。したがって、定理 3.3 により

$$\bar{X} = \frac{1}{n}(X_1 + X_2 + \dots + X_n) = \frac{1}{n}X_1 + \frac{1}{n}X_2 + \dots + \frac{1}{n}X_n$$

は $N\left(\mu, \frac{\sigma^2}{n}\right)$ 型に従う。

(例 2) 確率変数 X, Y が独立で、それぞれ $N(\mu_1, \sigma_1^2)$ 型、 $N(\mu_2, \sigma_2^2)$ 型に従うならば、 $\alpha X + \beta Y$ は $N(\alpha\mu_1 + \beta\mu_2, \alpha^2\sigma_1^2 + \beta^2\sigma_2^2)$ 型に従う。

(解) X が $N(\mu_1, \sigma_1^2)$ 型に従うから、 αX は $N(\alpha\mu_1, \alpha^2\sigma_1^2)$ 型に従う。同様に βY は $N(\beta\mu_2, \beta^2\sigma_2^2)$ 型に従う。したがって、定理 3.2 より $\alpha X + \beta Y$ は $N(\alpha\mu_1 + \beta\mu_2, \alpha^2\sigma_1^2 + \beta^2\sigma_2^2)$ 型に従う。

(例 3) ある学年の男子生徒の身長を X cm, 女子生徒の身長を Y cm とし、 X, Y がそれぞれ $N(138.6, 5.6^2)$ 型、 $N(135.4, 5.2^2)$ 型に従うとすると、男女の身長の差 $X - Y$ が 5 cm 以内である確率を求めよ。

(解) 例 2 において $\alpha=1, \beta=-1, \mu_1=138.6, \mu_2=135.4, \sigma_1^2=5.6^2, \sigma_2^2=5.2^2$ とお

けば $X - Y$ は $N(3.2, 7.6^2)$ 型に従う。ゆえに

$$\begin{aligned} P(-5 < X - Y < 5) &= P(X - Y < 5) - P(X - Y < -5) \\ &= \Phi\left(\frac{5-3.2}{7.6}\right) - \Phi\left(\frac{-5-3.2}{7.6}\right) = \Phi(0.24) - \Phi(-1.08) \\ &= \Phi(0.24) + \Phi(1.08) - 1 = 0.4548 \end{aligned}$$

問1 ある大学の入学試験の点数は平均 300 点，標準偏差 50 点の正規分布に従う。

(1) 450 点以上の受験生の割合，(2) 200 点以下の受験生の割合
をそれぞれ求めよ。

問2 ある学校で身体検査を行った。体重は平均 57 kg，標準偏差 4 kg の正規分布をしていると仮定して，体重 65 kg 以上の学生は何%いると考えてよいか。また平均と 1 kg 以上相違しない学生は何%いると考えてよいか。

§3.4 母集団と標本

観測・調査・実験などの対象から得られるすべての測定値(データ)の集まりを母集団という。全体のデータの数 N とするとき $N \rightarrow \infty$ ならば，その母集団を有限母集団といい， N をその母集団の大きさという。 $N \rightarrow \infty$ と考えられるとき，その母集団を無限母集団という。母集団の平均，分散をそれぞれ母平均，母分散という。

母集団から任意に 1 つの要素をとり出すとき，その 1 つの要素を標本という。無作為(ランダム)に標本を得ることを無作為抽出という。

いま母集団が与えられたとき，その母集団を定める確率変数を X とし，その母集団から n 個の標本を無作為抽出することを考える。個々の標本を定める確率変数を $X_i (i=1, 2, \dots, n)$ とすると， n 個の確率変数の組

$$(X_1, X_2, \dots, X_n)$$

が得られ，個々の標本の分布は母集団の分布と同様の法則に従うであろう。

無作為抽出された n 個の標本を得る目的は，これらを用いて，母平均や母分散などの母集団の特性値の推定を行うことにある。このためには統計量と呼ばれる確率変数の分布を調べておくことと便利である。たとえば $\bar{X}, S^2, S$ を

$$\bar{X} = \frac{1}{n}(X_1 + X_2 + \dots + X_n)$$

$$S^2 = \frac{1}{n-1} \{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \cdots + (X_n - \bar{X})^2\}$$

$$S = \sqrt{S^2}$$

で定義し、それぞれ標本平均、標本分散、標本標準偏差と呼び、これらを統計量と総称する。

いま n 個の標本を実際に得て、これを

$$(x_1, x_2, \dots, x_n)$$

とすれば、すなわち $X_i = x_i (i=1, 2, \dots, n)$ とすれば、 x_i を X_i の実現値という。

$$\bar{x} = \frac{1}{n}(x_1 + x_2 + \cdots + x_n)$$

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

とおけば、 $\bar{x}$ は $\bar{X}$ の実現値であり、 s^2 は S^2 の実現値である。なお n 個の標本の組

$$(X_1, X_2, \dots, X_n)$$

を、簡単に、大きさ n の標本と呼んでいる。

統計量の分布を順を追って調べてみよう。

母集団を定める確率変数 X が $N(\mu, \sigma^2)$ 型に従うとき、その母集団を $N(\mu, \sigma^2)$ 型の正規母集団という。

【定理 3.5】 正規母集団 $N(\mu, \sigma^2)$ からの、大きさ n の標本の標本平均 $\bar{X}$ の分布は $N\left(\mu, \frac{\sigma^2}{n}\right)$ 型に従う。

(証明) この母集団からの大きさ n の標本を

$$(X_1, X_2, \dots, X_n)$$

とすれば、 $X_i (i=1, 2, \dots, n)$ もまた $N(\mu, \sigma^2)$ 型に従う。

ゆえに定理 3.4 より

$$\bar{X} = \frac{1}{n}(X_1 + X_2 + \cdots + X_n)$$

は $N\left(\mu, \frac{\sigma^2}{n}\right)$ 型に従う。

なお次節で母集団が正規母集団でなくとも, n が大きければ $\bar{X}$ は近似的に正規分布に従うことが示される(中心極限定理)。

(例) 20歳の日本人男子の身長は戦前は, 平均 160 cm, 標準偏差 6 cm の正規分布に従っていた. この中から抽出した 100 人の平均が母平均と 0.5 cm 以上相違する確率を求めよ.

(解) 100 人の平均 $\bar{X}$ の分布は

$$N\left(160, \frac{6^2}{100}\right) = N\left(160, \left(\frac{6}{10}\right)^2\right)$$

であるから, 求める確率は

$$\begin{aligned} P(|\bar{X} - 160| > 0.5) &= 2P(\bar{X} > 160.5) = 2\{1 - P(\bar{X} < 160.5)\} \\ &= 2\left\{1 - \Phi\left(\frac{160.5 - 160}{0.6}\right)\right\} = 2\{1 - \Phi(0.833)\} \\ &= 2(1 - 0.7967) = 0.4066 \end{aligned}$$

[定理 3.6] 母平均 μ , 母分散 σ^2 の大きさ N の母集団からの大きさ n の標本の標本平均 $\bar{X}$ の平均 $E(\bar{X})$ および分散 $\sigma^2(\bar{X})$ は

$$E(\bar{X}) = \mu, \quad \sigma^2(\bar{X}) = \frac{N-n}{N-1} \cdot \frac{\sigma^2}{n}$$

である.

$$(\text{証明}) \quad E(\bar{X}) = E\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right)$$

$$= \frac{1}{n} \{E(X_1) + E(X_2) + \cdots + E(X_n)\}$$

$$= \frac{1}{n} \cdot n\mu = \mu$$

$$\sigma^2(\bar{X}) = E\{(\bar{X} - \mu)^2\} = E\left\{\frac{(X_1 + X_2 + \cdots + X_n - n\mu)^2}{n^2}\right\}$$

$$= \frac{1}{n^2} E\left[\{(X_1 - \mu) + (X_2 - \mu) + \cdots + (X_n - \mu)\}^2\right]$$

$$= \frac{1}{n^2} E\left\{\sum_{i=1}^n (X_i - \mu)^2 + \sum_{\substack{i,j=1 \\ i \neq j}}^n (X_i - \mu)(X_j - \mu)\right\}$$

$$= \frac{1}{n^2} \left[\sum_{i=1}^n E\{(X_i - \mu)^2\} + \sum_{\substack{i,j=1 \\ i \neq j}}^n E\{(X_i - \mu)(X_j - \mu)\} \right]$$

ここで

$$\begin{aligned}
 \sum_{i=1}^n E\{(X_i - \mu)^2\} &= n\sigma^2 \\
 \sum_{\substack{i, j=1 \\ i \neq j}}^n E\{(X_i - \mu)(X_j - \mu)\} \\
 &= n(n-1)E\{(X_k - \mu)(X_l - \mu)\} \quad (k \neq l) \\
 &= n(n-1) \sum_{\substack{i, j=1 \\ i \neq j}}^N (x_i - \mu)(x_j - \mu)P(X_k = x_i, X_l = x_j) \\
 &= n(n-1) \frac{1}{N(N-1)} \sum_{\substack{i, j=1 \\ i \neq j}}^N (x_i - \mu)(x_j - \mu) \\
 &= \frac{n(n-1)}{N(N-1)} \sum_{i=1}^N \left\{ (x_i - \mu) \sum_{\substack{j=1 \\ i \neq j}}^N (x_j - \mu) \right\} \\
 &= \frac{n(n-1)}{N(N-1)} \sum_{i=1}^N \left[(x_i - \mu) \left\{ \sum_{j=1}^N (x_j - \mu) - (x_i - \mu) \right\} \right] \\
 &= \frac{n(n-1)}{N(N-1)} \sum_{i=1}^N \left\{ -(x_i - \mu)^2 \right\} \\
 &= -\frac{n(n-1)}{N-1} \cdot \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2 \\
 &= -\frac{n(n-1)}{N-1} \sigma^2
 \end{aligned}$$

$$\text{したがって } \sigma^2(\bar{X}) = \frac{1}{n^2} \left\{ n\sigma^2 - \frac{n(n-1)}{N-1} \sigma^2 \right\} = \frac{N-n}{N-1} \cdot \frac{\sigma^2}{n}$$

〔定理 3.7〕 無限母集団からの大きさ n の標本の標本平均 $\bar{X}$ の平均 $E(\bar{X})$ 、分散 $\sigma^2(\bar{X})$ は

$$E(\bar{X}) = \mu, \quad \sigma^2(\bar{X}) = \frac{\sigma^2}{n}$$

(証明) 定理 3.6 で $N \rightarrow \infty$ にすればよい。いま、統計量 U^2 を

$$\begin{aligned}
 U^2 &= \frac{1}{n} \{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \cdots + (X_n - \bar{X})^2\} \\
 &= \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2
 \end{aligned}$$

$$= \frac{n-1}{n} S^2$$

と定義し, $E(U^2)$ を求めてみよう.

$$\begin{aligned} \sum_{i=1}^n (X_i - \bar{X})^2 &= \sum_{i=1}^n \{(X_i - \mu) - (\bar{X} - \mu)\}^2 \\ &= \sum_{i=1}^n \{(X_i - \mu)^2 - 2(\bar{X} - \mu)(X_i - \mu) + (\bar{X} - \mu)^2\} \\ &= \sum_{i=1}^n (X_i - \mu)^2 - 2(\bar{X} - \mu)n(\bar{X} - \mu) + n(\bar{X} - \mu)^2 \\ &= \sum_{i=1}^n (X_i - \mu)^2 - n(\bar{X} - \mu)^2 \end{aligned}$$

$$\begin{aligned} \text{より} \quad E(U^2) &= E\left\{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2\right\} \\ &= E\left\{\frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2 - (\bar{X} - \mu)^2\right\} \\ &= \frac{1}{n} \sum_{i=1}^n E\{(X_i - \mu)^2\} - E\{(\bar{X} - \mu)^2\} \\ &= \sigma^2 - \sigma^2(\bar{X}) \end{aligned}$$

もし考えている母集団が無限母集団であるならば, 定理 3.7 より

$$\sigma^2(\bar{X}) = \frac{\sigma^2}{n}$$

であるから

$$E(U^2) = \frac{n-1}{n} \sigma^2$$

を得る. したがって

$$E\left(\frac{n}{n-1} U^2\right) = E(S^2) = \sigma^2$$

が成り立つ. これより母分散を推定する統計量として標本分散を

$$S^2 = \frac{1}{n-1} \{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \cdots + (X_n - \bar{X})^2\}$$

と定義した理由がわかるであろう.

問1 ある学年の実力テストの結果が, 平均点 72 点, 標準偏差 16 点の正規分布をし

ていることを知って、任意抽出した 36 人の生徒の平均点が 80 点以上である確率および 70 点以下である確率を求めよ。

問 2 母集団から任意抽出した標本

4.5, 4.4, 3.8, 3.7, 4.0, 4.7, 4.4, 4.6, 5.0, 4.7 (kg)

から、標本平均と標本分散を求めよ。

§3.5 中心極限定理

前節で $N(\mu, \sigma^2)$ 型の母集団からの大きさ n の標本平均 $\bar{X}$ は、正規分布 $N(\mu, \sigma^2/n)$ 型に従うことが示された。この節では、母集団の分布がどうであろうと、 n が大きいときは、 $\bar{X}$ は近似的に $N(\mu, \sigma^2/n)$ 型に従うことを示す。

〔定理 3.8〕 (中心極限定理) 母平均 μ , 母分散 σ^2 の任意の無限母集団から大きさ n の標本をとるとき、その標本平均 $\bar{X}$ は n が大きいとき、近似的に正規分布 $N(\mu, \sigma^2/n)$ に従う。

(証明) $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ とするとき、 $\bar{X}$ が $N(\mu, \sigma^2/n)$ に近似することを示すためには、

$$Y = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

で与えられる確率変数 Y が、 n が大きいとき、 $N(0, 1)$ に近似することを示せばよい。

$$Y = \frac{\frac{1}{n} \sum_{i=1}^n X_i - \mu}{\sigma/\sqrt{n}} = \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma} \right)$$

であるから

$$Y_i = \frac{X_i - \mu}{\sigma} \quad (i=1, 2, \dots, n)$$

とおき、確率変数 Y_i のモーメント母関数を $M_i(\theta)$ とすると(すべての i に共通)

$$\frac{1}{\sqrt{n}} Y_i$$

のモーメント母関数は $M_i(\theta/\sqrt{n})$ である(定理 2.19)。したがって

$$Y = \sum_{i=1}^n \frac{1}{\sqrt{n}} Y_i$$

のモーメント母関数を $M(\theta)$ とすると

$$M(\theta) = \left\{ M_1\left(\frac{\theta}{\sqrt{n}}\right) \right\}^n \quad (\text{定理 2.18})$$

となる。一方 $M_1(\theta)$ をテーラー展開すると

$$M_1(\theta) = M_1(0) + M_1'(0)\theta + \frac{1}{2} M_1''(\zeta)\theta^2 \quad (0 < \zeta < \theta)$$

また

$$M_1(0) = 1, \quad M_1'(0) = E(Y_i) = 0$$

であるから

$$\begin{aligned} M(\theta) &= \left\{ 1 + \frac{1}{2} M_1''(\zeta) \frac{\theta^2}{n} \right\}^n \quad \left(0 < \zeta < \frac{\theta}{\sqrt{n}} \right) \\ &= \left[\left\{ 1 + \frac{1}{2n/M_1''(\zeta)\theta^2} \right\}^{2n/M_1''(\zeta)\theta^2} \right]^{M_1''(\zeta)\theta^2/2} \end{aligned}$$

$n \rightarrow \infty$ のとき, $\zeta \rightarrow 0, M_1''(\zeta) \rightarrow M_1''(0) = \sigma^2(Y_i) + \{M_1'(0)\}^2 = 1$ であるから

$$\lim_{n \rightarrow \infty} M(\theta) = \lim_{n \rightarrow \infty} \left[\left\{ 1 + \frac{1}{2n/M_1''(\zeta)\theta^2} \right\}^{2n/M_1''(\zeta)\theta^2} \right]^{M_1''(\zeta)\theta^2/2} = e^{\theta^2/2}$$

$e^{\theta^2/2}$ は $N(0, 1)$ 型の確率変数のモーメント母関数であるから, 定理 2.16, 定理 2.17 により Y は, n が大きいとき, $N(0, 1)$ 型の正規分布に近似する。

図 3-2 が示すような分布から, 大きさ 5 の標本を 100 組無作為抽出して標本平均のヒストグラムを作成したら図 3-3 が得られた。

標本抽出数を $n=5$ としたが, n を大きくすれば図 3-3 のヒストグラムの形は正規分布に近づく ($n > 30$ ならば, 正規分布と考えてほとんど間違いはない)。

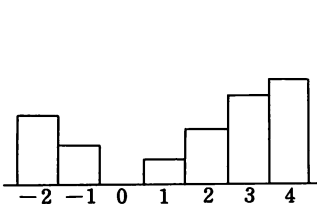


図 3-2

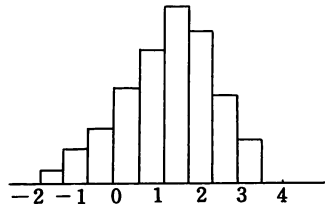


図 3-3

(例 1) ある年の 20 歳の男子の体重平均は 60 kg, 標準偏差は 5.4 kg であったとする. この中から 49 人を任意抽出して, その 49 人の体重平均が 62 kg 以上である確率を求めよ.

(解) 49 人の体重平均 $\bar{X}$ は中心極限定理によって $N(60, (5.4/7)^2)$ に近似するから

$$P(\bar{X} > 62) = 1 - P(\bar{X} \leq 62) = 1 - \Phi\left(\frac{62-60}{5.4/7}\right) = 1 - \Phi(2.59) \doteq 0.0048$$

問 1 ある年の満 20 歳の男の子の身長平均値が 159.72 cm, 標準偏差が 5.94 cm であった. いま任意に抽出した 64 人の男の子の身長を測定するとき, つぎの確率を求めよ.

(1) 64 人の平均身長が 162.00 cm 以上である確率

(2) 64 人の平均身長の母平均 159.72 cm からのずれが 1.00 cm 以上である確率を求めよ.

〔定理 3.9〕 (ラプラスの定理) 二項分布において, n が大きければ $N(np, npq)$ 型の正規分布とみなせる ($p+q=1$).

(証明) 1 と 0 からなる無限母集団を考え, その中に含まれる 1, 0 の割合をそれぞれ $p, q(p+q=1)$ とする. このとき, 母平均 μ , 母分散 σ^2 は

$$\mu = 1 \times p + 0 \times q = p$$

$$\sigma^2 = (1-p)^2 p + (0-p)^2 q = pq$$

である. この母集団からの大きさ n の標本の標本平均

$$\bar{Y} = \frac{1}{n}(Y_1 + Y_2 + \cdots + Y_n)$$

は, 中心極限定理により, n が大きいときは $N(p, pq/n)$ に近似する. いま確率変数 X を

$$X = Y_1 + Y_2 + \cdots + Y_n$$

とすると, X は n 個の標本中の 1 の個数を与える確率変数であるから,

$$P(X=k) = {}_n C_k p^k q^{n-k}$$

である. したがって, X は二項分布に従う. 一方,

$$X = n\bar{X}$$

であるから, 二項分布に従う確率変数 X は, n が大きいとき $N(np, npq)$ に近似している.

図 3-4 は $p=0.1$ とし, $n=10, 20, 50, 100$ の場合の二項分布のグラフである。 n が大きくなるに従い正規分布のグラフに近づいていることがわかる。

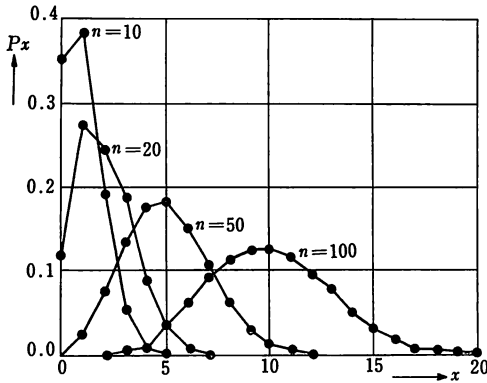


図 3-4

(例 2) 1600 枚の貨幣を投げて, 表の向く枚数を X 枚とするとき

$$X \leq 750, X \leq 800, X \leq 850$$

の場合の確率をそれぞれ求めよ。

(解) X はラプラスの定理により $N(np, npq)$ 型の正規分布に近似し

$$np = 1600 \times \frac{1}{2} = 800,$$

$$npq = 1600 \times \frac{1}{2} \times \frac{1}{2} = 20^2$$

であるから

$$P(X \leq 750) = \Phi\left(\frac{750-800}{20}\right) = \Phi(-2.5) = 1 - \Phi(2.5) = 0.0062$$

$$P(X \leq 800) = \Phi\left(\frac{800-800}{20}\right) = \Phi(0) = 0.5000$$

$$P(X \leq 850) = \Phi\left(\frac{850-800}{20}\right) = \Phi(2.5) = 0.9938$$

[注] 離散分布を連続分布に近似させているのであるから, より正確な値を求めるためには $P(X \leq 750)$ の代わりに $P(X \leq 750.5)$ とするほうがよい。

(例 3) ある国で, ある政策を支持する国民が 60% いるとき, 任意に選んだ人のうち半数以上がこの政策に反対する確率を求めよ。

(解) 政策を支持する者が60%であるから、40%が反対と考えよう。100人のうち反対する人数を X 人とすれば

$$P(X=k) = {}_{100}C_k \left(\frac{4}{10}\right)^k \left(\frac{6}{10}\right)^{100-k}$$

となり、 X は二項分布に従う。

$$n = 100, p = \frac{4}{10}, q = \frac{6}{10}$$

とおけば、

$$np = 40, npq = 24 \div (4.9)^2$$

であるから X は $N(40, 24)$ 型の正規分布に近似する。

$$\begin{aligned} P(X > 50) &= 1 - P(X \leq 50) = 1 - \Phi\left(\frac{50-40}{4.9}\right) \\ &= 1 - \Phi(2.04) = 1 - 0.979 = 0.021 \end{aligned}$$

問2 サイコロを360回投げて、1の目が

- (1) 70回以上 (2) 60回以上 (3) 40回以下

表れる確率をそれぞれ求めよ。

問3 不良率1%の製品を1000個入りのケースに入れたとき、そのケースの中に不良品が入っていない確率と15個以下である確率を求めよ。

〔定理 3.10〕 大きさが N で、母平均 μ 、母分散 σ^2 の任意の有限母集団から大きさ n の標本をとるとき、その標本平均 $\bar{X}$ は n が大きければ、 $N\left(\mu, \frac{\sigma^2}{n} \cdot \frac{N-n}{N-1}\right)$ 型の正規分布に近似する。(ただし $n \geq 50$ とする)

〔定理 3.11〕 N 個の中に性質 E をもつものの割合が p のとき、この中から n 個を抽出し、その中に性質 E をもつ個数を X 個とすると、 n が大きければ X は、 $N\left(np, npq \frac{N-n}{N-1}\right)$ 型の正規分布に近似する。(ただし $n \geq 50$ とする)

定理 3.10 は中心極限定理の、定理 3.11 はラプラスの定理のそれぞれ特別な場合である。

問4 定理 3.11 において、性質 E をもつものに1、性質 E をもたないものに0を結びつけ、定理 3.10 を用いて定理 3.11 を証明せよ。

(例4) ある学年の200人に対し一斉テストを行って平均78点、標準偏差16点を得た。この200人から任意抽出した50人の平均点が80点以上である確率を求めよ。

(解) 定理 3.10 より

$$\mu = 78, \quad \frac{\sigma^2}{n} \cdot \frac{N-n}{N-1} = \frac{16^2}{50} \cdot \frac{200-50}{200-1} \doteq 3.86$$

であるから, 求める確率は

$$1 - \Phi\left(\frac{80-78}{\sqrt{3.86}}\right) \doteq 1 - \Phi(1.02) \doteq 1 - 0.8461 = 0.1539$$

(例5) 1000個の製品の中に不良品が1%入っている. この中から200個抽出して不良品が1個以下である確率を求めよ.

(解) 定理 3.11において

$$p = 0.01, \quad q = 0.99, \quad N = 1000, \quad n = 200$$

$$np = 2, \quad npq \cdot \frac{N-n}{N-1} \doteq 1.584 \doteq (1.26)^2$$

であるから, 求める確率は

$$\Phi\left(\frac{1-2}{1.26}\right) \doteq \Phi(-0.79) = 1 - \Phi(0.79) \doteq 1 - 0.7852 = 0.2148$$

[注] 例2の注意と同様, この問題も離散分布を連続分布で近似させているのであるから, $\Phi\left(\frac{1-2}{1.26}\right)$ を用いるよりも, $\Phi\left(\frac{1.5-2}{1.26}\right)$ を用いた方がよい.

問5 ある学校の1000人の学生の体重平均は58 kg, 標準偏差6 kgであった. この中から100人を抽出して, 体重平均が60 kg以上である確率を求めよ.

問6 1000人の中に, ある政策に賛成するものが60%いるとする. この中から抽出した100人の中で賛成が半数以下である確率を求めよ.



第4章 その他の分布関数

§4.1 ガンマ関数とベータ関数

この章で χ^2 -分布, t -分布, F -分布を取り扱うが, その際にガンマ関数, ベータ関数を用いる. この節においてはこれらの関数について説明しよう.

$$\Gamma(t) = \int_0^{\infty} x^{t-1} e^{-x} dx \quad (t > 0) \quad (4.1.1)$$

で定義される t の関数 $\Gamma(t)$ をガンマ関数という. これより

$$\Gamma(1) = \int_0^{\infty} e^{-x} dx = 1$$

$$\Gamma(t+1) = \int_0^{\infty} x^t e^{-x} dx = [-e^{-x} x^t]_0^{\infty} + t \int_0^{\infty} x^{t-1} e^{-x} dx = t\Gamma(t)$$

したがって

$$\Gamma(1) = 1, \quad \Gamma(t+1) = t\Gamma(t) \quad (4.1.2)$$

が成り立つ. さらに (4.1.2) より, n が正の整数ならば,

$$\Gamma(n+1) = n! \quad (4.1.3)$$

が得られる. また, 等式

$$\int_0^{\infty} x^{t-1} e^{-ax} dx = \frac{\Gamma(t)}{a^t} \quad (a > 0) \quad (4.1.4)$$

が成り立つ (左辺において $ax=y$ とおけば右辺が導かれる).

つぎに

$$B(\alpha, \beta) = \int_0^1 t^{\alpha-1} (1-t)^{\beta-1} dt \quad (\alpha > 0, \beta > 0) \quad (4.1.5)$$

で定義される α, β の関数 $B(\alpha, \beta)$ をベータ関数という. (4.1.5) において $t = \frac{x}{1+x}$ とおけば,

$$B(\alpha, \beta) = \int_0^{\infty} \frac{x^{\alpha-1}}{(1+x)^{\alpha+\beta}} dx$$

$\Gamma(t)$ と $B(\alpha, \beta)$ との間の関係を調べてみよう.

(4.1.4) において, t を α , a を t に書き換えて両辺に $t^{\alpha+\beta-1} e^{-t}$ をかけ

$$\Gamma(\alpha)t^{\beta-1}e^{-t} = \int_0^\infty t^{\alpha+\beta-1}x^{\alpha-1}e^{-(1+x)t}dx$$

つぎに、この式の両辺を t について 0 から ∞ まで積分すれば、

$$\text{左辺} = \int_0^\infty \Gamma(\alpha)t^{\beta-1}e^{-t}dt = \Gamma(\alpha)\Gamma(\beta)$$

$$\text{右辺} = \int_0^\infty \left\{ \int_0^\infty t^{\alpha+\beta-1}x^{\alpha-1}e^{-(1+x)t}dx \right\} dt = \int_0^\infty \left\{ x^{\alpha-1} \int_0^\infty t^{\alpha+\beta-1}e^{-(1+x)t}dt \right\} dx$$

ここで、(4.1.4)において t を $\alpha+\beta$, x を t , a を $1+x$ に書き換えると

$$\int_0^\infty t^{\alpha+\beta-1}e^{-(1+x)t}dt = \frac{\Gamma(\alpha+\beta)}{(1+x)^{\alpha+\beta}}$$

であるから、

$$\text{右辺} = \int_0^\infty x^{\alpha-1} \cdot \frac{\Gamma(\alpha+\beta)}{(1+x)^{\alpha+\beta}} dx = \Gamma(\alpha+\beta)B(\alpha, \beta)$$

したがって

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} \quad (4.1.6)$$

(4.1.6)において $\alpha=\beta=1/2$ とすれば、

$$B\left(\frac{1}{2}, \frac{1}{2}\right) = \frac{\Gamma(1/2)\Gamma(1/2)}{\Gamma(1)}$$

ところで

$$B\left(\frac{1}{2}, \frac{1}{2}\right) = \int_0^1 t^{-1/2}(1-t)^{-1/2}dt = \int_0^1 \frac{dt}{\sqrt{t(1-t)}} = \pi$$

$$\Gamma(1) = 1$$

であるから

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi} \quad (4.1.7)$$

つぎに $a>0$, $t>0$ に対して x の関数 $f(x)$ を

$$f(x) = \begin{cases} \frac{a^t}{\Gamma(t)} x^{t-1} e^{-ax} & (x>0) \\ 0 & (x\leq 0) \end{cases} \quad (4.1.8)$$

で定義すると

$$\int_{-\infty}^{\infty} f(x)dx = \frac{a^t}{\Gamma(t)} \int_0^{\infty} x^{t-1} e^{-ax} dx = \frac{a^t}{\Gamma(t)} \cdot \frac{\Gamma(t)}{a^t} = 1$$

かつ, $f(x) \geq 0$ であるから, $f(x)$ はある確率変数の確率密度関数と考えられる。
このとき, この確率変数のモーメント母関数 $M(\theta)$ は

$$M(\theta) = \int_0^{\infty} e^{\theta x} \cdot \frac{a^t}{\Gamma(t)} x^{t-1} e^{-ax} dx = \frac{a^t}{\Gamma(t)} \int_0^{\infty} x^{t-1} e^{-(a-\theta)x} dx$$

$a-\theta > 0$ のとき, (4.1.4) を用いて

$$M(\theta) = \frac{a^t}{\Gamma(t)} \cdot \frac{\Gamma(t)}{(a-\theta)^t} = \frac{1}{(1-\theta/a)^t} \quad (4.1.9)$$

§4.2 χ^2 - 分 布

n 個の確率変数 $X_1, X_2, \dots, X_n$ が互いに独立で, それぞれ正規分布 $N(0, 1)$ に従うとき,

$$X = X_1^2 + X_2^2 + \dots + X_n^2$$

の確率密度関数 $f(x)$ を求めてみよう。まず

$$Y = X_i^2 \quad (i=1, 2, \dots, n)$$

とおき, Y の分布関数, 確率密度関数を $G(y), g(y)$ とすれば,

$y > 0$ のとき

$$\begin{aligned} G(y) &= P(Y \leq y) = P(-\sqrt{y} \leq X_i \leq \sqrt{y}) \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\sqrt{y}}^{\sqrt{y}} e^{-x^2/2} dx = \frac{2}{\sqrt{2\pi}} \int_0^{\sqrt{y}} e^{-x^2/2} dx \end{aligned}$$

ここで, $x = \sqrt{t}$ とおき

$$G(y) = \int_0^y \frac{1}{\sqrt{2\pi t}} e^{-t/2} dt$$

この両辺を y で微分して

$$g(y) = \frac{1}{\sqrt{2\pi y}} e^{-y/2} \quad (y > 0)$$

また, $y \leq 0$ のとき $G(y) = 0$ であるから

$$g(y) = \begin{cases} \frac{1}{\sqrt{2\pi y}} e^{-y/2} & (y > 0) \\ 0 & (y \leq 0) \end{cases} \quad (4.2.1)$$

したがって、 Y のモーメント母関数 $M_y(\theta)$ は

$$M_y(\theta) = \int_{-\infty}^{\infty} e^{\theta y} g(y) dy = \int_0^{\infty} e^{\theta y} \cdot \frac{1}{\sqrt{2\pi} y} e^{-y/2} dy = \frac{1}{\sqrt{2\pi}} \int_0^{\infty} y^{1/2-1} e^{-(1/2-\theta)y} dy$$

$1/2-\theta > 0$ のとき、(4.1.4)を用いて

$$M_y(\theta) = \frac{1}{\sqrt{2\pi}} \cdot \frac{\Gamma(1/2)}{(1/2-\theta)^{1/2}} = \frac{1}{(1-2\theta)^{1/2}}$$

ゆえに、 X のモーメント母関数 $M(\theta)$ は定理 2.18 を用いて

$$M(\theta) = \frac{1}{(1-2\theta)^{n/2}} \quad (4.2.2)$$

と書ける。この式は(4.1.9)において、 $a=1/2$, $t=n/2$ としたものであるから定理 2.16 により

$$f(x) = \begin{cases} \frac{1}{2^{n/2}\Gamma(n/2)} x^{n/2-1} e^{-x/2} & (x > 0) \\ 0 & (x \leq 0) \end{cases} \quad (4.2.3)$$

を得る。

確率変数 X の確率密度関数が(4.2.3)で表されるとき、 X は自由度 n の χ^2 -分布に従うという。

自由度 n の χ^2 -分布に従う確率変数 X と確率 α に対し

$$P(X \geq \lambda) = \alpha \quad (0 < \alpha < 1)$$

をみたす λ の値を求める表がある。この表を χ^2 -分布表といい、巻末にのせてある。今後この λ を $\chi^2_{n,\alpha}$ と書くことにする。

自由度 n の χ^2 -分布に従う確率変数 X の確率密度関数 $f(x)$ のグラフは図 4-1 によって与えられる(X が χ^2 -分布に従うとき、 X を χ^2 で表すこともある)。

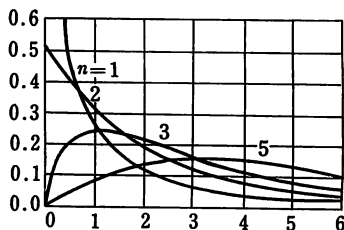


図 4-1

(例 1) χ_1^2, χ_2^2 が互いに独立でそれぞれ自由度 n_1, n_2 の χ^2 -分布に従うとき、 $\chi_1^2 + \chi_2^2$

は自由度 $n_1 + n_2$ の χ^2 -分布に従う(この性質を χ^2 -分布の再生性という)。

(解) (4.2.2)より、 χ_1^2, χ_2^2 のモーメント母関数はそれぞれ

$$\frac{1}{(1-2\theta)^{n_1/2}}, \quad \frac{1}{(1-2\theta)^{n_2/2}}$$

であるから, $\chi_1^2 + \chi_2^2$ のモーメント母関数は

$$\frac{1}{(1-2\theta)^{n_1/2}} \times \frac{1}{(1-2\theta)^{n_2/2}} = \frac{1}{(1-2\theta)^{(n_1+n_2)/2}}$$

となり, これは $\chi_1^2 + \chi_2^2$ が自由度 $n_1 + n_2$ の χ^2 -分布に従うことを示している.

§4.3 t -分布

$n+1$ 個の確率変数 $X_1, X_2, \dots, X_n, Y$ が, 互いに独立でそれぞれ正規分布 $N(0, \sigma^2)$ に従うとき,

$$T = \frac{Y}{\sqrt{\frac{1}{n} \sum_{i=1}^n X_i^2}}$$

の確率密度関数 $p(t)$ を求めてみよう.

X_i/σ ($i=1, 2, \dots, n$) は正規分布 $N(0, 1)$ に従うから

$$X = \sum_{i=1}^n \left(\frac{X_i}{\sigma} \right)^2$$

の確率密度関数 $f(x)$ は (4.2.3) により

$$f(x) = \begin{cases} \frac{1}{2^{n/2} \Gamma(n/2)} x^{n/2-1} e^{-x/2} & (x > 0) \\ 0 & (x \leq 0) \end{cases}$$

である. したがって,

$$Z = \frac{1}{n} \sum_{i=1}^n X_i^2 = \frac{\sigma^2}{n} X$$

の確率密度関数 $h(z)$ は

$$\begin{aligned} h(z) &= \frac{n}{\sigma^2} f\left(\frac{nz}{\sigma^2}\right) \\ &= \begin{cases} \frac{(n/2)^{n/2}}{\sigma^n \Gamma(n/2)} z^{n/2-1} e^{-nz/2\sigma^2} & (z > 0) \\ 0 & (z \leq 0) \end{cases} \end{aligned} \quad (4.3.1)$$

つぎに,

$$U = \sqrt{\frac{1}{n} \sum_{i=1}^n X_i^2} = \sqrt{Z}$$

とおき, U の確率密度関数を $k(u)$ とすれば, $u > 0$ のとき,

$$\int_0^u k(u) du = P(U \leq u) = P(Z \leq u^2) = \int_0^{u^2} h(z) dz$$

であるから, 上式の両辺を u で微分して

$$k(u) = 2uh(u^2)$$

$$\therefore k(u) = \begin{cases} \frac{2(n/2)^{n/2}}{\sigma^n \Gamma(n/2)} u^{n-1} e^{-nu^2/2\sigma^2} & (u > 0) \\ 0 & (u \leq 0) \end{cases}$$

一方 Y の確率密度関数 $g(y)$ は

$$g(y) = \frac{1}{\sqrt{2\pi} \sigma} e^{-y^2/2\sigma^2}$$

で, Y と U は互いに独立であるから, Y と U の二次元確率密度関数 $p(y, u)$ は

$$p(y, u) = g(y) \cdot k(u)$$

したがって

$$p(y, u) = \begin{cases} \frac{\sqrt{2}}{\sqrt{\pi}} \cdot \frac{(n/2)^{n/2}}{\sigma^{n+1} \Gamma(n/2)} u^{n-1} e^{-(y^2+nu^2)/2\sigma^2} & (u > 0) \\ 0 & (u \leq 0) \end{cases}$$

である. ゆえに定理 2.10 を用いて, 確率変数

$$T = \frac{Y}{\sqrt{\frac{1}{n} \sum_{i=1}^n X_i^2}} = \frac{Y}{U}$$

の確率密度関数 $p(t)$ は

$$\begin{aligned} p(t) &= \int_{-\infty}^{\infty} p(tv, v) |v| dv = \int_0^{\infty} p(tv, v) v dv \\ &= \int_0^{\infty} \frac{\sqrt{2}}{\sqrt{\pi}} \cdot \frac{(n/2)^{n/2}}{\sigma^{n+1} \Gamma(n/2)} v^n e^{-(t^2 v^2 + n v^2)/2\sigma^2} dv \end{aligned}$$

ここで, $\left(\frac{t^2+n}{2\sigma^2}\right)v^2 = z$ とおけば

$$2\left(\frac{t^2+n}{2\sigma^2}\right)v dv = dz, \quad v = \left(\frac{2\sigma^2}{t^2+n}\right)^{1/2} \cdot z^{1/2}$$

であるから

$$\begin{aligned}
 p(t) &= \frac{1}{\sqrt{n\pi} \cdot \Gamma(n/2)} \cdot \left(\frac{n}{t^2 + n} \right)^{(n+1)/2} \int_0^\infty z^{(n+1)/2-1} e^{-z} dz \\
 &= \frac{\Gamma((n+1)/2)}{\sqrt{n\pi} \cdot \Gamma(n/2)} \left(1 + \frac{t^2}{n} \right)^{-(n+1)/2} \quad (4.3.2)
 \end{aligned}$$

確率変数 T の確率密度関数が (4.3.2) で表されるとき, T は自由度 n の t -分布に従うという ($p(t)$ の中に σ が含まれていないことに注意).

自由度 n の t -分布に従う確率変数 T と確率 α に対し,

$$P(T \geq \lambda) = \alpha \quad (0 < \alpha < 1)$$

をみtus λ の値を求める表がある. この表を t -分布表といい, 巻末にのせてある. 今後この λ を $t_{n,\alpha}$ と書くことにする.

自由度 n の t -分布に従う確率変数 T の確率密度関数 $p(t)$ のグラフは図 4-2 によって与えられる.

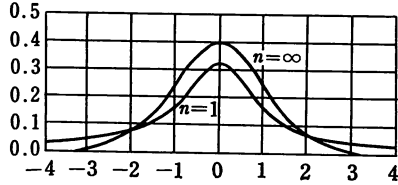


図 4-2

このグラフは直線 $t=0$ に関して対称である. また自由度 n の t -分布は,

n が限りなく大きくなると正規分布 $N(0, 1)$ に近づく. いいかえれば

$$\lim_{n \rightarrow \infty} p(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$$

が成り立つ. これをスターリングの近似式

$$\Gamma(m+1) \doteq \sqrt{2\pi m} m^m e^{-m} \quad (m \doteq \infty)$$

を用いて証明してみよう.

$$\begin{aligned}
 p(t) &= \frac{1}{\sqrt{n\pi}} \cdot \frac{\Gamma((n+1)/2)}{\Gamma(n/2)} \left(1 + \frac{t^2}{n} \right)^{-(n+1)/2} \\
 &= \frac{1}{\sqrt{n\pi}} \cdot \frac{\Gamma\left(\frac{n-1}{2} + 1\right)}{\Gamma\left(\frac{n-2}{2} + 1\right)} \left(1 + \frac{t^2}{n} \right)^{-(n+1)/2} \\
 &\doteq \frac{1}{\sqrt{n\pi}} \cdot \frac{\sqrt{(n-1)\pi} \left(\frac{n-1}{2} \right)^{(n-1)/2} e^{-(n-1)/2}}{\sqrt{(n-2)\pi} \left(\frac{n-2}{2} \right)^{(n-2)/2} e^{-(n-2)/2}} \left(1 + \frac{t^2}{n} \right)^{-(n+1)/2}
 \end{aligned}$$

$$= \frac{1}{\sqrt{2\pi}} e^{-1/2} \left\{ \left(1 + \frac{1}{n-2} \right)^{n-2} \right\}^{1/2} \left(1 + \frac{1}{n-2} \right)^{1/2} \left(1 - \frac{1}{n} \right)^{1/2} \\ \times \left(1 + \frac{t^2/2}{n/2} \right)^{-n/2} \left(1 + \frac{t^2}{n} \right)^{-1/2}$$

ところで

$$\left\{ \left(1 + \frac{1}{n-2} \right)^{n-2} \right\}^{1/2} \rightarrow e^{1/2}, \quad \left(1 + \frac{1}{n-2} \right)^{1/2} \rightarrow 1, \quad \left(1 - \frac{1}{n} \right)^{1/2} \rightarrow 1, \\ \left(1 + \frac{t^2/2}{n/2} \right)^{-n/2} \rightarrow e^{-t^2/2}, \quad \left(1 + \frac{t^2}{n} \right)^{-1/2} \rightarrow 1$$

であるから

$$\lim_{n \rightarrow \infty} p(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$$

§4.4 F - 分 布

$m+n$ 個の確率変数 $X_1, X_2, \dots, X_m; Y_1, Y_2, \dots, Y_n$ が互いに独立で、それぞれ正規分布 $N(0, \sigma^2)$ に従うとき

$$F = \frac{\frac{1}{m} \sum_{i=1}^m X_i^2}{\frac{1}{n} \sum_{j=1}^n Y_j^2}$$

の確率密度関数 $p(f)$ を求めてみよう。

$$X = \sum_{i=1}^m X_i^2, \quad Y = \sum_{j=1}^n Y_j^2$$

とおき、 X, Y の確率密度関数をそれぞれ $f(x), g(y)$ とおけば、

$$X = \sigma^2 \sum_{i=1}^m \left(\frac{X_i}{\sigma} \right)^2, \quad Y = \sigma^2 \sum_{j=1}^n \left(\frac{Y_j}{\sigma} \right)^2$$

また、 X_i/σ ($i=1, 2, \dots, m$) および Y_j/σ ($j=1, 2, \dots, n$) はそれぞれ正規分布 $N(0, 1)$ に従うから、(4.3.1)を導いたのと同じ方法で

$$f(x) = \begin{cases} \frac{1}{2^{m/2} \sigma^m \Gamma(m/2)} x^{m/2-1} e^{-x/2\sigma^2} & (x > 0) \\ 0 & (x \leq 0) \end{cases}$$

$$g(y) = \begin{cases} \frac{1}{2^{n/2}\sigma^n\Gamma(n/2)} y^{n/2-1} e^{-y/2\sigma} & (y>0) \\ 0 & (y\leq 0) \end{cases}$$

が得られる。また、 X と Y は互いに独立であるから、 X と Y の二次元確率密度関数を $p(x, y)$ とすれば、

$$p(x, y) = f(x) \cdot g(y) \\ = \begin{cases} \frac{1}{2^{(m+n)/2}\sigma^{m+n}\Gamma(m/2)\Gamma(n/2)} x^{m/2-1} y^{n/2-1} e^{-(x+y)/2\sigma} & (x>0 \text{ かつ } y>0) \\ 0 & (x\leq 0 \text{ または } y\leq 0) \end{cases}$$

ゆえに $U=X/Y$ とおけば U の確率密度関数 $h(u)$ は定理 2.10 を用いて

$$h(u) = \int_{-\infty}^{\infty} p(uv, v) |v| dv = \int_0^{\infty} p(uv, v) v dv \quad (4.4.1)$$

したがって、 $u>0$ のとき

$$h(u) = \int_0^{\infty} \frac{1}{2^{(m+n)/2}\sigma^{m+n}\Gamma(m/2)\Gamma(n/2)} (uv)^{m/2-1} v^{n/2-1} e^{-(uv+v)/2\sigma^2} v dv \\ = \frac{u^{m/2-1}}{2^{(m+n)/2}\sigma^{m+n}\Gamma(m/2)\Gamma(n/2)} \int_0^{\infty} v^{(m+n)/2-1} e^{-((1+u)v)/2\sigma^2} dv$$

ここで、 $\frac{1+u}{2\sigma^2} v = z$ とおけば

$$v = \frac{2\sigma^2}{1+u} z \quad \text{かつ} \quad dv = \frac{2\sigma^2}{1+u} dz$$

であるから

$$h(u) = \frac{u^{m/2-1}}{\Gamma(m/2)\Gamma(n/2)} \cdot \frac{1}{(1+u)^{(m+n)/2}} \int_0^{\infty} z^{(m+n)/2-1} e^{-z} dz \\ = \frac{\Gamma((m+n)/2)}{\Gamma(m/2)\Gamma(n/2)} \cdot \frac{u^{m/2-1}}{(1+u)^{(m+n)/2}} = \frac{1}{B(m/2, n/2)} \cdot \frac{u^{m/2-1}}{(1+u)^{(m+n)/2}}$$

また、 $u\leq 0$ のときは (4.4.1) より

$$h(u) = 0$$

が得られるから

$$h(u) = \begin{cases} \frac{1}{B(m/2, n/2)} \cdot \frac{u^{m/2-1}}{(1+u)^{(m+n)/2}} & (u>0) \\ 0 & (u\leq 0) \end{cases}$$

$$\text{ところで } F = \frac{\frac{1}{m} \sum_{i=1}^m X_i^2}{\frac{1}{n} \sum_{j=1}^n Y_j^2} = \frac{n}{m} U \text{ より}$$

$$p(f) = \frac{m}{n} h\left(\frac{m}{n} f\right)$$

$$\therefore p(f) = \begin{cases} \frac{m^{m/2} n^{n/2}}{B(m/2, n/2)} \cdot \frac{f^{m/2-1}}{(mf+n)^{(m+n)/2}} & (f > 0) \\ 0 & (f \leq 0) \end{cases} \quad (4.4.2)$$

確率変数 F の確率密度関数が (4.4.2) で表されるとき, F は自由度 m, n の F -分布に従うという。

自由度 m, n の F -分布に従う確率変数 F と確率 α に対し,

$$P(F \geq \lambda) = \alpha \quad (\alpha = 0.05, 0.025, 0.01)$$

をみたす λ を求める表がある。この表を F -分布表といい、巻末にのせてある。今後この λ を $F_{n, \alpha}^m$ と書くことにする。

自由度 m, n の F -分布をする確率変数 F の確率密度関数 $p(f)$ のグラフは図 4-3 によって与えられる。

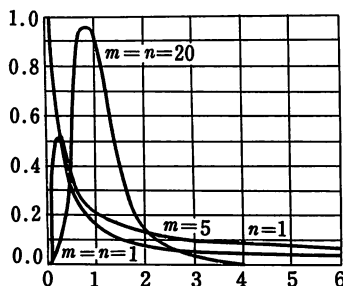


図 4-3

§ 4.5 χ^2 -分布, t -分布, F -分布をする標本分布

この節においては、4.2 節、4.3 節、4.4 節で扱った χ^2 -分布, t -分布, F -分布をする標本分布を調べよう。そのためにはつぎの定理を必要とする。

〔定理 4.1〕 正規母集団 $N(\mu, \sigma^2)$ から大きさ n の標本を任意抽出し、その標本平均を $\bar{X}$ 、標本標準偏差を S とすれば、 $\bar{X}, S$ の二次元確率密度関数 $p(\bar{x}, s)$ は

$$p(\bar{x}, s) = \begin{cases} \frac{n^{1/2}(n-1)^{(n-1)/2}}{2^{(n-2)/2} \sqrt{\pi} \Gamma\left(\frac{n-1}{2}\right) \sigma^n} s^{n-2} e^{-\frac{n(\bar{x}-\mu)^2 + (n-1)s^2}{2\sigma^2}} & (s > 0) \\ 0 & (s \leq 0) \end{cases}$$

によって与えられる。(証明は本書の程度を越えるので省略する)

〔定理 4.2〕 定理 4.1 のもとで $U=S^2$ とおけば, U の確率密度関数 $p(u)$ は

$$p(u) = \begin{cases} \frac{(n-1)^{(n-1)/2}}{2^{(n-1)/2} \Gamma\left(\frac{n-1}{2}\right) \sigma^{n-1}} u^{(n-3)/2} e^{-\frac{(n-1)u}{2\sigma^2}} & (u > 0) \\ 0 & (u \leq 0) \end{cases}$$

によって与えられる。(証明は本書の程度を越えるので省略する)

〔定理 4.3〕 正規母集団 $N(\mu, \sigma^2)$ から大きさ n の標本を任意抽出し, その標本平均を $\bar{X}$, 標本標準偏差を S とすれば,

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

は自由度 $n-1$ の t -分布をする.

この定理における T は母分散 σ^2 が未知である正規母集団の母平均 μ の推定や検定の際に使われる理論上の確率変数である(後述).

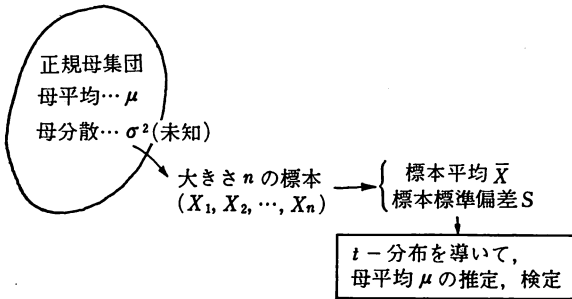


図 4-4 母平均 μ の推定, 検定

(証明) 定理 4.1 により, $\bar{X}, S$ の二次元確率密度関数 $p(\bar{x}, s)$ は

$$p(\bar{x}, s) = \frac{n^{1/2}(n-1)^{(n-1)/2}}{2^{(n-2)/2} \sqrt{\pi} \Gamma\left(\frac{n-1}{2}\right) \sigma^n} s^{n-2} e^{-\frac{n(\bar{x}-\mu)^2 + (n-1)s^2}{2\sigma^2}} \quad (s > 0)$$

によって与えられる. したがって

$$U = \sqrt{n}(\bar{X} - \mu) \quad (\bar{X} = U/\sqrt{n} + \mu)$$

とおけば, 定理 2.8 により, U, S の二次元確率密度関数 $p_1(u, s)$ は

$$p_1(u, s) = p\left(\frac{u}{\sqrt{n}} + \mu, s\right) |J| \quad \left(J = \begin{vmatrix} \frac{1}{\sqrt{n}} & 0 \\ 0 & 1 \end{vmatrix} = \frac{1}{\sqrt{n}}\right)$$

$$= \begin{cases} \frac{(n-1)^{(n-1)/2}}{2^{(n-2)/2} \sqrt{\pi} \Gamma\left(\frac{n-1}{2}\right) \sigma^n} s^{n-2} e^{-\frac{u^2 + (n-1)s^2}{2\sigma^2}} & (s > 0) \\ 0 & (s \leq 0) \end{cases}$$

ゆえに $T = \frac{\bar{X} - \mu}{S/\sqrt{n}} = \frac{U}{S}$ の確率密度関数 $p(t)$ は定理 2.10 を用いて

$$p(t) = \int_{-\infty}^{\infty} p_1(tv, v) |v| dv$$

$$= \int_0^{\infty} \frac{(n-1)^{(n-1)/2}}{2^{(n-2)/2} \sqrt{\pi} \Gamma\left(\frac{n-1}{2}\right) \sigma^n} v^{n-2} e^{-\frac{t^2 v^2 + (n-1)v^2}{2\sigma^2}} v dv$$

ここで, $\frac{t^2 + (n-1)}{2\sigma^2} v^2 = z$ とおけば $\frac{t^2 + (n-1)}{\sigma^2} v dv = dz$ であるから,

$$p(t) = \frac{(n-1)^{(n-1)/2}}{\sqrt{\pi} \Gamma\left(\frac{n-1}{2}\right) \{t^2 + (n-1)\}^{n/2}} \int_0^{\infty} z^{(n-2)/2} e^{-z} dz$$

$$= \frac{1}{\sqrt{(n-1)\pi} \Gamma\left(\frac{n-1}{2}\right)} \left(1 + \frac{t^2}{n-1}\right)^{-n/2} \int_0^{\infty} z^{n/2-1} e^{-z} dz$$

$$= \frac{\Gamma(n/2)}{\sqrt{(n-1)\pi} \Gamma\left(\frac{n-1}{2}\right)} \left(1 + \frac{t^2}{n-1}\right)^{-n/2}$$

これは確率変数 T が自由度 $n-1$ の t -分布に従うことを示している。

(例 1) 正規母集団 $N(\mu, \sigma^2)$ から大きさ 10 の標本を無作為抽出したとき,

$$T = \frac{\bar{X} - \mu}{S/\sqrt{10}}$$

について, つぎの式をみたす λ の値を求めよ。

$$(1) P(T \geq \lambda) = 0.05, \quad (2) P(|T| < \lambda) = 0.95$$

(解) T は自由度 9 の t -分布に従うから, t -分布表より,

$$(1) \lambda = 1.8331$$

(2) $P(|T| < \lambda) = 0.95$ であるから T の確率密度関数の対称性を用いて,

$$P(T \geq \lambda) = 0.025 \quad \therefore \lambda = 2.2622$$

〔定理 4.4〕 正規母集団 $N(\mu, \sigma^2)$ から大きさ n の標本を無作為抽出し、その標本分散を S^2 とすれば

$$X = \frac{(n-1)S^2}{\sigma^2}$$

は自由度 $n-1$ の χ^2 -分布に従う。

(証明) 定理 4.2 により, $U=S^2$ の確率密度関数 $p(u)$ は

$$p(u) = \begin{cases} \frac{(n-1)^{(n-1)/2}}{2^{(n-1)/2} \Gamma\left(\frac{n-1}{2}\right) \sigma^{n-1}} u^{(n-3)/2} e^{-\frac{(n-1)u}{2\sigma^2}} & (u > 0) \\ 0 & (u \leq 0) \end{cases}$$

によって与えられ、また

$$X = \frac{(n-1)S^2}{\sigma^2} = \frac{n-1}{\sigma^2} U$$

であるから, X の確率密度関数を $f(x)$ とおけば,

$$f(x) = \frac{\sigma^2}{n-1} p\left(\frac{\sigma^2 x}{n-1}\right)$$

ゆえに, $x > 0$ のとき

$$f(x) = \frac{1}{2^{(n-1)/2} \Gamma\left(\frac{n-1}{2}\right)} x^{(n-1)/2-1} e^{-x/2}$$

で, $x \leq 0$ のとき $f(x)=0$ である。したがって, X は自由度 $n-1$ の χ^2 -分布に従う。

この定理の X は正規母集団の母分散 σ^2 の推定や検定の際に使われる理論上の確率変数である(後述)。

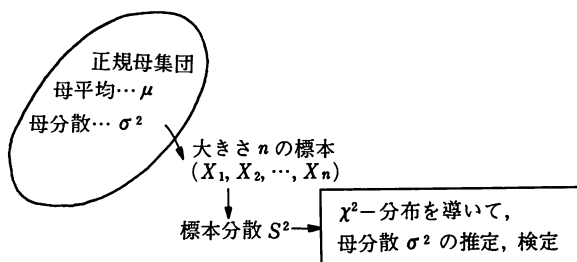


図 4-5 母分散 σ^2 の推定, 検定

(例 2) 正規母集団 $N(\mu, \sigma^2)$ から大きさ 5 の標本を無作為抽出したとき,

$$\chi^2 = \frac{4S^2}{\sigma^2}$$

について, つぎの式をみたす λ の値を求めよ.

$$(1) P(\chi^2 \geq \lambda) = 0.05, \quad (2) P(\chi^2 < \lambda) = 0.975$$

(解) χ^2 は自由度 4 の χ^2 -分布に従うから, χ^2 -分布表より,

$$(1) \lambda = 9.49$$

$$(2) P(\chi^2 < \lambda) = 0.975 \text{ であるから } P(\chi^2 \geq \lambda) = 0.025$$

ゆえに, $\lambda = 11.14$

つぎに χ^2 -分布をする確率変数と F -分布をする確率変数との間の関係を調べよう.

$m+n$ 個の確率変数 $X_1, X_2, \dots, X_m; Y_1, Y_2, \dots, Y_n$ が互いに独立で, それぞれ正規分布 $N(0, \sigma^2)$ に従うとき

$$F = \frac{\frac{1}{m} \sum_{i=1}^m X_i^2}{\frac{1}{n} \sum_{j=1}^n Y_j^2}$$

は自由度 m, n の F -分布をし,

$$\chi_1^2 = \frac{1}{\sigma^2} \sum_{i=1}^m X_i^2, \quad \chi_2^2 = \frac{1}{\sigma^2} \sum_{j=1}^n Y_j^2$$

はそれぞれ自由度 m , 自由度 n の χ^2 -分布をする. 一方

$$F = \frac{\chi_1^2/m}{\chi_2^2/n}$$

と書けるから, これからつぎの定理が得られる.

〔定理 4.5〕 2つの確率変数 χ_1^2, χ_2^2 が互いに独立でそれぞれ自由度 n_1 , 自由度 n_2 の χ^2 -分布に従うとき,

$$F = \frac{\chi_1^2/n_1}{\chi_2^2/n_2}$$

は自由度 n_1, n_2 の F -分布に従う.

〔定理 4.6〕 母分散がそれぞれ σ_1^2, σ_2^2 である互いに独立な 2つの正規母集団から, 大きさ n_1, n_2 の標本を無作為抽出し, その標本分散を S_1^2, S_2^2 とする

とき

$$F = \frac{\sigma_2^2 S_1^2}{\sigma_1^2 S_2^2}$$

は自由度 n_1-1 , n_2-1 の F -分布に従う.

$$(\text{証明}) \quad \chi_1^2 = \frac{(n_1-1)S_1^2}{\sigma_1^2}, \quad \chi_2^2 = \frac{(n_2-1)S_2^2}{\sigma_2^2}$$

とおけば, 定理 4.4 により χ_1^2, χ_2^2 はそれぞれ自由度 n_1-1 , 自由度 n_2-1 の χ^2 -分布に従うから, 定理 4.5 を用いて

$$F = \frac{\chi_1^2/(n_1-1)}{\chi_2^2/(n_2-1)} = \frac{\sigma_2^2 S_1^2}{\sigma_1^2 S_2^2}$$

は自由度 n_1-1 , n_2-1 の F -分布に従う.

この定理は互いに独立な2つの正規母集団 A, B の母分散の比率を推定するときや, 等分散であるかどうかの検定などの際に使われる理論上の確率変数である(後述). χ^2 -分布が母分散そのものの推定や検定に使われたことと対比しておぼえておくといよい.

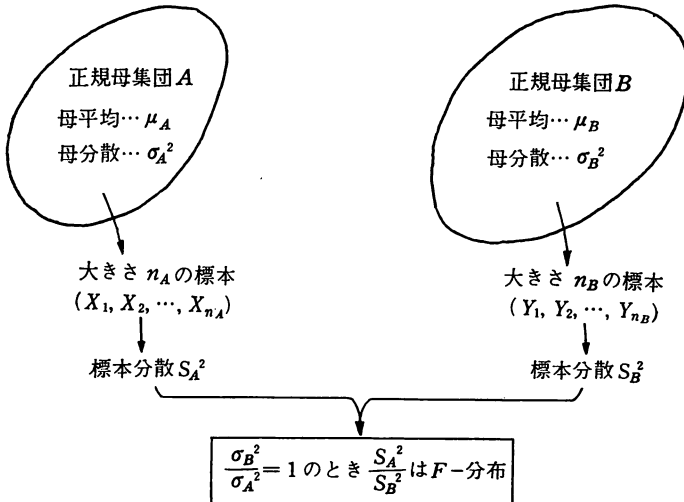


図 4-6

(例 3) 母分散の等しい互いに独立な2つの正規母集団から, 大きさ 10, 15 の標本を無作為抽出し, その標本分散を S_1^2, S_2^2 とするとき

$$F = S_1^2/S_2^2$$

について、つぎの式をみたす λ の値を求めよ。

$$(1) P(F \geq \lambda) = 0.05, \quad (2) P(F \geq \lambda) = 0.01$$

(解) $F = S_1^2/S_2^2$ は自由度 9, 14 の F -分布に従うから、 F -分布表より

$$(1) \lambda = 2.646, \quad (2) \lambda = 4.030$$

(例 4) F が自由度 n_1, n_2 の F -分布に従うならば、 $1/F$ は自由度 n_2, n_1 の F -分布に従う。

(証明) 定理 4.5 により、 F は自由度がそれぞれ n_1, n_2 の χ^2 -分布に従う確率変数 χ_1^2, χ_2^2 に対し、

$$F = \frac{\chi_1^2/n_1}{\chi_2^2/n_2}$$

と書ける。したがって

$$\frac{1}{F} = \frac{\chi_2^2/n_2}{\chi_1^2/n_1}$$

であるから、 $1/F$ は自由度 n_2, n_1 の F -分布に従う。

[定理 4.7] 母分散の等しい互いに独立な 2 つの正規母集団 $N(\mu_1, \sigma^2)$, $N(\mu_2, \sigma^2)$ から、大きさ n_1, n_2 の標本を無作為抽出し、その標本平均、標本分散をそれぞれ $\bar{X}_1, \bar{X}_2, S_1^2, S_2^2$ とすれば、

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}} \sqrt{\frac{n_1 n_2 (n_1 + n_2 - 2)}{n_1 + n_2}}$$

は自由度 $n_1 + n_2 - 2$ の t -分布をする。

(証明) $\bar{X}_1, \bar{X}_2$ はそれぞれ正規分布 $N(\mu_1, \sigma^2/n_1)$, $N(\mu_2, \sigma^2/n_2)$ に従うから、 $\bar{X}_1 - \bar{X}_2$ は正規分布 $N(\mu_1 - \mu_2, \sigma^2/n_1 + \sigma^2/n_2)$ に従う。したがって、

$$U = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sigma \sqrt{1/n_1 + 1/n_2}}$$

は正規分布 $N(0, 1)$ に従い、これより U^2 は自由度 1 の χ^2 -分布に従う。

一方、定理 4.4 により $\frac{(n_1 - 1)S_1^2}{\sigma^2}$, $\frac{(n_2 - 1)S_2^2}{\sigma^2}$ はそれぞれ自由度 $n_1 - 1$, $n_2 - 1$ の χ^2 -分布に従うから、 $\chi^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{\sigma^2}$ は自由度 $n_1 + n_2 - 2$ の χ^2 -分布に従う。

以上から, 定理 4.5 を用いて,

$$\begin{aligned} T^2 &= \frac{\{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)\}^2}{n_1 S_1^2 + n_2 S_2^2} \cdot \frac{n_1 n_2 (n_1 + n_2 - 2)}{n_1 + n_2} \\ &= \frac{U^2/1}{\chi^2/(n_1 + n_2 - 2)} \end{aligned}$$

は自由度 1, $n_1 + n_2 - 2$ の F -分布に従う. ゆえに, T は自由度 $n_1 + n_2 - 2$ の t -分布に従う.

〔注〕 一般に F が自由度 1, n の F -分布に従うとき, $\sqrt{F}$ は自由度 n の t -分布に従う.

〔定理 4.8〕 二項分布に従う確率変数 X が

$$P(X=k) = {}_n C_k p^k (1-p)^{n-k} \quad (k=0, 1, 2, \dots, n)$$

のとき, 自由度が $2(n-k+1)$, $2k$ の F -分布に従う確率変数 F と

$$\lambda = \frac{k(1-p)}{(n-k+1)p}$$

である λ に対し,

$$P(X \geq k) = P(F \geq \lambda)$$

(証明) 定理 3.1 より,

$$P(X \geq k) = \frac{n!}{(k-1)!(n-k)!} \int_0^p t^{k-1} (1-t)^{n-k} dt$$

ここで, $m_1 = 2(n-k+1)$, $m_2 = 2k$ とおき, m が正の整数のとき $\Gamma(m) = (m-1)!$ であることを用いて

$$\begin{aligned} P(X \geq k) &= \frac{\Gamma\left(\frac{m_1 + m_2}{2}\right)}{\Gamma(m_1/2)\Gamma(m_2/2)} \int_0^p t^{m_2/2-1} (1-t)^{m_1/2-1} dt \\ &= \frac{1}{B(m_1/2, m_2/2)} \int_0^p t^{m_2/2-1} (1-t)^{m_1/2-1} dt \end{aligned}$$

ここで, $t = \frac{m_2}{m_1 f + m_2}$ とおけば,

$$P(X \geq k) = \frac{m_1^{m_1/2} m_2^{m_2/2}}{B(m_1/2, m_2/2)} \int_1^\infty \frac{f^{m_1/2-1}}{(m_1 f + m_2)^{(m_1+m_2)/2}} df$$

したがって, 自由度が m_1, m_2 の F -分布に従う確率変数 F に対して

$$P(X \geq k) = P(F \geq \lambda)$$

(例 5) 無作為抽出した 20 人の中に、男子の数が 15 人以上いる確率は 5% 以下といえるか。ただし全体の男子、女子の割合は等しいものとする。

(解) $n=20$, $k=15$, $p=1/2$ であるから、

$$2(n-k+1) = 12, \quad 2k = 30, \quad \lambda = 2.5$$

ゆえに、20 人中の男子の数を X 人とおけば、自由度 12, 30 の F -分布に従う確率変数 F に対し、

$$P(X \geq 15) = P(F \geq 2.5)$$

と書ける。一方、 F -分布表より

$$P(F \geq 2.092) = 0.05$$

であるから

$$P(F \geq 2.5) < 0.05$$

したがって

$$P(X \geq 15) < 0.05$$

〔定理 4.9〕 二項分布に従う確率変数 X が

$$P(X=k) = {}_n C_k p^k (1-p)^{n-k} \quad (k=0, 1, 2, \dots, n)$$

のとき、自由度が $2(k+1)$, $2(n-k)$ の F -分布に従う確率変数 F' と

$$\lambda' = \frac{(n-k)p}{(k+1)(1-p)}$$

である λ' に対し

$$P(X \leq k) = P(F' \geq \lambda')$$

$$(\text{証明}) \quad P(X \leq k) = \sum_{r=0}^k {}_n C_r p^r (1-p)^{n-r} = \sum_{r=0}^k {}_n C_{n-r} (1-p)^{n-r} p^r$$

ここで、 $n-r=x$, $1-p=q$ とおけば、

$$P(X \leq k) = \sum_{x=n-k}^n {}_n C_x q^x (1-q)^{n-x}$$

したがって、定理 4.8 により、自由度が $2(k+1)$, $2(n-k)$ の F -分布に従う確率変数 F' に対し

$$\lambda' = \frac{(n-k)(1-q)}{\{n-(n-k)+1\}q} = \frac{(n-k)p}{(k+1)(1-p)}$$

とおけば

$$P(X \leq k) = P(F' \geq \lambda')$$

(例6) 1つのサイコロを32回投げて、1の目が出る回数が2回以下である確率は5%以下といえるか。

$$(\text{解}) \quad 2(k+1) = 6, \quad 2(n-k) = 60, \quad \lambda' = \frac{30 \times 1/6}{3 \times 5/6} = 2$$

ゆえに、32回投げて1の目が出る回数を X とおけば

$$P(X \leq 2) = P(F' \geq 2)$$

ただし、 F' は自由度 6, 60 の F -分布に従う確率変数である。

一方、 F -分布表より $P(F' \geq 2.254) = 0.05$ であるから

$$P(F' \geq 2) > 0.05$$

したがって

$$P(X \leq 2) > 0.05$$

〔定理 4.10〕 無限母集団が k 個の階級 $C_1, C_2, \dots, C_k$ に分かれていて、その $C_1, C_2, \dots, C_k$ に属するものの割合がそれぞれ $p_1, p_2, \dots, p_k$ ($p_1 + p_2 + \dots + p_k = 1$) とする。このときこの母集団から大きさ n の標本を無作為抽出し、階級 C_i に属するものの個数を N_i ($i=1, 2, \dots, k$) とすれば、

$$X = \sum_{i=1}^k \frac{(N_i - np_i)^2}{np_i}$$

は n が大きいとき、近似的に自由度 $k-1$ の χ^2 -分布に従う。

(証明) $k=2$ のとき

$$X = \frac{(N_1 - np_1)^2}{np_1} + \frac{(N_2 - np_2)^2}{np_2}$$

ここで、 $N_2 = n - N_1$, $p_2 = 1 - p_1$ を代入して、

$$X = \frac{(N_1 - np_1)^2}{np_1} + \frac{(N_1 - np_1)^2}{n(1-p_1)} = \frac{(N_1 - np_1)^2}{np_1(1-p_1)}$$

一方、 n が大きいとき、定理 3.9 により N_1 は近似的に正規分布 $N(np_1, np_1(1-p_1))$ に従うから

$$\frac{N_1 - np_1}{\sqrt{np_1(1-p_1)}}$$

は近似的に正規分布 $N(0, 1)$ に従う。ゆえに X は近似的に自由度 $2-1$ の χ^2 -分布に従う。

($k > 2$ については本書の程度を越えるので証明は省略)

(例 7) 貨幣を 200 回投げて、表の出る回数が 115 回以上となる確率は 5% より大きい。 χ^2 -分布表より調べよ。

(解) 表の出る回数を N_1 、裏の出る回数を N_2 とする。一方

$$np_1 = np_2 = 200 \times \frac{1}{2} = 100$$

であるから

$$\chi^2 = \frac{(N_1 - 100)^2}{100} + \frac{(N_2 - 100)^2}{100}$$

は自由度 1 の χ^2 -分布に従う。

上式から、 $N_1 \geq 115$ のとき $\chi^2 \geq 4.5$ となるから、

$$P(N_1 \geq 115) = P(\chi^2 \geq 4.5)$$

また、 χ^2 -分布表より

$$P(\chi^2 \geq 3.84) = 0.05$$

したがって

$$P(N_1 \geq 115) < 0.05$$

第5章 母数の推定

§5.1 点推定と区間推定

標本調査の主な目的は、標本から得られた知識をもとにして、母集団の性質を推定することである。たとえば、全国の小学生の学力を知りたい場合に無作為抽出した何人かの小学生について学力を調べ、それによって全体の学力を推定する。また、ある実験において、何回か測定してそれによって結論を下したりする。

母数の推定において、点推定と区間推定とがある。これについて説明しよう。

母集団から標本を抽出して、標本から母平均、母分散を推定する場合を考える。いま標本値 $(x_1, x_2, \dots, x_n)$ を知って、母平均 μ の推定値として、

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

を用いたり、また母分散の推定値として、

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

を用いたりする。これにはそれなりの理由がある。大きさ n の標本を $(X_1, X_2, \dots, X_n)$ とすれば、標本値 $x_1, x_2, \dots, x_n$ はそれぞれ $X_1, X_2, \dots, X_n$ の実現値であるから、 $\bar{x}, s^2$ もそれぞれ標本平均 $\bar{X}$ 、標本分散 S^2 の実現値となり、かつ、

$$E(\bar{X}) = \mu \quad (\mu; \text{母平均}), \quad E(S^2) = \sigma^2 \quad (\sigma^2; \text{母分散})$$

であるからである(3.4節)。

一般に母数 θ を推定するときに用いる統計量 T を推定量といい、 T の実現値が θ の推定値である。 n 個の標本から計算される統計量 T を、 n を固定して T_n とすれば、 T_n の値は n 個の標本 $X_1, X_2, \dots, X_n$ の選び方によっていろいろ変化するが、 X_i の値のすべての場合を考えて平均したものが θ に等しくなることが望ましい。すなわち、

$$E\{T_n(X_1, X_2, \dots, X_n)\} = \theta$$

のとき、 T_n を母数 θ の不偏推定量、 T_n の実現値を不偏推定値という。したがって、 $E(\bar{X}) = \mu$ 、 $E(S^2) = \sigma^2$ であったから、標本平均 $\bar{X}$ 、標本分散 S^2 は母平均 μ 、母分散 σ^2 の不偏推定量であり、それらの実現値 $\bar{x}$ 、 s^2 は不偏推定値である。

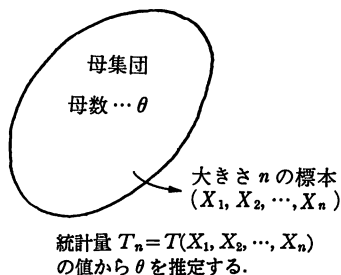


図5-1

不偏推定量によって母数を推定すること

を点推定という。点推定に対し区間推定とよばれるものがある。これについて例をあげて説明しよう。

「小学校1年生の身長は、標準偏差5 cm ぐらゐの正規分布をしていることが知られている。いま無作為抽出した16人の1年生について調べたら身長平均は113.1 cmであった」

これより、1年生全体の身長平均(母平均) μ をつぎのような手順で推定してみよう。そのために無作為抽出した16人の身長平均(標本平均)を $\bar{X}$ cm とおけば、113.1は $\bar{X}$ の実現値である。また定理3.5によって $\bar{X}$ は正規分布 $N(\mu, 5^2/16)$ 型に従うから、

$$Y = \frac{\bar{X} - \mu}{5/4}$$

は正規分布 $N(0, 1)$ 型に従う。したがって、

$$P(-\lambda < Y < \lambda) = 0.95$$

となるような λ ，すなわち，正規分布表において

$$\Phi(\lambda) = 0.975$$

であるような λ を求めると

$$\lambda = 1.96$$

であるから、

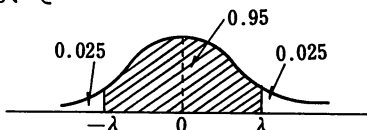


図5-2

$$P\left(-1.96 < \frac{\bar{X} - \mu}{5/4} < 1.96\right) = 0.95$$

が成り立つ。いいかえれば,

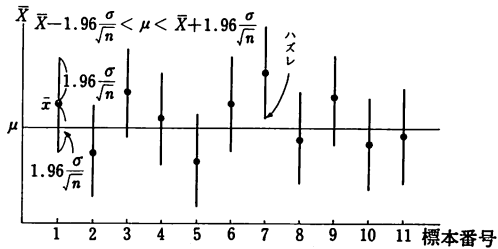
$$P(\bar{X}-2.45 < \mu < \bar{X}+2.45) = 0.95$$

が成り立つ。ここで $\bar{X}$ の代わりに $\bar{X}$ の実現値 113.1 を代入することにより, 95% の可能性で関係式

$$110.65 < \mu < 115.55$$

が成立する。このような形で母数を推定することを区間推定するという。この 95% を信頼度または信頼係数, 95% に対し $100\% - 95\% = 5\%$ を危険率または有意水準といい, 上の得られた区間 110.65~115.55 を信頼区間, その両端の限界 110.65, 115.55 を信頼限界という。

〔注〕 μ の値の 95% が区間 $\bar{X} - 1.96 \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + 1.96 \frac{\sigma}{\sqrt{n}}$ 内に含まれるということではない。 μ は定数であって, $\bar{X}$ が変数であるから標本の選び方によって $\bar{X}$ のとる値 $\bar{x}$ がいろいろ変わり, これにしたがって, 区間の位置も変わるが, これらの区間の 95% のものが μ の値をおおっているということである(図 5-3)。



(95% 以上の区間が μ をおおう)

図 5-3

§5.2 母平均の推定

大標本の場合

$N(\mu, \sigma^2)$ 型の正規母集団 (μ ; 未知) から抽出した大きさ n の標本平均 $\bar{X}$ は正規分布 $N(\mu, \sigma^2/n)$ 型に従うから

$$Y = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

とおけば, Y は正規分布 $N(0, 1)$ 型に従う. いま

$$P(-\lambda < Y < \lambda) = 1 - \alpha \quad (5.2.1)$$

なる λ を正規分布表から定めるならば,

$$P\left(-\lambda < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < \lambda\right) = 1 - \alpha$$

ゆえに

$$P\left(\bar{X} - \lambda \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + \lambda \cdot \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

したがって, 標本平均 $\bar{X}$ の実現値を $\bar{x}$ とすれば, 母平均 μ の信頼区間は, 信頼度 $1 - \alpha$ で

$$\bar{x} - \lambda \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + \lambda \cdot \frac{\sigma}{\sqrt{n}}$$

である. ここで λ は (5.2.1) より

$$\Phi(\lambda) = 1 - \frac{\alpha}{2}$$

と書ける. 今後この λ を $z_{\alpha/2}$ と書くことにすれば

$$\Phi(z_{\alpha/2}) = 1 - \frac{\alpha}{2}$$

が成り立つ. このとき母平均の信頼限界は信頼度 $1 - \alpha$ で,

$$\bar{x} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \quad (5.2.2)$$

である.

いま, 母分散が既知の正規母集団において, 母平均の信頼限界を求めたが, もし標本の大きさが大きければ(これを大標本という), どんな無限母集団であっても, 中心極限定理によって標本平均 $\bar{X}$ は近似的に正規分布 $N(\mu, \sigma^2/n)$ 型に従うから, 同様の結果が得られる. またこのとき, 母分散 σ^2 が未知であれば, σ^2 の代わりに標本分散 S^2 の実現値 s^2 で代用することができる. 以上をまとめて, 母平均の信頼限界は信頼度 $1 - \alpha$ で

正規母集団または $n \rightarrow \infty$ で, σ が既知のとき;

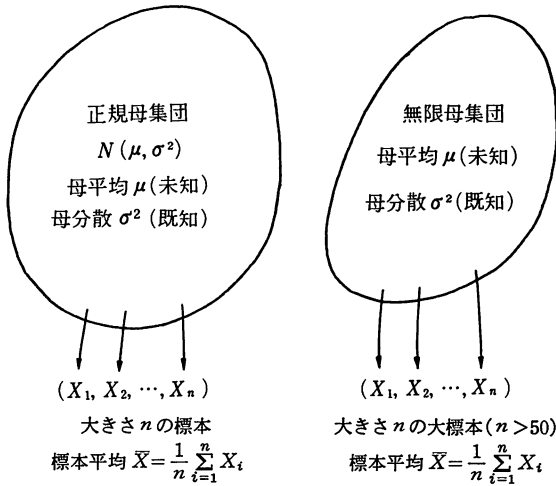


図 5-4 母平均 μ の推定

$$\bar{x} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \quad (5.2.3)$$

$n \rightarrow \infty$ で σ が未知のとき;

$$\bar{x} \pm z_{\alpha/2} \cdot \frac{s}{\sqrt{n}} \quad (5.2.4)$$

つぎに、母集団が有限母集団であれば、その母集団の大きさを N とするとき、母平均の信頼限界は

有限母集団, $n \rightarrow \infty$ で σ が既知のとき;

$$\bar{x} \pm z_{\alpha/2} \cdot \sqrt{\frac{N-n}{N-1}} \cdot \frac{\sigma}{\sqrt{n}} \quad (5.2.5)$$

有限母集団, $n \rightarrow \infty$ で σ が未知のとき;

$$\bar{x} \pm z_{\alpha/2} \cdot \sqrt{\frac{N-n}{N-1}} \cdot \frac{s}{\sqrt{n}} \quad (5.2.6)$$

となる。

(例 1) 勤労世帯 5,000 について、1 か月の実収入を調べたら平均 194,856 円、標本標準偏差 42,000 円であった。全国平均を信頼度 95% で区間推定せよ。

(解) $\bar{x}=194856$, $s=42000$, $n=5000$, $z_{\alpha/2}=1.96$

であるから、母平均 μ の信頼限界は

$$194856 \pm 1.96 \times \frac{42000}{\sqrt{5000}} \doteq 193692, 196020$$

$$\therefore 193692 < \mu < 196020 (\text{円})$$

(例 2) 1000 枚の答案の中から無作為抽出した 100 枚について調べたら、平均 65 点、標本標準偏差 30 点であった。信頼度 95% で全平均を区間推定せよ。

(解) $\bar{x}=65, s=30, N=1000, n=100, z_{\alpha/2}=1.96$ であるから、母平均 μ の信頼限界は

$$65 \pm 1.96 \times \sqrt{\frac{1000-100}{1000-1}} \times \frac{30}{\sqrt{100}} = 59.46, 70.51$$

$$\therefore 59.49 < \mu < 70.51 (\text{点})$$

小標本の場合

(a) 母分散 σ^2 が既知のとき

正規母集団 $N(\mu, \sigma^2)$ から大きさ n の標本を無作為抽出して、その標本平均 $\bar{X}$ の実現値 $\bar{x}$ を得たとする。このとき、(5.2.3) から母平均 μ の信頼限界は信頼度 $1-\alpha$ で

$$\bar{x} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \quad (5.2.7)$$

である。

(b) 母分散 σ^2 が未知のとき

正規母集団 $N(\mu, \sigma^2)$ (σ^2 ; 未知) から大きさ n の標本を無作為抽出して、その標本平均 $\bar{X}$ 、標本分散 S^2 の実現値 $\bar{x}, s^2$ を得たとする。このとき、母平均 μ を推定しよう。

定理 4.3 により $T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$ は自由度 $n-1$ の t -分布に従うから、 t -分布表を用いて $t_{n-1, \alpha/2}$ を求めれば、

$$P\left(-t_{n-1, \alpha/2} < \frac{\bar{X} - \mu}{S/\sqrt{n}} < t_{n-1, \alpha/2}\right) = 1 - \alpha$$

が成り立つ。この式を書き直して

$$P\left(\bar{X} - t_{n-1, \alpha/2} \cdot \frac{S}{\sqrt{n}} < \mu < \bar{X} + t_{n-1, \alpha/2} \cdot \frac{S}{\sqrt{n}}\right) = 1 - \alpha$$

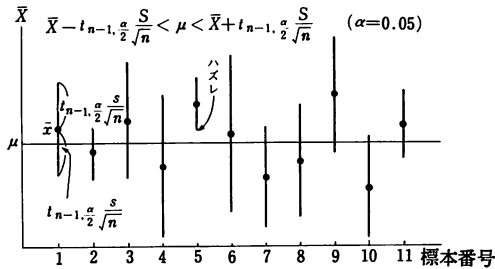
したがって、 $\bar{X}=\bar{x}$ 、 $S=s$ であるから、母平均 μ の信頼限界は信頼度 $1-\alpha$ で

$$\bar{x} \pm t_{n-1, \alpha/2} \cdot \frac{s}{\sqrt{n}} \quad (5.2.8)$$

である。

なお、この推定法は、標本数が少ない場合だけでなく、標本数が多くてもこのまま成り立つ。

〔注〕 μ の信頼区間においては、標本平均 $\bar{X}$ とともに、標本標準偏差 S の値も変動する。 n が小さければ、信頼区間の変動もはげしいが、それらのうちの95%のものが μ の値をおおうということである(図5-5)。



(区間の中心ばかりでなく区間の幅も変わる)

図5-5

(例3) 正規母集団 $N(\mu, 0.3^2)$ から大きさ4の標本を無作為抽出して、標本値

12.6, 13.4, 12.8, 13.2

を得た。これより母平均 μ を信頼度95%で区間推定せよ。

$$(解) \quad \bar{x} = \frac{52.0}{4} = 13.00, \quad \sigma = 0.3, \quad n = 4, \quad \alpha = 0.05, \quad z_{\alpha/2} = 1.96$$

であるから、 μ の信頼限界は

$$\bar{x} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} = 13.00 \pm 1.96 \times \frac{0.3}{\sqrt{4}} = 12.71, 13.29$$

$$\therefore 12.71 < \mu < 13.29$$

(例4) 正規母集団から無作為抽出した大きさ8の標本は

126.6, 126.1, 126.3, 126.7

126.0, 125.7, 125.9, 125.9

であった。これより母平均 μ を信頼度 95% で推定せよ。

(解) $\bar{x} = 126.15$, $s^2 = 0.1257$, $n = 8$, $t_{7, 0.025} = 2.365$

であるから、母平均の信頼限界は

$$126.15 \pm 2.365 \times \frac{\sqrt{0.1257}}{\sqrt{8}} = 126.15 \pm 0.30 = 126.45, 125.85$$

$$\therefore 125.85 < \mu < 126.45$$

問1 製針工場で、ある日の製品からとった大きさ 100 の標本の平均は 0.103 mm、標本標準偏差は 0.002 mm であった。信頼度 95% で母平均の信頼区間を求めよ。

問2 ある会社の従業員の数は 1000 人であった。無作為抽出した 100 人について調べたところ平均年齢は 35.4 歳、標本標準偏差は 14.2 歳であった。信頼度 95% で全従業員の平均年齢を区間推定せよ。

問3 ある工場で生産している金属板の中から 200 枚を無作為抽出して厚さ $x(\text{mm})$ を調べたところ、つぎの結果を得た。

x	6.8	7.0	7.2	7.4	7.6	7.8	8.0	8.2	8.4	8.6	8.8	計
度数	1	3	10	22	32	46	39	25	12	7	3	200

これより、厚さの母平均を信頼度 95% および 99% で推定せよ。

問4 正規母集団からの大きさ 5 の標本値が

10.5, 11.0, 11.2, 12.5, 12.8

であった。母分散 $\sigma^2 = 0.81$ を知って、この母平均 μ を信頼度 95% で推定せよ。

問5 あるメタノール溶液の含水率を 10 回測定したところ、 $\bar{x} = 0.552$ および $s = 0.037$ を得た。これよりメタノール溶液の真の含水率 μ を信頼度 90% で推定せよ。

問6 ある材料の表面かたさを測定した結果、つぎの測定値を得た。

14.2, 15.1, 13.8, 14.4, 14.9, 15.7, 14.3, 13.8

これより母平均 μ を信頼度 95% で推定せよ。

§5.3 比率の推定

大標本の場合

ある性質 E をもつものの比率が p (未知) である無限母集団から、大きさ n の標本を無作為抽出して、その中にその性質 E をもつものが k 個はいついていとき、 p の値を信頼度 $1 - \alpha$ で区間推定しよう。

無作為抽出した大きさ n の標本の中に性質 E をもつものの個数を X とすれば、 X は

$$P(X=r) = {}_nC_r p^r (1-p)^{n-r} \quad (0 \leq r \leq n)$$

をみたす確率変数であるから、 n が大きいときラプラスの定理によって近似的に正規分布 $N(np, np(1-p))$ 型に従う。ゆえに $\bar{X} = X/n$ とおけば、 $\bar{X}$ は近似的に正規分布 $N(p, \frac{p(1-p)}{n})$ 型に従う。したがって、大標本の母平均の推定の場合と同様に考えて

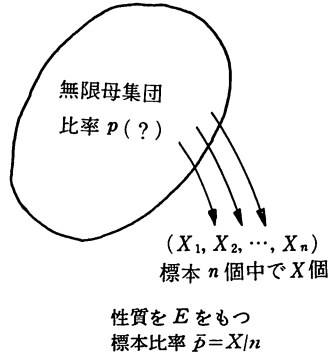


図 5-6 比率 p の推定

$$P\left(\bar{X} - z_{\alpha/2} \cdot \frac{\sqrt{p(1-p)}}{\sqrt{n}} < p < \bar{X} + z_{\alpha/2} \cdot \frac{\sqrt{p(1-p)}}{\sqrt{n}}\right) = 1 - \alpha$$

が成り立つ。ここで $\bar{X}$ の実現値は k/n であり、また $\sqrt{p(1-p)}$ の代わりに $\sqrt{k/n(1-k/n)}$ で代用すれば、 p の信頼限界は信頼度 $1 - \alpha$ で、

無限母集団で $n \doteq \infty$ のとき；

$$\frac{k}{n} \pm z_{\alpha/2} \cdot \sqrt{\frac{k}{n} \left(1 - \frac{k}{n}\right)} / \sqrt{n} \quad (5.3.1)$$

である。もし母集団が有限母集団であれば、その母集団の大きさを N とすれば

有限母集団で $n \doteq \infty$ のとき；

$$\frac{k}{n} \pm z_{\alpha/2} \cdot \sqrt{\frac{k}{n} \left(1 - \frac{k}{n}\right) \frac{N-n}{N-1}} / \sqrt{n} \quad (5.3.2)$$

となる。

(例 1) 1000 個の製品の中に 20 個の不良品がはいっていた。これより全体の不良率を信頼度 95% で区間推定せよ。

(解) $n=1000$, $k=20$, $z_{\alpha/2}=1.96$ であるから、比率 p の信頼限界は

$$\begin{aligned} \frac{20}{1000} \pm 1.96 \sqrt{\frac{20}{1000} \left(1 - \frac{20}{1000}\right)} / \sqrt{1000} &\doteq 0.0113, 0.0286 \\ \therefore 0.0113 < p < 0.0286 \end{aligned}$$

(例 2) 3600 人の学生の中から無作為抽出した 400 人中 240 人がある決定に

賛成した。これより全体の賛成者の比率を信頼度 95% で区間推定せよ。

(解) $N=3600$, $n=400$, $k=240$, $z_{\alpha/2}=1.96$ であるから, 比率 p の信頼限界は,

$$\frac{240}{400} \pm 1.96 \sqrt{\frac{240}{400} \left(1 - \frac{240}{400}\right) \cdot \frac{3600-400}{3600-1}} / \sqrt{400} \doteq 0.554, 0.644$$

$$\therefore 0.554 < p < 0.644$$

小標本の場合

ある性質 E をもつものの比率が p (未知) である無限母集団から, 大きさ n の標本を無作為抽出して, その中にその性質 E をもつものの個数を X 個とすると, いま $X=k$ であることを知って比率 p を推定しよう。

確率変数 X は二項分布に従うから, 定理 4.8 により, 自由度が $2(n-k+1)$, $2k$ の F -分布に従う確率変数 F と

$$\lambda = \frac{k(1-p)}{(n-k+1)p}$$

である λ に対して

$$P(X \geq k) = P(F \geq \lambda)$$

と書ける。また, 定理 4.9 より, 自由度が $2(k+1)$, $2(n-k)$ の F -分布に従う確率変数 F' と

$$\lambda' = \frac{(n-k)p}{(k+1)(1-p)}$$

である λ' に対して

$$P(X \leq k) = P(F' \geq \lambda')$$

と書けるから

$$P(F \geq F_0) = \frac{\alpha}{2}$$

を満足する F_0 を F -分布表より求めるならば, 信頼度 $1-\alpha/2$ として

$$\lambda = \frac{k(1-p)}{(n-k+1)p} < F_0$$

が成り立つ。これを解いて

$$\frac{k}{F_0(n-k+1)+k} < p \quad (5.3.3)$$

同様に

$$P(F' \geq F_0') = \frac{\alpha}{2}$$

を満足する F_0' を F -分布表より求めるならば, 信頼度 $1-\alpha/2$ として

$$\lambda' = \frac{(n-k)p}{(k+1)(1-p)} < F_0'$$

が成り立つ. これを解いて

$$p < \frac{(k+1)F_0'}{(n-k)+(k+1)F_0'} \quad (5.3.4)$$

したがって, (5.3.3), (5.3.4) より, 比率 p の信頼区間は信頼度 $1-\alpha$ として

$$\frac{k}{F_0(n-k+1)+k} < p < \frac{(k+1)F_0'}{(n-k)+(k+1)F_0'}$$

である. 以上をまとめると,

無限母集団から無作為抽出した大きさ n の標本の中に, 性質 E をもつものが k 個入っていたとする. このとき, その母集団の中に性質 E をもつものの比率 p の信頼区間は信頼度 $1-\alpha$ として

$$\underline{p} < p < \bar{p} \quad (5.3.5)$$

である. ただし

(1) $\underline{p}$ は $n_1=2(n-k+1)$, $n_2=2k$ として F -分布表より $F_{n_2, \alpha/2}^{n_1}$ を求めて

$$\underline{p} = \frac{n_2}{n_1 F_{n_2, \alpha/2}^{n_1} + n_2}$$

(2) $\bar{p}$ は $n_1'=2(k+1)$, $n_2'=2(n-k)$ として F -分布表より $F_{n_2', \alpha/2}^{n_1'}$ を求めて

$$\bar{p} = \frac{n_1' F_{n_2', \alpha/2}^{n_1'}}{n_1' F_{n_2', \alpha/2}^{n_1'} + n_2'}$$

である.

(例3) 金属製部品がある. 無作為抽出した15個の中の3個に, 材質上の欠点があることがわかった. この部品の不良率を信頼度95%で推定せよ.

(解) $n_1 = 2(15-3+1) = 26$, $n_2 = 2 \times 3 = 6$, $F_{6, 0.025}^{26} = 5.12$

$$\therefore \underline{p} = \frac{6}{26 \times 5.12 + 6} = 0.043$$

$$n_1' = 2(3+1) = 8, n_2' = 2(15-3) = 24, F_{24, 0.025}^8 = 2.78$$

$$\therefore \bar{p} = \frac{8 \times 2.78}{8 \times 2.78 + 24} = 0.48$$

ゆえに、求める不良率 p の信頼区間は

$$0.043 < p < 0.48$$

問1 全国から無作為抽出した 3500 人について、支持する政党を調べたところ、A 党を支持する者が 42% いた。A 党の支持率を信頼度 95% で区間推定せよ。

問2 1000 人の学生の中から無作為抽出した 100 人の学生について調べたところ、英数ともに得意とする学生が 20 人いた。これより英数ともに得意とする学生の割合を信頼度 95% で区間推定せよ。

問3 高校生 100 人について近視者を調べたら 30 人いた。これより、高校生の近視率を信頼度 95% で推定せよ。

問4 ある注射液のロットから無作為抽出した 30 本について検査した結果、4 本の不良品があった。このロットの母不良率を信頼度 95% で推定せよ。

問5 20 人の患者に、ある療法をほどこして 15 人が治癒した。治癒率を信頼度 95% で推定せよ。

§5.4 母平均差の推定

大標本の場合

2 つの正規母集団 A, B の母平均、母分散をそれぞれ μ_A, μ_B (ともに未知); σ_A^2, σ_B^2 とし、この A, B からそれぞれ大きさ n_A, n_B の標本を独立に無作為抽出し、その標本平均を $\bar{X}_A, \bar{X}_B$ とする。

このとき、 $\bar{X}_A, \bar{X}_B$ はそれぞれ正規分布 $N(\mu_A, \sigma_A^2/n_A), N(\mu_B, \sigma_B^2/n_B)$ に従うから、 $\bar{X}_A - \bar{X}_B$ も正規分布 $N(\mu_A - \mu_B, \sigma_A^2/n_A + \sigma_B^2/n_B)$ に従う (大標本ならば、任意の無限母集団でもよい)。したがって、 $\bar{X}_A, \bar{X}_B$ の実現値が $\bar{x}_A, \bar{x}_B$ であれば、(5.2.3)において、

$$\begin{aligned} & \bar{x} \text{ の代わりに } \bar{x}_A - \bar{x}_B \\ & \frac{\sigma}{\sqrt{n}} \text{ の代わりに } \sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}} \end{aligned}$$

を代入することにより、母平均差 $\mu_A - \mu_B$ の信頼限界は信頼度 $1 - \alpha$ で、

正規母集団または $n_A \doteq \infty, n_B \doteq \infty$ で、 σ_A, σ_B が既知のとき;

$$(\bar{x}_A - \bar{x}_B) \pm z_{\alpha/2} \cdot \sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}} \quad (5.4.1)$$

である。また、大標本のとき、母分散 σ_A^2, σ_B^2 が未知ならば、標本分散 S_A^2, S_B^2 の実現値で代用して、(5.4.1)はつぎのように書ける。

$n_A \doteq \infty, n_B \doteq \infty$ で σ_A^2, σ_B^2 が未知のとき；

$$(\bar{x}_A - \bar{x}_B) \pm z_{\alpha/2} \cdot \sqrt{\frac{S_A^2}{n_A} + \frac{S_B^2}{n_B}} \quad (5.4.2)$$

もし母集団が有限母集団であれば、それぞれの大きさを N_A, N_B とすると、(5.4.2)は

有限母集団で $n_A \doteq \infty, n_B \doteq \infty$ のとき；

$$(\bar{x}_A - \bar{x}_B) \pm z_{\alpha/2} \cdot \sqrt{\frac{S_A^2}{n_A} \left(\frac{N_A - n_A}{N_A - 1} \right) + \frac{S_B^2}{n_B} \left(\frac{N_B - n_B}{N_B - 1} \right)} \quad (5.4.3)$$

となる。

(例1) 2つの機械 A, B で同種の製品を作っている。 A, B の製品からそれぞれ無作為抽出した100個の標本について性能を調べたら、 A の製品は標本平均1910、標本標準偏差105、 B の製品は平均1850、標本標準偏差120であった。信頼度95%で両製品の性能の母平均差の信頼限界を求めよ。

$$\begin{aligned} (\text{解}) \quad & \bar{x}_A = 1910, \bar{x}_B = 1850, n_A = n_B = 100 \\ & s_A = 105, s_B = 120, z_{\alpha/2} = 1.96 \end{aligned}$$

であるから、両製品の性能の母平均差の信頼限界は

$$\begin{aligned} & (1910 - 1850) \pm 1.96 \cdot \sqrt{\frac{105^2}{100} + \frac{120^2}{100}} \\ & = 60 \pm 31.24 = 91.24, 28.76 \end{aligned}$$

小標本の場合

母分散の等しい2つの互いに独立な正規母集団 $N(\mu_A, \sigma^2), N(\mu_B, \sigma^2)$ (σ ; 未知) から、それぞれ大きさ n_A, n_B の標本を無作為抽出して、その標本平均 $\bar{X}_A, \bar{X}_B$ 、標本分散 S_A^2, S_B^2 の実現値 $\bar{x}_A, \bar{x}_B; s_A^2, s_B^2$ を知って、母平均差 $\mu_A - \mu_B$ を推定しよう。

定理4.7により

$$11 < 25$$

$$T = \frac{(\bar{X}_A - \bar{X}_B) - (\mu_A - \mu_B)}{\sqrt{(n_A - 1)S_A^2 + (n_B - 1)S_B^2}} \sqrt{\frac{n_A n_B (n_A + n_B - 2)}{n_A + n_B}}$$

は自由度 $n_1 + n_2 - 2$ の t -分布に従う。

したがって

$$S_w = \sqrt{\frac{(n_A + n_B)\{(n_A - 1)S_A^2 + (n_B - 1)S_B^2\}}{n_A n_B (n_A + n_B - 2)}}$$

とおけば

$$P\left(-t_{n_A+n_B-2, \alpha/2} < \frac{(\bar{X}_A - \bar{X}_B) - (\mu_A - \mu_B)}{S_w} < t_{n_A+n_B-2, \alpha/2}\right) = 1 - \alpha \quad (5.4.4)$$

が成り立つ。この式を書き直して

$$P(\bar{X}_A - \bar{X}_B - t_{n_A+n_B-2, \alpha/2} \cdot S_w < \mu_A - \mu_B < \bar{X}_A - \bar{X}_B + t_{n_A+n_B-2, \alpha/2} \cdot S_w) = 1 - \alpha$$

ところで $\bar{X}_A = \bar{x}_A$, $\bar{X}_B = \bar{x}_B$; $S_A^2 = s_A^2$, $S_B^2 = s_B^2$ であるから 母平均差 $\mu_A - \mu_B$ の信頼限界は信頼度 $1 - \alpha$ で

$$(\bar{x}_A - \bar{x}_B) \pm t_{n_A+n_B-2, \alpha/2} \cdot S_w \quad (5.4.5)$$

である。ただし、

$$S_w = \sqrt{\frac{(n_A + n_B)\{(n_A - 1)s_A^2 + (n_B - 1)s_B^2\}}{n_A n_B (n_A + n_B - 2)}}$$

(例 2) 2つの学校 A, B から、それぞれ無作為抽出した 15 人について成績を調べ下表を得た。

点数	0	20	40	60	80	100	計
A校	0	1	3	7	3	1	15
B校	0	2	3	6	4	0	15

これより 2つの学校の成績の母平均差を信頼度 95% で推定せよ。ただし、両校の母分散は等しいものとする。

(解) $n_A = n_B = 15$, $\alpha = 0.05$, $t_{28, 0.025} = 2.048$

$$\bar{x}_A = \frac{1}{15} \{20 \times 1 + 40 \times 3 + 60 \times 7 + 80 \times 3 + 100 \times 1\} = 60$$

$$\bar{x}_B = \frac{1}{15} \{20 \times 2 + 40 \times 3 + 60 \times 6 + 80 \times 4\} = 56$$

$$\begin{aligned}
14s_A^2 &= \{(20-60)^2 \times 1 + (40-60)^2 \times 3 + (60-60)^2 \times 7 \\
&\quad + (80-60)^2 \times 3 + (100-60)^2 \times 1\} \\
&= 5600 \\
14s_B^2 &= \{(20-56)^2 \times 2 + (40-56)^2 \times 3 + (60-56)^2 \times 6 + (80-56)^2 \times 4\} \\
&= 5760
\end{aligned}$$

ゆえに、求める成績の母平均差の信頼限界は

$$\begin{aligned}
(60-56) \pm 2.048 \times \sqrt{\frac{(15+15)(5600+5760)}{15 \times 15 \times (15+15-2)}} \\
\div 19.053, -11.053 \\
\therefore -11.053 < \mu_A - \mu_B < 19.053
\end{aligned}$$

(例3) ある錠剤のある成分の含有率を検出するのに2つの方法A, Bを用いた。その結果は

$$\text{方法A: } n_A = 5, \bar{x}_A = 3.298\%, s_A^2 = 0.000035$$

$$\text{方法B: } n_B = 5, \bar{x}_B = 3.225\%, s_B^2 = 0.000093$$

であった。両方法による母平均差を信頼度95%で推定せよ。ただし、その母分散はいずれの場合も等しいと仮定する。

$$(\text{解}) \quad \alpha = 0.05, t_{8, 0.025} = 2.306$$

であるから、求める母平均差 $\mu_A - \mu_B$ (%) の信頼限界は、

$$\begin{aligned}
(3.298 - 3.225) \pm 2.306 \sqrt{\frac{(5+5)(4 \times 0.000035 + 4 \times 0.000093)}{5 \times 5 \times (5+5-2)}} \\
= 0.043 \pm 0.012 = 0.031, 0.055 \\
\therefore 0.031 < \mu_A - \mu_B < 0.055(\%)
\end{aligned}$$

問1 全国の小学生に実力テストを行った。都市にある小学校から無作為抽出した500人の成績は

$$\text{平均} = 63.2 \text{ 点, 標本標準偏差} = 32 \text{ 点}$$

で、地方にある小学校から無作為抽出した300人の成績は

$$\text{平均} = 61.4 \text{ 点, 標本標準偏差} = 30 \text{ 点}$$

であった。両者の成績の平均差を信頼度95%で区間推定せよ。

問2 $N(\mu_A, 5^2)$ からの大きさ16の標本の標本平均が20.5で、 $N(\mu_B, 6^2)$ からの大きさ25の標本の標本平均が24.0であるとき、 $\mu_A - \mu_B$ を信頼度95%で推定せよ。

問3 仕事のテストを経験年数2年以上、2年以下に分けて行い、つぎの結果を得た。

$$2 \text{ 年以上; } n_A = 850, \bar{x}_A = 48.5, s_A = 5.8$$

2年以下; $n_B = 510$, $\bar{x}_B = 46.0$, $s_B = 5.9$

これより, 母平均差 $\mu_A - \mu_B$ を信頼度 95% で推定せよ.

問4 2つの工場 A, B で製造した同種の製品をそれぞれ 32 個, 30 個ずつ抽出して, その性質を調べたところ, A 工場の製品は平均 27.5, 標本分散 6.12, B 工場の製品は平均 24.1, 標本分散 8.24 であった. 両工場の製品の性能差を信頼度 95% で推定せよ. ただし, その母分散はいずれも等しいと仮定する.

問5 ある電球会社が A 型フィラメント入りの電球 10 個と B 型フィラメント入りの電球 10 個を試験し, つぎのような寿命データを得た.

A 型フィラメント; 1293, 1380, 1614, 1493, 1340,

1643, 1466, 1627, 1383, 1711

B 型フィラメント; 1061, 1065, 1092, 1017, 1021,

1138, 1143, 1094, 1270, 1029

A 型フィラメント電球と B 型フィラメント電球の平均寿命差を信頼度 95% で推定せよ. ただし $\sigma_A^2 = \sigma_B^2$ とする.

§5.5 比率差の推定

大標本の場合

ある性質 E をもつものの割合が p_A, p_B (ともに未知) である 2 つの無限母集団 A, B から, それぞれ大きさ $n_A \doteq \infty, n_B \doteq \infty$ の標本を無作為抽出し, その中に性質 E をもつものの個数を X_A, X_B とする.

$$\bar{X}_A = \frac{X_A}{n_A}, \quad \bar{X}_B = \frac{X_B}{n_B}$$

とおけば,

$$\bar{X}_A \text{ は正規分布 } N\left(p_A, \frac{p_A(1-p_A)}{n_A}\right)$$

$$\bar{X}_B \text{ は正規分布 } N\left(p_B, \frac{p_B(1-p_B)}{n_B}\right)$$

に近似するから, $\bar{X}_A - \bar{X}_B$ は正規分布

$$N\left(p_A - p_B, \frac{p_A(1-p_A)}{n_A} + \frac{p_B(1-p_B)}{n_B}\right).$$

に近似する. したがって, X_A, X_B の実現値がそれぞれ k_A, k_B ならば, $\bar{X}_A, \bar{X}_B$

の実現値は k_A/n_A , k_B/n_B となる。ゆえに、大標本の比率の推定の場合と同様の考えで

$$\frac{p_A(1-p_A)}{n_A} + \frac{p_B(1-p_B)}{n_B} \text{ の代わりに } \frac{\frac{k_A}{n_A}\left(1-\frac{k_A}{n_A}\right)}{n_A} + \frac{\frac{k_B}{n_B}\left(1-\frac{k_B}{n_B}\right)}{n_B}$$

を代用するならば、比率差 $p_A - p_B$ の信頼限界は信頼度 $1 - \alpha$ で

無限母集団で $n_A \div \infty, n_B \div \infty$ のとき;

$$\left(\frac{k_A}{n_A} - \frac{k_B}{n_B} \right) \pm z_{\alpha/2} \cdot \sqrt{\frac{\frac{k_A}{n_A}\left(1-\frac{k_A}{n_A}\right)}{n_A} + \frac{\frac{k_B}{n_B}\left(1-\frac{k_B}{n_B}\right)}{n_B}} \quad (5.5.1)$$

である。

(例1) 2つの機械 A, B で同種の製品を作っている。A, B の製品からそれぞれ大きさ 100 の標本を無作為抽出して調べたら、A の製品の中には不良品が 4 個、B の製品の中には 6 個はいていた。両製品の不良率の差を信頼度 95% として区間推定せよ。

(解) $k_A = 4, k_B = 6, n_A = n_B = 100, z_{\alpha/2} = 1.96$

であるから、両製品の不良率の差 $p_A - p_B$ の信頼限界は、

$$\left(\frac{4}{100} - \frac{6}{100} \right) \pm 1.96 \cdot \sqrt{\frac{\frac{4}{100}\left(1-\frac{4}{100}\right)}{100} + \frac{\frac{6}{100}\left(1-\frac{6}{100}\right)}{100}} = 0.0402, -0.0802$$

$$\therefore -0.0802 < p_A - p_B < 0.0402$$

問1 ある政策に賛成する者は、第1回の調査では無作為抽出した 400 人中 184 人であったが、第2回の調査では無作為抽出した 500 人中 280 人いた。これより賛成する者の比率の差の信頼限界を信頼度 95% で求めよ。

問2 ある市で1つの政策に賛成か反対かを男子 150 人、女子 200 人について聞いたところ、賛成は男子 85 人、女子は 95 人であった。これより、賛成率の男女差を信頼度 95% で推定せよ。

§5.6 母分散の推定

正規母集団 $N(\mu, \sigma^2)$ から大きさ n の標本を無作為抽出して、その標本分散 S^2 の実現値 s^2 を知り母分散 σ^2 を推定しよう。

このとき、定理 4.4 により

$$X = \frac{(n-1)S^2}{\sigma^2}$$

は自由度 $n-1$ の χ^2 -分布に従う。

ゆえに、 χ^2 -分布表より $\chi_{n-1, 1-\alpha/2}^2, \chi_{n-1, \alpha/2}^2$ を求めれば

$$P\left(\chi_{n-1, 1-\alpha/2}^2 < \frac{(n-1)S^2}{\sigma^2} < \chi_{n-1, \alpha/2}^2\right) = 1-\alpha$$

が成り立つ。この式を書き直して

$$P\left(\frac{(n-1)S^2}{\chi_{n-1, \alpha/2}^2} < \sigma^2 < \frac{(n-1)S^2}{\chi_{n-1, 1-\alpha/2}^2}\right) = 1-\alpha$$

したがって、 $S^2=s^2$ であるから、母分散 σ^2 の信頼区間は信頼度 $1-\alpha$ で

$$\frac{(n-1)s^2}{\chi_{n-1, \alpha/2}^2} < \sigma^2 < \frac{(n-1)s^2}{\chi_{n-1, 1-\alpha/2}^2} \quad (5.6.1)$$

である。

(例 1) ある溶液中のメタノール含有率を 4 回測定して、 $s=0.03$ を得た。母集団の正規性(近似的に)を仮定して母分散 σ^2 を信頼度 95% で推定せよ。

$$(解) \quad s^2 = 0.0009, \quad n-1 = 3, \quad \chi_{3, 0.025}^2 = 9.348, \quad \chi_{3, 0.975}^2 = 0.216$$

であるから、母分散 σ^2 の信頼区間は

$$\frac{3 \times 0.0009}{9.348} < \sigma^2 < \frac{3 \times 0.0009}{0.216}$$

$$\therefore 0.00029 < \sigma^2 < 0.0125$$

問 1 与えられた紡錘から無作為に選んだ 15 本の綿糸試料において、引張り強さ試験を行った。この 15 個の引張り強さの標本測定値に対する S の値は 11.2 ポンドであることがわかった。この母集団の σ^2 を信頼度 95% で推定せよ。ただし、母集団分布は近似的に正規分布に従うものと仮定する。

問 2 金属板の一边の長さを 8 回測って、つぎのような値を得た。

$$19.8, 20.0, 20.2, 19.9, 20.7, 20.2, 19.4, 20.5 \quad (\text{mm})$$

真の長さ、母分散(測定のバラツキ程度)を 95% の信頼度で推定せよ。

第6章 仮説の検定

標本調査の目的の中に、母数の推定と並んで仮説の検定がある。

われわれの周囲を見たとき、子供達の体格は年々向上している。このことから、わが国の子供の体格は年々向上していると判断してよいだろうか。このような問題を統計学的に調べるのがこの章の問題である。

§6.1 仮説の検定

いろいろの場合についての仮説の検定があるが、いま具体的に1つの例を取り上げて考えてみよう。

「全国一斉に実力テストを行った結果は平均 61.2 点、標準偏差 13.4 点である。ある都市から無作為抽出した 144 人の成績は 63.5 点であった。この都市の生徒の成績は全国平均より高いといえるだろうか」……

これはこの都市の生徒の成績の母平均が、61.2 より大きいかどうかを考える問題である。

まず、この都市の生徒の成績の母集団も、全国の生徒の成績の母集団も同質のものであると仮定してみよう。いいかえれば、この都市の生徒の成績の母平均を μ 、母分散を σ^2 とするとき、

$$[\mu=61.2, \sigma^2=13.4^2]$$

であると仮定してみよう。すると仮定から、この母集団から無作為抽出した大きさ 144 の標本の標本平均 $\bar{X}$ は、近似的に

$$\text{正規分布 } N\left(61.2, \frac{13.4^2}{144}\right)$$

に従い、実際に標本を抽出してみたら $\bar{X}=63.5$ であったということになる。

ところで、 $P(\bar{X} \geq 63.5)$ を計算してみると

$$\begin{aligned} P(\bar{X} \geq 63.5) &= 1 - F(63.5) = 1 - \Phi\left(\frac{63.5 - 61.2}{13.4/12}\right) \\ &\doteq 1 - \Phi(2.06) \doteq 0.02 \end{aligned}$$

ここで、0.02は非常に小さい数である。これはこの都市の成績の母集団を全国の母集団と同質とみなした場合、 $\bar{X}=63.5$ であることは非常にめずらしい現象が起ったということになる。このようなとき、この仮定は正しくないと考えてすてることにする。すなわち、この都市の成績の母集団は全国の母集団と異質のものであると考える。したがって、問題の意味から、この都市の成績は全国の成績より高いと考える。

確率 $P(\bar{X} \geq 63.5)$ が非常に小さいかどうかは何を規準にしたらよいか。一般に 0.05, 0.025 等を用い、これを危険率または有意水準という。また、すてることを予想して設定した仮定を帰無仮説または単に仮説という。仮説をすてる場合、仮説が正しいにもかかわらずすててしまったという誤りをおかす可能性がある。この誤りを第1種の誤りという。

先の例で、144人の平均を 62.5 点とすると、同じ仮説のもとで

$$\begin{aligned} P(\bar{X} \geq 62.5) &= 1 - F(62.5) = 1 - \Phi\left(\frac{62.5 - 61.2}{13.4/12}\right) \\ &\doteq 1 - \Phi(1.16) = 0.1230 \end{aligned}$$

となり、危険率 0.05 で仮説はすてられないことになる。この場合、仮説が誤りであるにもかかわらずすてられなかったという誤りをおかす可能性がある。この誤りを第2種の誤りという。

§6.2 母平均の検定

σ が既知の正規母集団、または任意の無限母集団で大標本の場合

無限母集団から大きさ n の標本を無作為抽出して標本平均 $\bar{X}$ の実現値 $\bar{x}$ を知り、

仮説「 $\mu = \mu_0$ 」

をたてる。このとき、この無限母集団が正規母集団か、または n が大きければ ($n \geq 36$)、定理 3.5 か、または定理 3.8 により $\bar{X}$ は正規分布 $N(\mu_0, \sigma^2/n)$ に従うから、 $\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$ は正規分布 $N(0, 1)$ に従う。ここで、 σ^2 はこの母集団の母分散

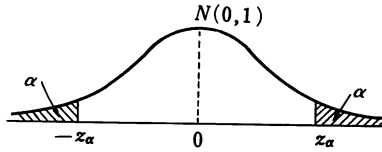


図 6-1

である。

ところで、 $\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$ の実現値は $\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$ であるから、危険率 α で

$$\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} > z_\alpha \quad \text{または} \quad \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} < -z_\alpha$$

のとき仮説をすてる。すなわち

$$\frac{|\bar{x}-\mu_0|}{\sigma/\sqrt{n}} > z_\alpha$$

のとき仮説をすてる。

以上のような検定を片側検定といい、危険率 α を片側危険率という。これに対し、危険率 α で

$$\frac{|\bar{x}-\mu_0|}{\sigma/\sqrt{n}} > z_{\alpha/2} \quad (6.2.1)$$

のとき仮説をすてることにする場合がある。

これを両側検定といい、 α を両側危険率という。片側危険率 $\alpha/2$ で検定することと、両側危険率 α で検定することは、結果として同じことであり、危険率は統計する者が

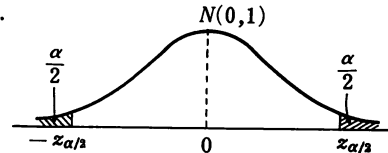


図 6-2

問題の内容に応じて決めるものであるから、今後特に断わらない限り両側検定を行うことにする。したがって、(6.2.1)によって検定を行う。

〔注〕 σ が未知で n が大きいとき、 σ の代わりに標本標準偏差 S の実現値 s を用いる。

σ が未知の正規母集団で小標本の場合

正規母集団から大きさ n の標本を無作為抽出して、標本平均 $\bar{X}$ 、標本標準偏差 S の実現値 $\bar{x}, s$ を知り、

仮説「 $\mu = \mu_0$ 」

をたてる。このとき、定理 4.3 により $\frac{\bar{X} - \mu_0}{S/\sqrt{n}}$ は自由度 $n-1$ の t -分布をする。

ところで、 $\frac{\bar{X} - \mu_0}{S/\sqrt{n}}$ の実現値は

$$\frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

であるから、危険率 α で、

$$\frac{\bar{x} - \mu_0}{s/\sqrt{n}} > t_{n-1, \alpha/2} \quad \text{または} \quad \frac{\bar{x} - \mu_0}{s/\sqrt{n}} < -t_{n-1, \alpha/2}$$

のとき仮説をすてる。すなわち、

$$\frac{|\bar{x} - \mu_0|}{s/\sqrt{n}} > t_{n-1, \alpha/2} \quad (6.2.2)$$

のとき仮説をすてる。

(例 1) ある機械による 1 日の製糸工程から同じ長さの糸 16 本を無作為抽出して、引張り強さを測定したところ、標本平均は 962 *psi* であった。これより全平均は 1000 *psi* を否定できるか。ただし、糸の引張り強さは正規分布に従い、標準偏差は 128 *psi* であることが経験的にわかっているものと仮定する (危険率 5%)。

(解) 帰無仮説「 $\mu = 1000$ 」をたてると

$$\bar{x} = 967, \mu_0 = 1000, \sigma = 128, n = 16, z_{0.025} = 1.96$$

$$\frac{|\bar{x} - \mu_0|}{\sigma/\sqrt{n}} = \frac{|967 - 1000|}{128/\sqrt{16}} \doteq 1.0 < z_{0.025}$$

ゆえに仮説はすてられない。すなわち、 $\mu = 1000$ を否定できない。

(例 2) ある工場で作っている電球の中から 36 個を無作為抽出して寿命時間を調べたところ、つぎの結果を得た。会社側ではこの電球の寿命時間の全平均は 1500 時間と公表しているが、これを否定できるか (危険率 5%)。

標本平均 = 1450 時間, 標本標準偏差 = 110 時間

(解)

仮説「 $\mu = 1500$ 」

をたてると

$$\bar{x} = 1450, \mu_0 = 1500, \sigma \doteq s = 110, n = 36, z_{0.025} = 1.96$$

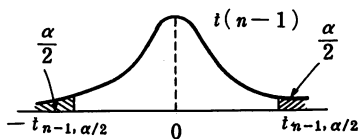


図 6-3

$$\frac{|\bar{x} - \mu_0|}{\sigma\sqrt{n}} = \frac{|1450 - 1500|}{110/6} = 2.73 > z_{0.025}$$

ゆえに、仮説はすてられる。すなわち、会社側の公表は否定される。

(例3) ある自動車メーカーの小型車の燃料効率について、メーカーでは標準状態において、1リットルあたり12km走行できると公表している。そこで9台の車について標準状態で走行テストを行ってみたところ、つぎの結果を得た。測定値の正規性を仮定してメーカーの公表を否定できるか(危険率5%)。

標本平均 = 11.8 km, 標本標準偏差 = 0.3 km

(解) 仮説「 $\mu = 12$ 」

をたてると

$$\bar{x} = 11.8, \mu_0 = 12, s = 0.3, n = 9, t_{8, 0.025} = 2.3060$$

$$\frac{|\bar{x} - \mu_0|}{s\sqrt{n}} = \frac{|11.8 - 12|}{0.3/3} = 2 < t_{8, 0.025}$$

ゆえに、仮説はすてられない。すなわち、メーカーの公表を否定することはできない。

問1 小学校新入生男子の身長在全国平均は114.1cm, 標準偏差は5cmである。ある都市で無作為抽出した新入生50人の平均身長を求めたところ115.6cmであった。この都市の平均身長は全国平均と異なるといえるか(危険率5%)。

問2 ある工場で作っている製品の性能は平均3.00である。ある日の製品の中から無作為抽出した100個の性能の平均は2.84, 標本標準偏差は0.89であった。この日の製品は普段と異なるといえるか(危険率5%)。

問3 ある工場で作っている蛍光灯の中から10本を無作為抽出して寿命時間を調べたところつぎの結果を得た。寿命時間は正規分布に従うものと仮定して、会社側で公表している寿命時間の全平均は1800時間であることを否定できるか(危険率1%)。

標本平均 = 1785時間, 標本標準偏差 = 11.5時間

問4 12オンスの表示のあるジャムかんを無作為に9個えらんで重さを測ったところつぎの結果を得た。このジャムかんの重さは正規分布に従うものと仮定して、重さの全平均は12オンスより少ないといえるか(危険率5%)。

標本平均 = 11.9オンス, 標本標準偏差 = 0.2オンス

問5 正規母集団から、大きさ5の標本を無作為抽出してつぎの結果を得た。母平均は3.26であることを否定できるか(危険率1%)。

3.25, 3.24, 3.25, 3.26, 3.25

§6.3 比率の検定

大標本の場合

性質 E をもつものを含む無限母集団から大きさ $n(n \geq 100)$ の標本を無作為抽出して、その中に性質 E をもつものが k 個入っていることを知り、

仮説「性質 E をもつものの比率は p である」

をたてる。このとき、大きさ n の標本の中の性質 E をもつものの個数 X は、定理 3.9 により近似的に正規分布 $N(np, npq)$ ($q=1-p$) に従うから、 X/n は正規分布 $N(p, pq/n)$ 、すなわち、 $\frac{X/n-p}{\sqrt{pq/n}}$ は正規分布 $N(0, 1)$ に従う。

ところで、 $\frac{X/n-p}{\sqrt{pq/n}}$ の実現値は $\frac{k/n-p}{\sqrt{pq/n}}$ であるから、危険率 α で、

$$\frac{k/n-p}{\sqrt{pq/n}} > z_{\alpha/2} \quad \text{または} \quad \frac{k/n-p}{\sqrt{pq/n}} < -z_{\alpha/2}$$

のとき仮説をすてる。すなわち

$$\frac{|k/n-p|}{\sqrt{pq/n}} > z_{\alpha/2} \quad (6.3.1)$$

のとき仮説をすてる。

小標本の場合

標本の大きさ n が大きくないときは、公式 (6.3.1) を用いることはできない。この場合にも、標本の中の性質 E をもつものの個数 X の実現値 k を知って

仮説「性質 E をもつものの比率は p である」

をたてると、このとき定理 4.8、定理 4.9 により

自由度が $2(n-k+1)$ 、 $2k$ の F -分布に従う確率変数 F と、 $\lambda = \frac{k(1-p)}{(n-k+1)p}$ である λ に対して

$$P(X \geq k) = P(F \geq \lambda)$$

が成り立ち、また、自由度が $2(k+1)$ 、 $2(n-k)$ の F -分布に従う確率変数 F' と $\lambda' = \frac{(n-k)p}{(k+1)(1-p)}$ である λ' に対して

$$P(X \leq k) = P(F' \geq \lambda')$$

が成り立つ。したがって、危険率 α で

(1) $k > np$ のとき.

$$P(X \geq k) = F(F \geq \lambda) < \frac{\alpha}{2}$$

ならば仮説をすてるから

$$\frac{n_2(1-p)}{n_1 p} > F_{n_2, \alpha/2}^{\quad n_1} \quad (n_1 = 2(n-k+1), \quad n_2 = 2k) \quad (6.3.2)$$

のとき仮説をすてることになる.

(2) $k < np$ のとき:

$$P(X \leq k) = P(F' \geq \lambda') < \frac{\alpha}{2}$$

ならば仮説をすてるから

$$\frac{n_2' p}{n_1'(1-p)} > F_{n_2', \alpha/2}^{\quad n_1'} \quad (n_1' = 2(k+1), \quad n_2' = 2(n-k)) \quad (6.3.3)$$

のとき仮説をすてることになる.

(例1) ある商品について, 買いたいと思っているかどうかを無作為抽出した 400 人について調べたところ, 買いたいと思っている人が 118 人いた. 買いたいと思っている人が全体の 25% 以上いるといえるか(危険率 5%).

(解) 仮説「買いたいと思っている人の比率は 0.25」をたてると

$$p = 0.25, \quad q = 0.75, \quad n = 400, \quad k = 118, \quad z_{0.025} = 1.96$$

$$\frac{|k/n - p|}{\sqrt{pq/n}} = \frac{|118/400 - 0.25|}{\sqrt{0.25 \times 0.75 \div 400}} \doteq 2.08 > z_{0.025}$$

ゆえに, 仮説はすてられる. すなわち, 買いたいと思っている人が全体の 25% 以上いるといえる.

[注] p を 0.25 より小さくすれば, 上式の分子は大きく, 分母は小さくなるのでなおさら $z_{0.025}$ より大きくなり仮説はすてられる.

(例2) ある団体の会員の中から 20 人を無作為抽出して調べたところ男子会員が 15 人いた. この団体は男子会員が半数以上いるといえるか(危険率 5%).

(解) 仮説「男子会員の比率は 0.5」をたてると

$$p = 0.5, \quad n = 20, \quad k = 15, \quad k > np$$

したがって,

$$n_1 = 2(20 - 15 + 1) = 12, \quad n_2 = 2 \times 15 = 30, \quad F_{30, 0.025}^{12} = 2.412$$

$$\frac{n_2(1-p)}{n_1 p} = \frac{30(1-0.5)}{12 \times 0.5} = 2.5 > F_{30, 0.025}^{12}$$

ゆえに、仮説はすてられる。すなわち、男子会員が半数以上いるといえる。

(例 3) 過去の実験ではある製品を製作するのに 5% の不良品が出た。ある工員が製作した製品 40 個については不良品が 1 個であった。この工員の技術は秀れているといえるか(危険率 10%)。

(解) 仮説「この工員の製作した製品の不良率は 0.05」をたてると

$$p = 0.05, n = 40, k = 1, k < np$$

したがって、

$$n_1' = 2(1+1) = 4, n_2' = 2(40-1) = 78, 2.447 < F_{78, 0.05}^4 < 2.525$$

$$\frac{n_2' p}{n_1'(1-p)} = \frac{78 \times 0.05}{4(1-0.05)} \doteq 1.03 < F_{78, 0.05}^4$$

ゆえに、仮説はすてられない。すなわち、この工員の技術は秀れているとはいえない。

問 1 無作為抽出した 900 人についてアンケートをとったところ、A 案に賛成するものが 490 人いた。A 案に賛成するものは半数以上いるといえるか(危険率 5%)。

問 2 ある地方で、400 人の出生児中 212 人が男子であった。この地方は男子と女子の出生率が異なるといえるか(危険率 5%)。

問 3 ある地区で無作為抽出した 20 人中 6 人が十二指腸虫保菌者であった。この地区は寄生率 20% 以上といえるか。また、50% 以下といえるか(危険率 5%)。

問 4 ある選挙区で選挙人名簿の中から 30 人を無作為抽出して面接調査したところ 9 人がある候補者を支持した。この選挙区でこの候補者の支持率は 40% 以下といえるか。また、20% 以上といえるか(危険率 5%)。

§ 6.4 母分散の検定

正規母集団から大きさ n の標本を無作為抽出し、その標本分散 $S^2 = s^2$ を知って母分散 σ^2 の検定を考えよう。

帰無仮説「 $\sigma^2 = \sigma_0^2$ 」

をたてたとき、定理 4.4 により $\frac{(n-1)S^2}{\sigma_0^2}$ は自由度 $n-1$ の χ^2 -分布に従う。

ところで、 $\frac{(n-1)S^2}{\sigma_0^2}$ の実現値は $\frac{(n-1)s^2}{\sigma_0^2}$ であるから危険率 α で

$$\frac{(n-1)s^2}{\sigma_0^2} > \chi_{n-1, \alpha/2}^2 \quad \text{または} \quad \frac{(n-1)s^2}{\sigma_0^2} < \chi_{n-1, 1-\alpha/2}^2 \quad (6.4.1)$$

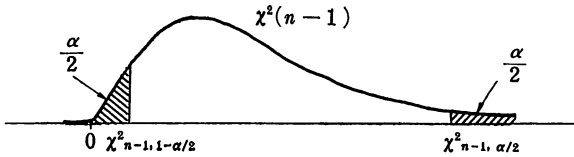


図 6-4

のとき仮説をすてる。

(例 1) 正規母集団から大きさ 11 の標本を無作為抽出して調べたところ、標本分散は 154.6 であった。これより母分散 σ^2 は 140 であるを否定できるか(危険率 5%)。

(解) 仮説「 $\sigma^2=140$ 」をたてると

$$s^2 = 154.6, \sigma_0^2 = 140, n = 11, \chi_{10, 0.025}^2 = 20.5, \chi_{10, 0.975}^2 = 3.25$$

$$\frac{(n-1)s^2}{\sigma_0^2} = \frac{10 \times 154.6}{140} = 11.04$$

$$\therefore \chi_{10, 0.975}^2 < \frac{(n-1)s^2}{\sigma_0^2} < \chi_{10, 0.025}^2$$

ゆえに、仮説はすてられない。すなわち、母分散は 140 であるを否定できない。

問 1 ある薬品の A 成分の含有率は、従来の経験から分散 0.42% の正規分布に従っていることが知られている。いま製造方法を変えて、大きさ 10 の標本について調べたところ標本分散は 0.64% であった。母分散に変化が生じたか(危険率 5%)。

問 2 ある陶器を作っている工場で、熟練工の作った陶器の直径は過去の実績から標準偏差 0.075 mm の正規分布に従うことが知られている。最近雇い入れた新入工員について、12 個の製品の標本標準偏差を調べたところ 0.082 mm であった。バラツキに差が認められるか(危険率 5%)。

§ 6.5 等平均の検定 ($\mu_A = \mu_B$ の検定)

σ が既知の正規母集団、または任意の無限母集団で大標本の場合

母平均が μ_A (未知)、母分散が σ_A^2 である無限母集団 A と、 μ_B (未知)、 σ_B^2 である無限母集団 B からそれぞれ大きさ n_A, n_B の標本を無作為抽出して、標本平均 $\bar{X}_A, \bar{X}_B$ の実現値 $\bar{x}_A, \bar{x}_B$ を知り、

$$\text{仮説「}\mu_A = \mu_B (= \mu)\text{」}$$

をたてる。このとき、無限母集団 A, B が正規母集団か、または n_A, n_B が大き

ければ ($n_A \geq 36, n_B \geq 36$), 定理 3.5 か, または定理 3.8 により, $\bar{X}_A, \bar{X}_B$ はそれぞれ正規分布 $N(\mu, \sigma_A^2/n_A)$, $N(\mu, \sigma_B^2/n_B)$ に従うから, $\bar{X}_A - \bar{X}_B$ は正規分布 $N(0, \sigma_A^2/n_A + \sigma_B^2/n_B)$, すなわち, $\frac{\bar{X}_A - \bar{X}_B}{\sqrt{\sigma_A^2/n_A + \sigma_B^2/n_B}}$ は正規分布 $N(0, 1)$ に従う.

ところで, $\frac{\bar{X}_A - \bar{X}_B}{\sqrt{\sigma_A^2/n_A + \sigma_B^2/n_B}}$ の実現値は $\frac{\bar{x}_A - \bar{x}_B}{\sqrt{\sigma_A^2/n_A + \sigma_B^2/n_B}}$ であるから, 危険率 α で

$$\frac{\bar{x}_A - \bar{x}_B}{\sqrt{\sigma_A^2/n_A + \sigma_B^2/n_B}} > z_{\alpha/2} \quad \text{または} \quad \frac{\bar{x}_A - \bar{x}_B}{\sqrt{\sigma_A^2/n_A + \sigma_B^2/n_B}} < -z_{\alpha/2}$$

のとき仮説をすてる. すなわち

$$\frac{|\bar{x}_A - \bar{x}_B|}{\sqrt{\sigma_A^2/n_A + \sigma_B^2/n_B}} > z_{\alpha/2} \quad (6.5.1)$$

のとき仮説をすてる.

[注] σ_A, σ_B が未知のとき, n_A, n_B が大きければ標本標準偏差 S_A, S_B の実現値 s_A, s_B で代用する.

σ が未知の正規母集団で小標本の場合

母平均が μ_A, μ_B (ともに未知) で, 母分散が等しいと考えられる 2 つの正規母集団 A, B からそれぞれ大きさ n_A, n_B の標本を無作為抽出して, 標本平均 $\bar{X}_A, \bar{X}_B$, 標本分散 S_A^2, S_B^2 の実現値 $\bar{x}_A, \bar{x}_B, s_A^2, s_B^2$ を知り,

仮説「 $\mu_A = \mu_B$ 」

をたてる. このとき

$$S_w = \sqrt{\frac{(n_A + n_B) \{ (n_A - 1) S_A^2 + (n_B - 1) S_B^2 \}}{n_A n_B (n_A + n_B - 2)}}$$

とおけば, 定理 4.7 により, $\frac{\bar{X}_A - \bar{X}_B}{S_w}$ は自由度 $n_A + n_B - 2$ の t -分布に従う.

ここで, $s_w = \sqrt{\frac{(n_A + n_B) \{ (n_A - 1) s_A^2 + (n_B - 1) s_B^2 \}}{n_A n_B (n_A + n_B - 2)}}$ とおけば, $\frac{\bar{X}_A - \bar{X}_B}{S_w}$ の実現値は $\frac{\bar{x}_A - \bar{x}_B}{s_w}$ であるから, 危険率 α で

$$\frac{\bar{x}_A - \bar{x}_B}{s_w} > t_{n_A + n_B - 2, \alpha/2} \quad \text{または} \quad \frac{\bar{x}_A - \bar{x}_B}{s_w} < -t_{n_A + n_B - 2, \alpha/2}$$

のとき仮説をすてる. すなわち,

$$\frac{|\bar{x}_A - \bar{x}_B|}{s_w} > t_{n_A+n_B-2, \alpha/2} \quad (6.5.2)$$

のとき仮説をすてる。

(例1) 2つの工場AとBで同一の製品を作っている。それぞれの工場の製品の中から100個ずつ無作為抽出して性能を調べたところつぎの結果を得た。両工場の製品に性能差が認められるか(危険率5%)。

A工場；標本平均=26.4，標本標準偏差=2.6

B工場；標本平均=25.6，標本標準偏差=2.4

(解) $\bar{x}_A = 26.4$, $\bar{x}_B = 25.6$, $\sigma_A \doteq s_A = 2.6$, $\sigma_B \doteq s_B = 2.4$, $n_A = n_B = 100$,
 $z_{0.025} = 1.96$

仮説「母平均は等しい」

をたてて

$$\frac{|\bar{x}_A - \bar{x}_B|}{\sqrt{\sigma_A^2/n_A + \sigma_B^2/n_B}} \doteq \frac{26.4 - 25.6}{\sqrt{2.6^2/100 + 2.4^2/100}} \doteq 2.26 > z_{0.025}$$

ゆえに，仮説はすてられる。すなわち，差異が認められる。

(例2) 中学生に同一問題で一斉テストを行った。A, Bの2つの都市からそれぞれ21人ずつ抽出して成績を調べたところつぎの結果を得た。両都市の中学生の成績に差異が認められるか(危険率5%)。ただし，成績はともに正規分布に従い，母分散は等しいものと仮定する。

A；標本平均=63.5，標本標準偏差=11.4

B；標本平均=61.2，標本標準偏差=10.6

(解) $\bar{x}_A = 63.5$, $\bar{x}_B = 61.2$, $s_A = 11.4$, $s_B = 10.6$, $n_A = n_B = 21$,
 $t_{40, 0.025} = 2.0211$

$$\therefore s_w = \sqrt{\frac{(21+21)\{20 \times 11.4^2 + 20 \times 10.6^2\}}{21 \times 21 \times (21+21-2)}} \doteq 3.397$$

仮説「母平均は等しい」

をたてて

$$\frac{|\bar{x}_A - \bar{x}_B|}{s_w} = \frac{|63.5 - 61.2|}{3.397} \doteq 0.677 < t_{40, 0.025}$$

ゆえに，仮説はすてられない。すなわち，両都市の中学生の成績に差異があるとはいえない。

[注] 母分散が等しいかどうか分からないときは §6.6 例 2 を参照。

問 1 2つの機械 A, B で、370 g 入りの缶詰を作っている。それぞれの製品の中から 100 個ずつ無作為抽出して容量を調べつぎの結果を得た。 A の製品の母平均と B の製品の母平均に差異があるか(危険率 5%, 1%)。

A ; 標本平均 = 371.2 g, 標本標準偏差 = 3.8 g

B ; 標本平均 = 370.1 g, 標本標準偏差 = 3.4 g

問 2 2つの地区 A, B からそれぞれ無作為抽出した 100 名の高校生に同一問題でテストを行いつぎの結果を得た。母平均に差異があるか(危険率 5%)。

A ; 標本平均 = 62 点, 標本標準偏差 = 6.2 点

B ; 標本平均 = 58 点, 標本標準偏差 = 5.8 点

問 3 午前の製品の中から 20 個, 午後の製品の中から 30 個を無作為抽出して強さを調べたところつぎの結果を得た。午前と午後の製品の強さに差が認められるか(危険率 5%)。

ただし、製品の強さはともに正規分布に従い、母分散は等しいものと仮定する。

午前 ; 標本平均 = 24.6, 標本標準偏差 = 3.2

午後 ; 標本平均 = 27.4, 標本標準偏差 = 2.6

問 4 同一規格でメーカーの異なる A, B 2種の電球について、それぞれ 8 個を無作為抽出して、その寿命を試験したところつぎの結果を得た。平均寿命時間に差があるか(危険率 5%)。

ただし、寿命時間は正規分布に従い、母分散は等しいものと仮定する。

A : 1260, 1200, 1330, 1310, 1260, 1250, 1350, 1280

B : 1320, 1250, 1400, 1370, 1280, 1330, 1280, 1370

§6.6 等分散の検定

母分散 σ_A^2, σ_B^2 が未知の 2つの正規母集団 A, B から、大きさ n_A, n_B の標本を無作為抽出して標本分散 S_A^2, S_B^2 の実

現値 s_A^2, s_B^2 を知り、

仮説「 $\sigma_A^2 = \sigma_B^2$ 」

をたてる。このとき、定理 4.6 により、 S_A^2/S_B^2 は自由度 n_A-1, n_B-1 の F -分布に従う。

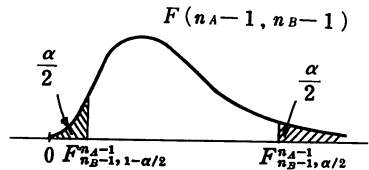


図 6-5

ところで, S_A^2/S_B^2 の実現値は s_A^2/s_B^2 であるから, 危険率 α で,

$$\frac{s_A^2}{s_B^2} > F_{n_B-1, \alpha/2}^{n_A-1} \quad \text{または} \quad \frac{s_A^2}{s_B^2} < F_{n_B-1, 1-\alpha/2}^{n_A-1} \quad (6.6.1)$$

のとき仮説をすてる.

ここで, あらかじめ s_A^2, s_B^2 のうち大きい方が分子にくるように順序を変えておけば, 図 6-5 の右側だけを考えればよいことになるので, (6.6.1) はつぎのようになる.

危険率 α で

$$(1) \quad s_A^2 > s_B^2 \text{ ならば, } \frac{s_A^2}{s_B^2} > F_{n_B-1, \alpha/2}^{n_A-1} \quad (6.6.2)$$

のとき仮説をすてる.

$$(2) \quad s_A^2 < s_B^2 \text{ ならば, } \frac{s_B^2}{s_A^2} > F_{n_A-1, \alpha/2}^{n_B-1} \quad (6.6.3)$$

のとき仮説をすてる.

(例 1) ある陶器を作っている工場で, 熟練工の作った陶器の中から 20 個, 新入工具の作った陶器の中から 25 個を無作為抽出して直径を測ったところ, それぞれの標本標準偏差は

熟練工 ; 0.07 mm, 新入工具 ; 0.11 mm

であった. 正規性を仮定して直径のバラツキに差異があるか (危険率 5%).

(解) 熟練工の製品を A , 新入工具の製品を B で表せば, $s_A=0.07$, $s_B=0.11$, $s_A < s_B$, $n_A=20$, $n_B=25$, $F_{19, 0.025}^{24}=2.452$

仮説「母分散は等しい」をたてて

$$\frac{s_B^2}{s_A^2} = \frac{0.11^2}{0.07^2} \doteq 2.496 > F_{19, 0.025}^{24}$$

ゆえに, 仮説はすてられる. すなわち, バラツキに差異がある.

(例 2) 男女からなるクラスで数学のテストを行い右の表を得た. これを標本と考えて男女の成績に差異があるといえるか (危険率 5%). ただし, 成績は男女とも正規分布に従うものと仮定する.

	人数	標本平均	標本分散
男(A)	25	63	140
女(B)	21	60	120

(解) $\bar{x}_A = 63$, $\bar{x}_B = 60$, $s_A^2 = 140$, $s_B^2 = 120$, $n_A = 25$, $n_B = 21$

この問題はつぎの2つの段階に分けて検定する。

(i) 仮説「母分散は等しい」

をたてて

$$\frac{s_A^2}{s_B^2} = \frac{140}{120} = 1.1\bar{6} < F_{20, 0.025}^{24} = 2.408$$

ゆえに、仮説はすてられない。したがって、男子、女子の成績の母分散は等しいとみなしてつぎの検定に入る。

(ii) 仮説「母平均は等しい」

をたてて

$$s_w = \sqrt{\frac{(25+21)\{24 \times 140 + 20 \times 120\}}{25 \times 21 \times (25+21-2)}} \doteq 3.387$$

$2.0003 < t_{44, 0.025} < 2.0211$ であるから

$$\frac{|\bar{x}_A - \bar{x}_B|}{s_w} = \frac{|63 - 60|}{3.387} \doteq 0.886 < t_{44, 0.025}$$

ゆえに、仮説はすてられない。すなわち、男女の成績に差異があるとはいえない。

[注] 上の問題のような場合に、(i)の検定を第1次検定、(ii)の検定を第2次検定という。

問1 中学1年生の中から無作為抽出した16人に対して、夏休み前と後に同程度の数学のテストを行いつぎの結果を得た。この結果はそれぞれ正規母集団からの標本であると仮定して、夏休み後に成績は落ちたといえるか(危険率5%)。

夏休み前；標本平均=65.2点，標本標準偏差=13.4点

夏休み後；標本平均=61.0点，標本標準偏差=12.6点

問2 A, B 2人の店員による1つの商品の包装時間を調べてつぎの結果を得た。包装時間の正規性を仮定して、両者の平均の間に差が認められるか(危険率5%)。

A : 38, 39, 37, 38, 31, 40, 36

B : 33, 41, 28, 39, 34

§6.7 $p_A = p_B$ の検定

性質 E をもつものを含む比率が p_A, p_B (ともに未知) である2つの無限母集団 A, B からそれぞれ大きさ n_A, n_B の標本を無作為抽出して、その中に含まれる性質 E をもつものの個数 X_A, X_B の実現値 k_A, k_B を知り、

仮説「 $p_A = p_B (=p)$ 」

をたてる。このとき n_A, n_B が大きければ ($n_A \geq 100, n_B \geq 100$), 定理 3.9 により, X_A, X_B はそれぞれ正規分布 $N(n_A p, n_A p q), N(n_B p, n_B p q)$ ($q=1-p$) に従うから, $X_A/n_A - X_B/n_B$ は正規分布 $N(0, (1/n_A + 1/n_B)pq)$, すなわち, $\frac{X_A/n_A - X_B/n_B}{\sqrt{(1/n_A + 1/n_B)pq}}$ は正規分布 $N(0, 1)$ に従う。

ここで, pq の値を近似的に $p_0(1-p_0)$ (ただし, $p_0 = \frac{k_A + k_B}{n_A + n_B}$) と考えれば, $\frac{X_A/n_A - X_B/n_B}{\sqrt{(1/n_A + 1/n_B)pq}}$ の実現値は $\frac{k_A/n_A - k_B/n_B}{\sqrt{(1/n_A + 1/n_B)p_0(1-p_0)}}$ である。したがって, この実現値を e とおけば, 危険率 α で,

$$e > z_{\alpha/2} \quad \text{または} \quad e < -z_{\alpha/2}$$

のとき仮説をすてる。すなわち,

$$|e| > z_{\alpha/2}$$

のとき仮説をすてる。以上のことをいい直すと,

n_A, n_B が大きいとき, 危険率 α で

$$\frac{|k_A/n_A - k_B/n_B|}{\sqrt{(1/n_A + 1/n_B)p_0(1-p_0)}} > z_{\alpha/2} \quad (6.7.1)$$

のとき仮説をすてる。ただし, $p_0 = \frac{k_A + k_B}{n_A + n_B}$

[注] n_A, n_B が小さいときは計算が大変複雑になるので, 本書では省略する。

(例 1) ある食品会社が自分のところの製品について, 知っているかどうかをテレビで宣伝放送する前と後で, それぞれ無作為抽出した 400 人の主婦について調査したところ, 放送前では 180 人, 放送後では 260 人が知っていると答えた。宣伝効果があったか(危険率 1%)。

(解) 放送前の主婦を母集団 A , 放送後の主婦を母集団 B と考えれば,

$$k_A = 180, k_B = 260, n_A = n_B = 400, p_0 = \frac{180+260}{400+400} = 0.55, z_{0.025} = 2.58$$

仮説「放送前と後で, その製品について知っていた人の割合は等しい」をたてて

$$\frac{|k_A/n_A - k_B/n_B|}{\sqrt{(1/n_A + 1/n_B)p_0(1-p_0)}} = \frac{|180/400 - 260/400|}{\sqrt{(1/400 + 1/400) \times 0.55 \times 0.45}} \doteq 5.71 > z_{0.025}$$

ゆえに, 仮説はすてられる。すなわち, テレビの宣伝効果はあった。

問 1 ある知事選挙で A 候補の支持者は, 第 1 回の調査では無作為抽出した 900 人中 420 人, 第 2 回の調査では 900 人中 485 人であった。 A 候補の支持者は増したといえる

か(危険率 2%).

問 2 ある市で中学 1 年生の男女各 400 人を無作為抽出して生活調査をしたところ、睡眠時間が 8.0 時間以上である人数は、男子 92 人、女子 80 人であった。男子と女子では 8.0 時間以上睡眠する人の比率は異なるといえるか(危険率 5%).

§ 6.8 適合度の検定

ある大企業の従業員を母集団にとったとき、その従業員の学歴によってこの母集団をいくつかの階級に分類することができる。

一般に、ある無限母集団がある性質によって k 個の階級 $C_1, C_2, \dots, C_k$ に分類されているとする。この無限母集団から大きさ n の標本を無作為抽出して、その標本の中で階級 C_i ($i=1, 2, \dots, k$) に属するものの個数 N_i の実現値 n_i を知り、この無限母集団での $C_1, C_2, \dots, C_k$ に属するものの比率の検定を考えよう。

仮説「この無限母集団で階級 $C_1, C_2, \dots, C_k$ に属するものの比率は $p_1, p_2, \dots, p_k$ である」

をたてる。このとき定理 4.10 によ

り、 $\sum_{i=1}^k \frac{(N_i - np_i)^2}{np_i}$ は自由度 $k-1$ の χ^2 -分布に従う。

ところで、 $\sum_{i=1}^k \frac{(N_i - np_i)^2}{np_i}$ の実現値は $\sum_{i=1}^k \frac{(n_i - np_i)^2}{np_i}$ で、この実現値が 0 から遠ざかるにしたがって仮説は不都合となるから、危険率 α で

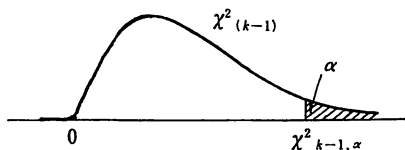


図 6-6

$$\sum_{i=1}^k \frac{(n_i - np_i)^2}{np_i} > \chi^2_{k-1, \alpha} \quad (6.8.1)$$

のとき仮説をすてる。

この適合度の検定は近似的に処理するので、検定を行う場合につぎの 2 つの条件を満足しなければならない。

(1) n は十分に大きい ($n \geq 50$).

(2) n_i は小さくない ($n_i \geq 5$).

仮りに、 n_i の中で 4 以下のものがあれば、適当に階級を合併して 5 以上とな

るようにすればよい。

〔注〕 N_i の実現値 n_i を実測度数, np_i を理論度数という。

(例1) サイコロを120回振ってつぎの結果を得た。このサイコロは正しくないといえるか(危険率5%)。

目の数	1	2	3	4	5	6	計
出た回数	27	12	14	28	24	15	120

(解) 仮説「このサイコロは正しい。すなわち, $p_1=p_2=\cdots=p_6=1/6$ 」をたてる。

$k=6$, $n=120$, $np_i=20 (i=1, 2, \cdots, 6)$, $\chi_{5, 0.05}^2=11.07$

$$\therefore \sum_{i=1}^6 \frac{(n_i - np_i)^2}{np_i} = 12.7 > \chi_{5, 0.05}^2$$

C_i	n_i	np_i	$(n_i - np_i)^2$	$(n_i - np_i)^2 / np_i$
1	27	20	49	49/20
2	12	20	64	64/20
3	14	20	36	36/20
4	28	20	64	64/20
5	24	20	16	16/20
6	15	20	25	25/20
計	120	120	254	12.7

ゆえに仮説はすてられる。すなわち, このサイコロは正しくない。

(例2) 1条に10粒あて, 60条に農作物の種子をまいて発芽試験を行いつぎの結果を得た。発芽状況は二項分布に合っているか(危険率5%)。

1条の発芽数	0	1	2	3	4	5	6	7	8	9	10	計
実測した条の数	4	7	12	18	9	5	2	2	0	1	0	60

(解) 全発芽数 = $0 \times 4 + 1 \times 7 + 2 \times 12 + 3 \times 18 + 4 \times 9 + 5 \times 5 + 6 \times 2 + 7 \times 2 + 8 \times 0$
 $+ 9 \times 1 + 10 \times 0 = 181$

$$\therefore \text{発芽率} = \frac{181}{10 \times 60} \doteq \frac{3}{10}$$

仮説「発芽状況は二項分布に従う」

をたてると、1条に*i*個発芽する確率 p_i は

$$p_i = {}_{10}C_i \left(\frac{3}{10}\right)^i \left(\frac{7}{10}\right)^{10-i}$$

$$\therefore p_0 = 0.028, p_1 = 0.121, p_2 = 0.234, p_3 = 0.267, \dots,$$

また、実測した条の数のうち4以下のものは適当にまとめて表を作ると

i	n_i	p_i	np_i	$n_i - np_i$	$\frac{(n_i - np_i)^2}{np_i}$
0	4	0.028	1.7	2.0	0.444
1	7	0.121	7.3		
2	12	0.234	14.0	-2.0	0.286
3	18	0.267	16.0	2.0	0.250
4	9	0.200	12.0	-3.0	0.750
5	5	0.103	6.2	-1.2	0.232
6	2	0.037	2.2	2.2	1.729
7	2	0.009	0.5		
8	0	0.001	0.1		
9	1	0.000	0.0		
10	0	0.000	0.0		
計	60	1.000	60		3.691

$$\therefore \sum_{i=1}^6 \frac{(n_i - np_i)^2}{np_i} = 3.691 < \chi_{5, 0.05}^2$$

ゆえに仮説はすてられない。すなわち、二項分布に適合していないとはいえない。

問1 エンドウの交配実験によってつぎの結果を得た。これは9:3:3:1に適合しているか(危険率5%)。

種 類	円形黄色	角形黄色	円形緑色	角形緑色	計
度 数	318	102	110	30	560

問2 ある製品の150個の仕切りを検査してつぎの結果を得た。1つの仕切りに含まれる不良品の個数(X)は $m=2$ のポアソン分布に従っているといえるか(危険率5%)。

不良品の個数	0	1	2	3	4	5	6	7	計
観 測 度 数	23	39	43	23	10	7	4	1	150

§6.9 $k \times h$ 分割表の独立性の検定

ある無限母集団が2つの性質 A, B について、つぎのように分類されている。 A については1から k までの k 個の階級、 B については1から h までの h 個の階級、全体としては hk 個の階級に分類されている。 X については i の階級に属し、 B については j の階級に属するような階級を C_{ij} で表す。

この無限母集団から大きさ n の標本を無作為抽出して、その中の階級 C_{ij} に属するものの個数を N_{ij} 、また

$$N_{i1} + N_{i2} + \cdots + N_{ih} = N_{i\cdot} \quad (i=1, 2, \cdots, k)$$

$$N_{1j} + N_{2j} + \cdots + N_{kj} = N_{\cdot j} \quad (j=1, 2, \cdots, h)$$

とおいて右の表を作る。この表を $k \times h$ 分割表という。

この節では N_{ij} の実現値 n_{ij} を

$k \times h$ 分割表

知って、2つの性質 A, B が互いに独立であるかどうかを検定する。

$k \times h$ 分割表

2つの性質 A, B が互いに独立

のとき、 F_{ij} を

$$F_{ij} = \frac{N_{i\cdot} \times N_{\cdot j}}{n}$$

A \ B					
	1	2		h	計
1	N_{11}	N_{12}		N_{1h}	$N_{1\cdot}$
2	N_{21}	N_{22}		N_{2h}	$N_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
k	N_{k1}	N_{k2}		N_{kh}	$N_{k\cdot}$
計	$N_{\cdot 1}$	$N_{\cdot 2}$		$N_{\cdot h}$	n

によって定めると、近似的に $\sum_{i=1}^k \sum_{j=1}^h \frac{(N_{ij} - F_{ij})^2}{F_{ij}}$ は自由度 $(k-1)(h-1)$ の χ^2 -分布に従う(証明省略)。これより

仮説「 A と B は互いに独立である」

をたてたとき、 N_{ij} の実現値 n_{ij} に対する F_{ij} の値を f_{ij} として、危険率 α で

$$\sum_{i=1}^k \sum_{j=1}^h \frac{(n_{ij} - f_{ij})^2}{f_{ij}} > \chi^2_{(k-1)(h-1), \alpha} \quad (6.9.1)$$

のとき仮説をすてる。

$2 \times h$ 分割表

N_{ij} の実現値 n_{ij} に対する $N_{i\cdot}, N_{\cdot j}$ の値をそれぞれ $n_{i\cdot}, n_{\cdot j}$ とすれば、 $k=2$

のとき(6.9.1)の左辺は $\frac{n^2}{n_{1\cdot} n_{2\cdot}} \left\{ \sum_{j=1}^h \frac{n_{1j}^2}{n_{\cdot j}} - \frac{n_{1\cdot}^2}{n} \right\}$ と変形できるから、この場合は

$$\frac{n^2}{n_{1.}n_{2.}} \left\{ \sum_{j=1}^h \frac{n_{1j}^2}{n_{.j}} - \frac{n_{1.}^2}{n} \right\} > \chi^2_{h-1, \alpha} \quad (6.9.2)$$

のとき仮説をすてる。

2×2分割表

$k=h=2$ のとき, (6.9.2)の左辺は $\frac{(n_{11}n_{22}-n_{12}n_{21})^2}{n_{1.}n_{2.}n_{.1}n_{.2}} \times n$ と変形できるから, この場合は

$$\frac{(n_{11}n_{22}-n_{12}n_{21})^2}{n_{1.}n_{2.}n_{.1}n_{.2}} \times n > \chi^2_{1, \alpha} \quad (6.9.3)$$

のとき仮説をすてる。

以上の検定で使う式(6.9.1), (6.9.2), (6.9.3)はすべて $n_{ij} \geq 5$ のときに限る。もし, n_{ij} の中で5以下のものがあれば適当に階級を合併する。2×2分割表では合併することができないので, このときはつぎの式を使って検定する。

$$\frac{(n_{11}n_{22}-n_{12}n_{21} \pm n/2)^2}{n_{1.}n_{2.}n_{.1}n_{.2}} \times n > \chi^2_{1, \alpha} \quad (6.9.4)$$

のとき仮説をすてる。

ここで, 左辺の複号は分子が小さくなるように, いずれか一方をとる。

(例1) A, B, C, D の4工場で作っている皮革を検査し, 1級品, 2級品, 3級品に分類してつぎの表を得た。

	1 級品	2 級品	3 級品	計
A	54	28	18	100
B	42	44	14	100
C	41	49	10	100
D	49	30	21	100
計	186	151	63	400

A, B, C, D の4工場によって差異が認められるか(危険率5%)。

(解) 仮説「 A, B, C, D の4工場によって差異がない」をたてる。

$\sum \sum \frac{(n_{ij}-f_{ij})^2}{f_{ij}}$ の値をつぎの手順で表を作って求める。

f_{ij} の値

	1 級品	2 級品	3 級品
A	46.5	37.75	15.75
B	46.5	37.75	15.75
C	46.5	37.75	15.75
D	46.5	37.75	15.75

 $(n_{ij}-f_{ij})$ の値

	1 級品	2 級品	3 級品
A	7.5	-9.75	2.25
B	-4.5	6.25	-1.75
C	-5.5	11.25	-5.75
D	2.5	-7.75	5.25

 $(n_{ij}-f_{ij})^2$ の値

	1 級品	2 級品	3 級品
A	56.25	95.06	5.06
B	20.25	39.06	3.06
C	30.25	126.56	33.06
D	6.25	60.06	27.56

 $(n_{ij}-f_{ij})^2/f_{ij}$ の値

	1 級品	2 級品	3 級品	計
A	1.21	2.518	0.321	4.049
B	0.435	1.035	0.194	1.664
C	0.650	3.353	2.099	6.102
D	0.134	1.591	1.75	3.475
計	2.429	8.497	4.364	15.29

$$\therefore \sum_{i=1}^4 \sum_{j=1}^3 \frac{(n_{ij}-f_{ij})^2}{f_{ij}} \doteq 15.29 > \chi_{3,0.05}^2 = 12.59$$

ゆえに、仮説はすてられる。すなわち、A, B, C, D の 4 工場によって差異が認められる。

(例 2) 無作為抽出した 200 人の有権者について、3 つの政党 A, B, C の支持者を調べたところつぎの表を得た。40 歳以上と 39 歳以下とで、支持する政党に差異が認められるか (危険率 5%)。

	A	B	C	計
40 歳以上	60	30	15	105
39 歳以下	37	36	22	95
計	97	66	37	200

(解) 仮説「40 歳以上と 39 歳以下とで、支持する政党に差異がない」をたてる。

$$\frac{n^2}{n_{1.}n_{2.}} \left\{ \sum_{j=1}^3 \frac{n_{1j}^2}{n_{.j}} - \frac{n_{1.}^2}{n} \right\}$$

$$= \frac{200^2}{105 \times 95} \left\{ \frac{60^2}{97} + \frac{30^2}{66} + \frac{15^2}{37} - \frac{105^2}{200} \right\} \doteq 6.84 > \chi_{2, 0.05}^2 = 5.99$$

ゆえに、仮説はすてられる。すなわち、40歳以上と39歳以下とで、支持する政党に差異が認められる。

(例3) ある学校で近視者について調査したところつぎの表を得た。近視になりやすさに男女の差があるか(危険率5%)。

	近 視	近視でない	計
男 子	31	139	170
女 子	36	124	160
計	67	263	330

(解) 仮説「近視になりやすさに男女の差はない」をたてる。

$$\frac{(n_{11}n_{22} - n_{12}n_{21})^2}{n_{1.}n_{2.}n_{.1}n_{.2}} \times n = \frac{(31 \times 124 - 139 \times 36)^2}{170 \times 160 \times 67 \times 263} \times 330 \doteq 0.93 < \chi_{1, 0.05}^2 = 3.84$$

ゆえに、仮説はすてられない。すなわち、近視になりやすさに男女の差が認められない。

(例4) ある食品の嗜好について、男女それぞれ85人、55人について、好き、嫌いのどちらかを調べて右表を得た。この食品に対する嗜好の度合いと性別は関係があるといえるか(危険率5%)。

	好 き	嫌 い	計
男	16	69	85
女	4	51	55
計	20	120	140

(解) 仮説「この食品に対する嗜好の度合いと性別は関係がない」をたてる。

$$\frac{(n_{11}n_{22} - n_{12}n_{21} \pm n/2)^2}{n_{1.}n_{2.}n_{.1}n_{.2}} \times n = \frac{(16 \times 51 - 69 \times 4 - 140/2)^2}{85 \times 55 \times 20 \times 120} \times 140$$

$$\doteq 2.76 < \chi_{1, 0.05}^2 = 3.84$$

ゆえに、仮説はすてられない。すなわち、この食品に対する嗜好の度合いと性別は関係があるとはいえない。

問1 100人の生徒について英、数の成績によって分類したところつぎの表を得た。両科目の成績に関係があるといえるか(危険率5%)。

英 数	優	良	可	計
優	16	8	8	32
良	10	22	9	41
可	7	8	12	27
計	33	38	29	100

問 2 5 班が作った製品を調べて上, 中, 下の 3 階級に分けたところつぎの表を得た.
班によって技量に差異が認められるか(危険率 5%).

班 性能	1	2	3	4	5	計
上	25	7	7	15	26	80
中	93	11	11	45	53	213
下	29	5	10	12	11	67
計	147	23	28	72	90	360

問 3 あるテレビドラマを見た 170 人について批評を求めたところつぎの表を得た.
男女の批評に差異が認められるか(危険率 5%).

	良 い	普 通	つまらない	計
男	28	54	18	100
女	22	28	20	70
計	50	82	38	170

問 4 A, B, C, D の 4 種類の薬品をそれぞれ 100 個ずつの被試験体に試みて, 効力の
有無を調べたところつぎの表を得た. 薬品の効力に差異が認められるか(危険率 5%).

	A	B	C	D	計
有 効	63	85	71	57	276
無 効	37	15	29	43	124
計	100	100	100	100	400

問5 2つの工場 A, B で製造している製品について、検査の結果は右表の通りであった。両工場に差異が認められるか(危険率5%)。

	合 格	不合格	計
A	120	30	150
B	145	35	180
計	265	65	330

問6 50人の患者の症状を A, B の2組に分けて、薬効の有無を調べたところ右表を得た。症状により薬効は違うか(危険率5%)。

	有 効	無 効	計
A	19	11	30
B	17	3	20
計	36	14	50

〔付表2〕 ポアソン分布表(2)

k	m									
	3.1	3.2	3.3	3.4	3.5	3.6	3.7	3.8	3.9	4.0
0	.0450	.0408	.0369	.0334	.0302	.0273	.0247	.0224	.0202	.0183
1	.1397	.1304	.1217	.1135	.1057	.0984	.0915	.0850	.0789	.0733
2	.2165	.2087	.2008	.1929	.1850	.1771	.1692	.1615	.1539	.1465
3	.2237	.2226	.2209	.2186	.2158	.2125	.2087	.2046	.2001	.1954
4	.1733	.1781	.1822	.1858	.1888	.1912	.1931	.1944	.1951	.1954
5	.1075	.1140	.1203	.1263	.1322	.1377	.1429	.1477	.1522	.1563
6	.0555	.0608	.0662	.0716	.0771	.0826	.0881	.0936	.0989	.1042
7	.0246	.0278	.0312	.0348	.0385	.0425	.0466	.0508	.0551	.0595
8	.0095	.0111	.0129	.0148	.0169	.0191	.0215	.0241	.0269	.0298
9	.0033	.0039	.0047	.0056	.0065	.0076	.0089	.0102	.0116	.0132
10	.0010	.0013	.0016	.0019	.0023	.0027	.0033	.0039	.0045	.0053
11	.0003	.0004	.0005	.0006	.0007	.0009	.0011	.0013	.0016	.0019
12	.0001	.0001	.0001	.0002	.0002	.0003	.0003	.0004	.0005	.0006
13	—	—	—	—	.0001	.0001	.0001	.0001	.0002	.0002
14	—	—	—	—	—	—	—	—	—	.0001
k	4.1	4.2	4.3	4.4	4.5	4.6	4.7	4.8	4.9	5.0
0	.0166	.0150	.0136	.0123	.0111	.0101	.0091	.0082	.0074	.0067
1	.0679	.0630	.0583	.0540	.0500	.0462	.0427	.0395	.0365	.0337
2	.1393	.1323	.1254	.1188	.1125	.1063	.1005	.0948	.0894	.0842
3	.1904	.1852	.1798	.1743	.1687	.1631	.1574	.1517	.1460	.1404
4	.1951	.1944	.1933	.1917	.1898	.1875	.1849	.1820	.1789	.1755
5	.1600	.1633	.1662	.1687	.1708	.1725	.1738	.1747	.1753	.1755
6	.1093	.1143	.1191	.1237	.1281	.1323	.1362	.1398	.1432	.1462
7	.0640	.0686	.0732	.0778	.0824	.0869	.0914	.0959	.1002	.1044
8	.0328	.0360	.0393	.0428	.0463	.0500	.0537	.0575	.0614	.0653
9	.0150	.0168	.0188	.0209	.0232	.0255	.0281	.0307	.0334	.0363
10	.0061	.0071	.0081	.0092	.0104	.0118	.0132	.0147	.0164	.0181
11	.0023	.0027	.0032	.0037	.0042	.0049	.0056	.0064	.0073	.0082
12	.0008	.0009	.0011	.0014	.0016	.0019	.0022	.0026	.0030	.0034
13	.0002	.0003	.0004	.0005	.0006	.0007	.0008	.0009	.0011	.0013
14	.0001	.0001	.0001	.0001	.0002	.0002	.0003	.0003	.0004	.0005
15	—	—	—	—	.0001	.0001	.0001	.0001	.0001	.0002

[附表4] χ^2 -分 布 表 $\chi^2_{n,\alpha}$

α	n	.995	.99	.975	.95	.90	.75	.50	.25	.10	.05	.025	.01	.005	α	n
1	1	0.000393	0.000157	0.000982	0.00393	0.0158	0.102	0.455	1.323	2.71	3.84	5.02	6.63	7.88	1	1
2	2	0.0100	0.0201	0.0506	0.103	0.211	0.575	1.386	2.77	4.61	5.99	7.38	9.21	10.60	2	2
3	3	0.0717	0.115	0.216	0.352	0.584	1.213	2.37	4.11	6.25	7.81	9.35	11.34	12.84	3	3
4	4	0.207	0.297	0.484	0.711	1.064	1.923	3.35	5.39	7.78	9.49	11.14	13.28	14.86	4	4
5	5	0.412	0.554	0.831	1.145	1.610	2.67	4.35	6.63	9.24	11.07	12.83	15.09	16.75	5	5
6	6	0.676	0.872	1.237	1.635	2.20	3.45	5.35	7.84	10.64	12.59	14.45	16.81	18.55	6	6
7	7	0.989	1.239	1.690	2.17	2.83	4.25	6.35	9.04	12.02	14.07	16.01	18.48	20.3	7	7
8	8	1.344	1.646	2.18	2.73	3.49	5.07	7.34	10.22	13.36	15.51	17.53	20.1	22.0	8	8
9	9	1.735	2.09	2.70	3.33	4.17	5.90	8.34	11.39	14.68	16.92	19.02	21.7	23.6	9	9
10	10	2.16	2.56	3.25	3.94	4.87	6.74	9.34	12.55	15.99	18.31	20.5	23.2	25.2	10	10
11	11	2.60	3.05	3.82	4.57	5.58	7.58	10.34	13.70	17.28	19.68	21.9	24.7	26.8	11	11
12	12	3.07	3.57	4.40	5.23	6.30	8.44	11.34	14.85	18.55	21.0	23.3	26.2	28.3	12	12
13	13	3.57	4.11	5.01	5.89	7.04	9.30	12.34	15.98	19.81	22.4	24.7	27.7	29.8	13	13
14	14	4.07	4.66	5.63	6.57	7.79	10.17	13.34	17.12	21.1	23.7	26.1	29.1	31.3	14	14
15	15	4.60	5.23	6.26	7.26	8.55	11.04	14.34	18.25	22.3	25.0	27.5	30.6	32.8	15	15
16	16	5.14	5.81	6.91	7.96	9.31	11.91	15.34	19.37	23.5	26.3	28.8	32.0	34.3	16	16
17	17	5.70	6.41	7.56	8.67	10.09	12.79	16.34	20.5	24.8	27.6	30.2	33.4	35.7	17	17
18	18	6.26	7.01	8.23	9.39	10.86	13.68	17.34	21.6	26.0	28.9	31.5	34.8	37.2	18	18
19	19	6.84	7.63	8.91	10.12	11.65	14.56	18.34	22.7	27.2	30.1	32.9	36.2	38.6	19	19
20	20	7.43	8.26	9.59	10.85	12.44	15.45	19.34	23.8	28.4	31.4	34.2	37.6	40.0	20	20
21	21	8.03	8.90	10.28	11.59	13.24	16.34	20.3	24.9	29.6	32.7	35.5	38.9	41.4	21	21
22	22	8.64	9.54	10.98	12.34	14.04	17.24	21.3	26.0	30.8	33.9	36.8	40.3	42.8	22	22
23	23	9.26	10.20	11.69	13.09	14.85	18.14	22.3	27.1	32.0	35.2	38.1	41.6	44.2	23	23
24	24	9.89	10.86	12.40	13.85	15.66	19.04	23.3	28.2	33.2	36.4	39.4	43.0	45.6	24	24
25	25	10.52	11.52	13.12	14.61	16.47	19.94	24.3	29.3	34.4	37.7	40.6	44.3	46.9	25	25
26	26	11.16	12.20	13.84	15.38	17.29	20.8	25.3	30.4	35.6	38.9	41.9	45.6	48.3	26	26
27	27	11.81	12.88	14.57	16.15	18.11	21.7	26.3	31.5	36.7	40.1	43.2	47.0	49.6	27	27
28	28	12.46	13.56	15.31	16.93	18.94	22.7	27.3	32.6	37.9	41.3	44.5	48.3	51.0	28	28
29	29	13.12	14.26	16.05	17.71	19.77	23.6	28.3	33.7	39.1	42.6	45.7	49.6	52.3	29	29
30	30	13.79	14.96	16.79	18.49	20.6	24.5	29.3	34.8	40.3	43.8	47.0	50.9	53.7	30	30
40	40	20.7	22.2	24.4	26.5	29.1	33.7	39.3	45.6	51.8	55.8	59.3	63.7	66.8	40	40
50	50	29.7	32.4	34.8	37.7	42.9	49.3	56.3	63.2	67.5	71.4	76.2	79.5	82.8	50	50
60	60	38.5	37.5	40.5	43.2	46.5	52.3	59.3	67.0	74.4	79.1	83.3	88.4	92.0	60	60
70	70	43.3	45.4	48.8	51.7	55.3	61.7	69.3	77.6	85.5	90.5	95.0	100.4	104.2	70	70
80	80	51.2	53.5	57.2	60.4	64.3	71.1	79.3	88.1	96.6	101.9	106.6	112.3	116.3	80	80
90	90	59.2	61.8	65.6	69.1	73.3	80.6	89.3	98.6	107.6	113.1	118.1	124.1	128.3	90	90
100	100	67.3	70.1	74.2	77.9	82.4	90.1	99.3	109.1	118.5	124.3	129.6	135.8	140.2	100	100

〔付表5〕 t -分 布 表 $t_{n,\alpha}$

$n \backslash \alpha$	0.05	0.025	0.0125	0.005	0.0025	0.0005
1	6.3138	12.706	25.472	63.657	127.32	636.62
2	2.9200	4.3027	6.2053	9.9248	14.089	31.598
3	2.3534	3.1825	4.1765	5.8409	7.4533	12.941
4	2.1318	2.7764	3.4954	4.6041	5.5976	8.610
5	2.0150	2.5706	3.1634	4.0321	4.7733	6.859
6	1.9432	2.4469	2.9687	3.7074	4.3168	5.959
7	1.8946	2.3646	2.8412	3.4995	4.0293	5.405
8	1.8595	2.3060	2.7515	3.3554	3.8325	5.041
9	1.8331	2.2622	2.6850	3.2498	3.6897	4.781
10	1.8125	2.2281	2.6338	3.1693	3.5814	4.587
11	1.7959	2.2010	2.5931	3.1058	3.4966	4.437
12	1.7823	2.1788	2.5600	3.0545	3.4284	4.318
13	1.7709	2.1604	2.5326	3.0123	3.3725	4.221
14	1.7613	2.1448	2.5096	2.9768	3.3257	4.140
15	1.7530	2.1315	2.4899	2.9467	3.2860	4.073
16	1.7459	2.1199	2.4729	2.9208	3.2520	4.015
17	1.7396	2.1098	2.4581	2.8982	3.2225	3.965
18	1.7341	2.1009	2.4450	2.8784	3.1966	3.922
19	1.7291	2.0930	2.4334	2.8609	3.1737	3.883
20	1.7247	2.0860	2.4231	2.8453	3.1534	3.850
21	1.7207	2.0796	2.4138	2.8314	3.1352	3.819
22	1.7171	2.0739	2.4055	2.8188	3.1188	3.792
23	1.7139	2.0687	2.3979	2.8073	3.1040	3.767
24	1.7109	2.0639	2.3910	2.7969	3.0905	3.745
25	1.7081	2.0595	2.3846	2.7874	3.0782	3.725
26	1.7056	2.0555	2.3788	2.7787	3.0669	3.707
27	1.7033	2.0518	2.3734	2.7707	3.0565	3.690
28	1.7011	2.0484	2.3685	2.7633	3.0469	3.674
29	1.6991	2.0452	2.3638	2.7564	3.0380	3.659
30	1.6973	2.0423	2.3596	2.7500	3.0298	3.646
40	1.6839	2.0211	2.3289	2.7045	2.9712	3.551
60	1.6707	2.0003	2.2991	2.6603	2.9146	3.460
120	1.6577	1.9799	2.2699	2.6174	2.8599	3.373
∞	1.6449	1.9600	2.2414	2.5758	2.8070	3.291

[付表6] F -分布表 $F_{n,0.05}^m$

$m \backslash n$	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9	243.9	246.0	248.0	249.0	250.1	251.1	252.2	253.3	254.3
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.39	19.40	19.41	19.43	19.45	19.45	19.46	19.47	19.48	19.49	19.50
3	10.13	9.552	9.227	9.117	9.014	8.941	8.887	8.845	8.812	8.786	8.745	8.703	8.660	8.639	8.617	8.594	8.572	8.549	8.527
4	7.709	6.944	6.591	6.388	6.256	6.163	6.094	6.041	5.999	5.964	5.912	5.858	5.803	5.774	5.746	5.717	5.688	5.658	5.628
5	6.608	5.786	5.410	5.192	5.050	4.950	4.876	4.818	4.773	4.735	4.678	4.619	4.558	4.527	4.496	4.464	4.431	4.398	4.365
6	5.987	5.143	4.757	4.534	4.387	4.284	4.207	4.147	4.099	4.060	4.000	3.938	3.874	3.842	3.808	3.774	3.740	3.705	3.669
7	5.591	4.737	4.347	4.120	3.972	3.866	3.787	3.726	3.677	3.637	3.575	3.511	3.445	3.411	3.376	3.340	3.304	3.267	3.230
8	5.318	4.459	4.066	3.838	3.688	3.581	3.501	3.438	3.388	3.347	3.284	3.218	3.150	3.115	3.079	3.043	3.005	2.967	2.928
9	5.117	4.257	3.863	3.633	3.482	3.374	3.293	3.230	3.179	3.137	3.073	3.006	2.937	2.901	2.864	2.826	2.787	2.748	2.707
10	4.965	4.103	3.708	3.478	3.326	3.217	3.136	3.072	3.020	2.978	2.913	2.845	2.774	2.737	2.700	2.661	2.621	2.580	2.538
11	4.844	3.982	3.587	3.357	3.204	3.095	3.012	2.948	2.896	2.854	2.788	2.719	2.646	2.609	2.571	2.531	2.490	2.448	2.405
12	4.747	3.885	3.490	3.259	3.106	2.996	2.913	2.849	2.796	2.753	2.687	2.617	2.544	2.506	2.466	2.425	2.384	2.341	2.296
13	4.667	3.806	3.411	3.179	3.025	2.915	2.832	2.767	2.714	2.671	2.604	2.533	2.459	2.420	2.380	2.339	2.297	2.252	2.206
14	4.600	3.739	3.344	3.112	2.958	2.848	2.764	2.699	2.646	2.602	2.534	2.463	2.388	2.349	2.308	2.266	2.223	2.178	2.131
15	4.543	3.682	3.287	3.056	2.901	2.791	2.707	2.641	2.588	2.544	2.475	2.404	2.328	2.288	2.247	2.204	2.160	2.114	2.066
16	4.494	3.634	3.239	3.007	2.852	2.741	2.657	2.591	2.538	2.494	2.425	2.352	2.276	2.235	2.194	2.151	2.106	2.059	2.010
17	4.451	3.592	3.197	2.965	2.810	2.699	2.614	2.548	2.494	2.450	2.381	2.308	2.230	2.189	2.148	2.104	2.058	2.011	1.960
18	4.414	3.555	3.160	2.928	2.773	2.661	2.577	2.510	2.456	2.412	2.342	2.269	2.191	2.150	2.107	2.063	2.017	1.968	1.917
19	4.381	3.522	3.127	2.895	2.740	2.628	2.544	2.477	2.423	2.378	2.308	2.234	2.156	2.114	2.071	2.026	1.980	1.930	1.878
20	4.351	3.493	3.098	2.866	2.711	2.599	2.514	2.447	2.393	2.348	2.278	2.203	2.124	2.083	2.039	1.994	1.946	1.896	1.843
21	4.325	3.467	3.073	2.840	2.685	2.573	2.488	2.421	2.366	2.321	2.250	2.176	2.096	2.054	2.010	1.965	1.917	1.866	1.812
22	4.301	3.443	3.049	2.817	2.661	2.549	2.464	2.397	2.342	2.297	2.226	2.151	2.071	2.028	1.984	1.938	1.890	1.838	1.783
23	4.279	3.422	3.028	2.796	2.640	2.528	2.442	2.375	2.320	2.275	2.204	2.128	2.048	2.005	1.961	1.914	1.865	1.813	1.757
24	4.260	3.403	3.009	2.776	2.621	2.508	2.423	2.355	2.300	2.255	2.183	2.108	2.027	1.984	1.939	1.892	1.842	1.790	1.733
25	4.242	3.385	2.991	2.758	2.603	2.490	2.405	2.337	2.282	2.237	2.165	2.089	2.008	1.964	1.919	1.872	1.822	1.768	1.711
26	4.225	3.369	2.975	2.742	2.587	2.474	2.389	2.321	2.266	2.220	2.148	2.072	1.990	1.946	1.901	1.853	1.803	1.749	1.691
27	4.210	3.354	2.960	2.727	2.572	2.459	2.373	2.305	2.250	2.204	2.132	2.056	1.974	1.930	1.884	1.836	1.785	1.731	1.672
28	4.196	3.340	2.947	2.714	2.558	2.445	2.359	2.291	2.236	2.190	2.118	2.041	1.959	1.915	1.869	1.820	1.769	1.714	1.654
29	4.183	3.328	2.934	2.701	2.545	2.432	2.346	2.278	2.223	2.177	2.105	2.028	1.945	1.901	1.854	1.806	1.754	1.698	1.638
30	4.171	3.316	2.922	2.690	2.534	2.421	2.334	2.266	2.211	2.165	2.092	2.015	1.932	1.887	1.841	1.792	1.740	1.684	1.622
40	4.085	3.232	2.839	2.606	2.450	2.336	2.249	2.180	2.124	2.077	2.004	1.925	1.839	1.793	1.746	1.693	1.637	1.577	1.509
60	4.001	3.150	2.758	2.525	2.368	2.254	2.167	2.097	2.040	1.993	1.917	1.836	1.748	1.700	1.649	1.594	1.534	1.467	1.389
120	3.920	3.072	2.680	2.447	2.290	2.175	2.087	2.016	1.959	1.911	1.834	1.751	1.659	1.608	1.554	1.495	1.429	1.352	1.254
∞	3.842	2.996	2.605	2.372	2.214	2.099	2.010	1.938	1.880	1.831	1.752	1.666	1.571	1.517	1.459	1.394	1.318	1.221	1.000

〔付表7〕 F -分布表 $F_{n,0.025}^m$

$m \backslash n$	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	647.8	799.5	864.2	899.6	921.9	937.1	948.2	956.7	963.3	968.6	976.7	984.9	993.1	997.3	1001.	1006.	1010.	1014.	1018.
2	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39	39.40	39.42	39.43	39.45	39.46	39.47	39.47	39.48	39.49	39.50
3	17.44	16.04	15.44	15.10	14.80	14.74	14.62	14.54	14.47	14.42	14.34	14.25	14.17	14.12	14.08	14.04	13.99	13.95	13.90
4	12.22	10.65	9.979	9.605	9.365	9.197	9.074	8.980	8.905	8.844	8.751	8.657	8.560	8.511	8.461	8.411	8.360	8.309	8.257
5	10.01	8.434	7.764	7.388	7.146	6.978	6.853	6.757	6.681	6.619	6.525	6.428	6.329	6.278	6.227	6.175	6.123	6.069	6.015
6	8.813	7.260	6.599	6.227	5.988	5.820	5.695	5.600	5.523	5.461	5.366	5.268	5.168	5.117	5.065	5.013	4.959	4.905	4.849
7	8.073	6.542	5.890	5.523	5.285	5.119	4.995	4.899	4.823	4.761	4.666	4.568	4.467	4.415	4.362	4.309	4.254	4.199	4.142
8	7.571	6.060	5.416	5.053	4.817	4.652	4.529	4.433	4.357	4.295	4.200	4.101	4.000	3.947	3.894	3.840	3.784	3.728	3.670
9	7.209	5.715	5.078	4.718	4.484	4.320	4.197	4.102	4.026	3.964	3.868	3.769	3.667	3.614	3.560	3.506	3.449	3.392	3.333
10	6.937	5.456	4.826	4.468	4.236	4.072	3.950	3.855	3.779	3.717	3.621	3.522	3.419	3.365	3.311	3.255	3.198	3.140	3.080
11	6.724	5.256	4.630	4.275	4.044	3.881	3.759	3.664	3.588	3.526	3.430	3.330	3.226	3.173	3.118	3.061	3.004	2.944	2.883
12	6.554	5.096	4.474	4.121	3.891	3.728	3.607	3.512	3.436	3.374	3.277	3.177	3.073	3.019	2.963	2.906	2.848	2.787	2.725
13	6.414	4.965	4.347	3.996	3.767	3.604	3.483	3.388	3.312	3.250	3.153	3.053	2.948	2.893	2.837	2.780	2.720	2.659	2.596
14	6.298	4.857	4.242	3.892	3.663	3.501	3.380	3.285	3.209	3.147	3.050	2.949	2.844	2.789	2.732	2.674	2.614	2.552	2.487
15	6.200	4.765	4.153	3.804	3.576	3.415	3.293	3.199	3.123	3.060	2.963	2.862	2.756	2.701	2.644	2.585	2.524	2.461	2.395
16	6.115	4.687	4.077	3.729	3.502	3.341	3.219	3.125	3.049	2.986	2.889	2.788	2.681	2.625	2.568	2.509	2.447	2.383	2.316
17	6.042	4.619	4.011	3.665	3.438	3.277	3.156	3.061	2.985	2.922	2.825	2.723	2.616	2.560	2.502	2.442	2.380	2.315	2.247
18	5.978	4.560	3.954	3.608	3.382	3.221	3.100	3.005	2.929	2.866	2.769	2.667	2.559	2.503	2.445	2.384	2.321	2.256	2.187
19	5.922	4.508	3.903	3.557	3.333	3.172	3.051	2.956	2.880	2.817	2.720	2.617	2.509	2.452	2.394	2.333	2.270	2.203	2.133
20	5.872	4.461	3.859	3.515	3.291	3.130	3.009	2.913	2.837	2.774	2.676	2.573	2.465	2.408	2.349	2.287	2.223	2.156	2.085
21	5.827	4.420	3.819	3.475	3.250	3.090	2.969	2.874	2.798	2.735	2.637	2.534	2.425	2.368	2.308	2.247	2.182	2.114	2.042
22	5.786	4.383	3.783	3.440	3.215	3.055	2.934	2.839	2.763	2.700	2.602	2.498	2.389	2.332	2.272	2.210	2.145	2.076	2.003
23	5.750	4.349	3.751	3.408	3.184	3.023	2.902	2.808	2.731	2.668	2.570	2.467	2.357	2.299	2.239	2.176	2.111	2.042	1.968
24	5.717	4.319	3.721	3.378	3.155	2.995	2.874	2.779	2.703	2.640	2.541	2.437	2.327	2.269	2.209	2.146	2.080	2.010	1.935
25	5.686	4.296	3.694	3.351	3.129	2.969	2.848	2.753	2.677	2.614	2.515	2.411	2.301	2.242	2.182	2.118	2.052	1.981	1.906
26	5.659	4.266	3.670	3.329	3.105	2.945	2.824	2.729	2.653	2.590	2.491	2.387	2.276	2.217	2.157	2.093	2.026	1.955	1.878
27	5.633	4.242	3.647	3.307	3.083	2.923	2.802	2.707	2.631	2.568	2.469	2.365	2.254	2.195	2.133	2.069	2.002	1.930	1.853
28	5.610	4.221	3.626	3.286	3.063	2.903	2.782	2.687	2.611	2.548	2.449	2.345	2.234	2.175	2.112	2.048	1.980	1.907	1.829
29	5.588	4.201	3.607	3.267	3.044	2.884	2.763	2.668	2.592	2.529	2.430	2.325	2.213	2.154	2.092	2.028	1.959	1.886	1.807
30	5.568	4.182	3.589	3.250	3.027	2.867	2.746	2.651	2.575	2.511	2.412	2.307	2.195	2.136	2.074	2.009	1.940	1.866	1.787
40	5.424	4.051	3.463	3.126	2.904	2.744	2.624	2.529	2.452	2.388	2.289	2.182	2.068	2.007	1.943	1.875	1.803	1.724	1.637
60	5.286	3.925	3.343	2.998	2.776	2.617	2.500	2.405	2.328	2.264	2.165	2.057	1.942	1.880	1.815	1.744	1.667	1.581	1.482
120	5.152	3.805	3.227	2.884	2.674	2.515	2.398	2.293	2.216	2.152	2.053	1.945	1.829	1.766	1.699	1.614	1.530	1.433	1.310
∞	5.024	3.689	3.116	2.786	2.587	2.428	2.311	2.206	2.129	2.065	1.966	1.858	1.741	1.677	1.610	1.484	1.388	1.268	1.000

[付表8] F-分 布 表 $F_{n,0.01}^m$

$\frac{m}{n}$	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	4052	5000	5403	5625	5764	5859	5923	5982	6023	6056	6106	6157	6209	6235	6261	6287	6313	6339	6366
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40	99.42	99.43	99.45	99.46	99.47	99.47	99.48	99.49	99.50
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35	27.23	27.05	26.87	26.69	26.60	26.51	26.41	26.32	26.22	26.13
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55	14.37	14.20	14.02	13.93	13.84	13.75	13.65	13.56	13.46
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	9.888	9.722	9.553	9.467	9.379	9.291	9.202	9.112	9.020
6	13.75	10.93	9.780	9.148	8.746	8.466	8.260	8.102	7.976	7.874	7.718	7.559	7.396	7.313	7.229	7.143	7.057	6.969	6.880
7	12.25	9.547	8.451	7.847	7.460	7.191	6.993	6.840	6.719	6.620	6.469	6.314	6.155	6.074	5.992	5.908	5.824	5.737	5.650
8	11.26	8.649	7.591	7.006	6.632	6.371	6.178	6.029	5.911	5.814	5.667	5.515	5.359	5.279	5.198	5.116	5.032	4.946	4.859
9	10.56	8.022	6.992	6.422	6.057	5.802	5.613	5.467	5.351	5.257	5.111	4.962	4.808	4.729	4.649	4.567	4.483	4.398	4.311
10	10.04	7.559	6.552	5.994	5.636	5.386	5.200	5.057	4.942	4.849	4.706	4.558	4.405	4.327	4.247	4.165	4.082	3.997	3.909
11	9.646	7.206	6.217	5.668	5.316	5.069	4.886	4.745	4.632	4.539	4.397	4.251	4.098	4.021	3.941	3.860	3.776	3.690	3.603
12	9.330	6.927	5.953	5.412	5.064	4.821	4.640	4.499	4.388	4.296	4.155	4.010	3.858	3.781	3.701	3.619	3.536	3.449	3.361
13	9.074	6.701	5.739	5.205	4.862	4.620	4.441	4.302	4.191	4.100	3.960	3.815	3.663	3.587	3.507	3.425	3.341	3.255	3.165
14	8.862	6.515	5.564	5.035	4.695	4.456	4.278	4.140	4.030	3.939	3.800	3.656	3.505	3.427	3.348	3.266	3.181	3.094	3.004
15	8.683	6.359	5.417	4.893	4.556	4.318	4.142	4.005	3.895	3.805	3.666	3.522	3.372	3.294	3.214	3.132	3.047	2.960	2.868
16	8.531	6.226	5.292	4.773	4.437	4.202	4.026	3.890	3.780	3.691	3.553	3.409	3.259	3.181	3.101	3.018	2.933	2.845	2.753
17	8.400	6.112	5.185	4.669	4.336	4.102	3.927	3.791	3.682	3.593	3.455	3.312	3.162	3.084	3.003	2.921	2.835	2.746	2.653
18	8.285	6.013	5.092	4.579	4.248	4.015	3.841	3.705	3.597	3.508	3.371	3.227	3.077	2.999	2.919	2.835	2.749	2.660	2.566
19	8.185	5.926	5.010	4.500	4.171	3.939	3.765	3.631	3.523	3.434	3.297	3.153	3.003	2.925	2.844	2.761	2.674	2.584	2.489
20	8.096	5.849	4.938	4.431	4.103	3.871	3.699	3.564	3.457	3.368	3.231	3.088	2.938	2.860	2.779	2.695	2.608	2.517	2.421
21	8.017	5.780	4.874	4.369	4.042	3.812	3.640	3.506	3.398	3.310	3.173	3.030	2.880	2.801	2.720	2.636	2.548	2.457	2.360
22	7.945	5.719	4.817	4.313	3.988	3.758	3.587	3.453	3.346	3.258	3.121	2.978	2.827	2.749	2.668	2.583	2.495	2.403	2.306
23	7.881	5.664	4.765	4.264	3.939	3.710	3.539	3.406	3.299	3.211	3.074	2.931	2.781	2.702	2.620	2.536	2.447	2.354	2.256
24	7.823	5.614	4.718	4.218	3.895	3.667	3.496	3.363	3.256	3.168	3.032	2.889	2.738	2.659	2.577	2.492	2.404	2.310	2.211
25	7.770	5.568	4.676	4.177	3.855	3.627	3.457	3.324	3.217	3.129	2.993	2.850	2.699	2.620	2.538	2.453	2.364	2.270	2.169
26	7.721	5.526	4.637	4.140	3.818	3.591	3.421	3.288	3.182	3.094	2.958	2.815	2.664	2.585	2.503	2.417	2.327	2.233	2.132
27	7.677	5.488	4.601	4.106	3.785	3.558	3.388	3.256	3.149	3.062	2.926	2.783	2.632	2.552	2.470	2.384	2.294	2.199	2.097
28	7.636	5.453	4.568	4.074	3.754	3.528	3.358	3.226	3.120	3.032	2.896	2.753	2.602	2.522	2.440	2.354	2.263	2.167	2.064
29	7.598	5.421	4.536	4.045	3.725	3.500	3.330	3.198	3.092	3.005	2.869	2.726	2.574	2.495	2.412	2.325	2.234	2.138	2.034
30	7.563	5.390	4.510	4.018	3.699	3.474	3.305	3.173	3.067	2.979	2.843	2.700	2.549	2.469	2.386	2.299	2.208	2.111	2.006
40	7.314	5.179	4.313	3.828	3.514	3.291	3.124	2.993	2.888	2.801	2.665	2.522	2.369	2.288	2.205	2.114	2.019	1.917	1.805
60	7.077	4.977	4.126	3.649	3.339	3.119	2.953	2.823	2.719	2.632	2.496	2.352	2.198	2.115	2.032	1.936	1.836	1.726	1.601
120	6.851	4.787	3.949	3.480	3.174	2.956	2.792	2.663	2.559	2.472	2.336	2.192	2.035	1.950	1.867	1.763	1.656	1.533	1.381
∞	6.635	4.605	3.782	3.319	3.017	2.802	2.639	2.511	2.407	2.321	2.185	2.039	1.878	1.791	1.706	1.592	1.473	1.335	1.000

解 答

—〈第 1 章〉—

- p. 8 問 1 平均 153.89, 標準偏差 5.46 問 2 略
 問 3 41 点から 65 点までの者は 25 人いる.
 p. 13 問 $\mu_x=72.8$, $\sigma_x^2=3.96$; $\mu_y=69.5$, $\sigma_y^2=32$; $\rho=0.746$

—〈第 2 章〉—

- p. 17 問 赤玉を R_1, R_2, R_3 , 白玉を W_1, W_2 と区別して
 $S = \{(R_1, R_2), \dots, (R_2, R_3), (R_1, W_1), \dots, (W_1, W_2)\}$
- p. 18 問 1 $\frac{{}_5C_4 \times 6 \times 5}{6^5} = \frac{25}{1296}$ 問 2 $\frac{{}_3P_3}{{}_4P_4} = \frac{1}{4}$ 問 3 $\frac{1}{8}, \frac{3}{8}, \frac{3}{8}, \frac{1}{8}$
 問 4 $\frac{{}_6C_1 \times {}_4C_1 \times {}_3C_1}{{}_{13}C_3} = \frac{36}{143}$ 問 5 $\frac{8}{{}_{10}C_3} = \frac{1}{15}$ 問 6 $\frac{8}{{}_{10}P_3} = \frac{1}{90}$
 問 7 $\frac{{}_{13}C_5 \times {}_4C_1}{{}_{52}C_5}$
- p. 22 問 (1) $\frac{91}{100}$ (2) $\frac{25}{100}$ (3) $\frac{44}{100}$ (4) $\frac{64}{100}$
- p. 24 問 1 (1) $\frac{{}_{90}C_{10}}{{}_{100}C_{10}}$ (2) $\frac{{}_{90}C_8 \times {}_{10}C_2}{{}_{100}C_{10}}$ 問 2 $1 - \frac{{}_7C_3}{{}_{10}C_3} = \frac{17}{24}$
 問 3 甲, 乙, 丙が勝つという事象をそれぞれ B_1, B_2, B_3 , n 回目に始めて表が出るという事象を $A_n (n=1, 2, \dots)$ とすると,

$$B_1 = \bigcup_{n=1}^{\infty} A_{3n-2}, \quad B_2 = \bigcup_{n=1}^{\infty} A_{3n-1}, \quad B_3 = \bigcup_{n=1}^{\infty} A_{3n}.$$
 これより, 甲は $\frac{4}{7}$, 乙は $\frac{2}{7}$, 丙は $\frac{1}{7}$
- 問 4 (1) $\frac{5}{18}$ (2) $\frac{1}{36}$ (3) $\frac{17}{36}$
- 問 5 $P(X=0) = \frac{1}{32}, \quad P(X=1) = \frac{5}{32}, \quad P(X=2) = \frac{10}{32}, \quad P(X=3) = \frac{10}{32},$
 $P(X=4) = \frac{5}{32}, \quad P(X=5) = \frac{1}{32}$
- 問 6 $P(X=10) = \frac{1}{12}, \quad P(X \leq 5) = \frac{5}{18}, \quad P(X > 9) = \frac{1}{6}$
- p. 25 問 7 $\sum_{r=0}^2 {}_{10}C_r \left(\frac{2}{100}\right)^r \left(\frac{98}{100}\right)^{10-r}$ 問 8 $\frac{5}{12}$
- p. 28 問 1 $P(X=0) = P(X=5) = \frac{1}{32}, \quad P(X=1) = P(X=4) = \frac{5}{32},$

$$P(X=2)=P(X=3)=\frac{10}{32}$$

$$F(x)=\begin{cases} 0 & (x<0) \\ \frac{1}{32} & (0\leq x<1) \\ \frac{6}{32} & (1\leq x<2) \\ \frac{16}{32} & (2\leq x<3) \\ \frac{26}{32} & (3\leq x<4) \\ \frac{31}{32} & (4\leq x<5) \\ 1 & (5\leq x) \end{cases}$$

問2 $P(X\leq 2)=F(2)=\frac{3}{5}$, $P(1<X\leq 3)=F(3)-F(2)=\frac{2}{5}$

問3 (1) $P(X\leq \alpha)$ (2) $P(\beta<X\leq \gamma)$

p. 35 問 定理 2.10 と同様.

p. 40 問1 2つのサイコロの目を表す確率変数をそれぞれ X_1, X_2 とするとき, 目の和 $X=X_1+X_2$, 目の積 $Y=X_1\cdot X_2$, かつ X_1 と X_2 は独立.

これらより $E(X)=7$, $\sigma^2(X)=\frac{35}{6}$

$$E(Y)=\frac{49}{4}, \sigma^2(Y)=79\frac{139}{144}$$

問2 $E(X)=1$, $\sigma^2(X)=1$

p. 48 問1 $M(\theta)=\left(\frac{1}{2}e^\theta+\frac{1}{2}\right)^3$, $E(X)=\frac{3}{2}$, $\sigma^2(X)=\frac{3}{4}$

問2 $\theta < a$ の時, $M(\theta)=\frac{a}{a-\theta}$, $E(X)=\frac{1}{a}$, $\sigma^2(X)=\frac{1}{a^2}$

——〈第 3 章〉——

p. 54 問1 $E(X)=\sum_{k=0}^{\infty} k \frac{m^k}{k!} e^{-m} = m e^{-m} \sum_{k=1}^{\infty} \frac{m^{k-1}}{(k-1)!} = m$

$$\begin{aligned} \sigma^2(X) &= \sum_{k=0}^{\infty} k^2 \frac{m^k}{k!} e^{-m} - m^2 = \sum_{k=1}^{\infty} k \frac{m^k}{(k-1)!} e^{-m} - m^2 \\ &= m e^{-m} \left\{ \sum_{k=1}^{\infty} (k-1) \frac{m^{k-1}}{(k-1)!} + \sum_{k=1}^{\infty} \frac{m^{k-1}}{(k-1)!} \right\} - m^2 \\ &= m e^{-m} (m e^m + e^m) - m^2 = m \end{aligned}$$

問2 100個の中に含まれる不良品の個数を表す確率変数を X とすると, 求める確率は $P(X\geq 1)$.

$$\therefore P(X \geq 1) = 1 - P(X = 0) = 1 - {}_{100}C_0 (0.002)^0 (0.998)^{100}$$

$$\doteq 1 - \frac{(0.2)^0}{0!} e^{-0.2} = 1 - 0.8187 = 0.1813$$

- p. 59 問 1 (1) 0.135% (2) 2.275% 問 2 2.275%, 19.741%
 p. 63 問 1 0.00135, 0.226627
 p. 64 問 2 標本平均 4.38, 標本分散 0.1773
 p. 66 問 1 (1) 0.001070 (2) 0.177016
 p. 68 問 2 (1) 0.079270 (2) 0.5 (3) 0.002327
 問 3 0.000736, 0.944083 問 4 略
 p. 69 問 5 0.000224 問 6 0.015778

——〈第 5 章〉——

- p. 98 問 1 $0.10261 < \mu < 0.10339$ (mm) 問 2 $32.76 < \mu < 38.04$ (歳)
 問 3 $\bar{x} = 7.845$, $s = 0.388$
 信頼度 95% のとき, $7.7912 < \mu < 7.8988$
 信頼度 99% のとき, $7.7742 < \mu < 7.9158$
 問 4 $\bar{x} = 11.6$, $10.811 < \mu < 12.389$ 問 5 $0.33755 < \mu < 0.76645$
 問 6 $\bar{x} = 14.525$, $s = 0.663$ $13.9706 < \mu < 15.0794$
 p. 102 問 1 $0.4038 < p < 0.4362$ 問 2 $0.126 < p < 0.274$
 問 3 $0.2102 < p < 0.3818$ 問 4 $0.038 < p < 0.308$
 問 5 $0.509 < p < 0.913$
 p. 105 問 1 $-2.604 < \mu_A - \mu_B < 6.204$ (点) 問 2 $-6.896 < \mu_A - \mu_B < -0.104$
 問 3 $1.8571 < \mu_A - \mu_B < 3.1429$
 p. 106 問 4 $2.042 < \mu_A - \mu_B < 4.758$
 問 5 $\bar{x}_A = 1495$, $9s_A^2 = 191588$, $\bar{x}_B = 1093$, $9s_B^2 = 52720$
 $292.59 < \mu_A - \mu_B < 511.41$
 p. 107 問 1 $-0.1647 < p_A - p_B < -0.0353$ 問 2 $-0.01355 < p_A - p_B < 0.19695$
 p. 108 問 1 $67.286 < \sigma^2 < 311.929$ 問 2 $0.0730 < \sigma^2 < 0.6916$

——〈第 6 章〉——

- p. 113 問 1 全国平均より高い。
 問 2 この日の製品が特に悪いとはいえない。
 問 3 会社側の公表を否定できる。
 問 4 全平均は 12 オンスより少ないとはいえない。
 問 5 母平均は 3.26 であるを否定できない。
 p. 116 問 1 A 案に賛成するものは半数以上である。

- 問2 男女の出生率が異なるとはいえない。
- 問3 20% 以上, 50% 以下とは限らない。
- 問4 20% 以上, 40% 以下とは限らない。
- p. 117 問1 母分散に変化があったとはいえない。
- 問2 バラツキに差があるとはいえない。
- p. 120 問1 (危険率 5%) 母平均に差異がある。
(危険率 1%) 母平均に差異があるとはいえない。
- 問2 母平均に差異がある。
- 問3 午前と午後の製品の強さに差が認められる。
- 問4 平均寿命に差があるとはいえない。
- p. 122 問1 成績に差があるとはいえない。
- 問2 平均時間に差があるとはいえない。
- p. 123 問1 支持者が増したといえる。
- p. 124 問2 男子と女子で, 8 時間以上睡眠する人の比率が異なるとはいえない。
- p. 126 問1 9 : 3 : 3 : 1 に適合していないとはいえない。
- 問2 $m=2$ のポアソン分布に従っていないとはいえない。
- p. 130 問1 両科目の成績に関係がある。
- 問2 技量に差異があるとはいえない。
- p. 131 問3 男女の批評に差異があるとはいえない。
- 問4 A, B, C, D の薬品の効力に差が認められる。
- p. 132 問5 両工場に差異があるとはいえない。
- 問6 A, B の薬効に差異があるとはいえない。

索引

ア行

 F -分布 78, 80 F -分布表 80

カ行

回帰係数 13

回帰直線 13

階級 3

 χ^2 -分布 74 χ^2 -分布表 74 χ^2 -分布の再生性 74

確率 15

確率分布 26

確率変数 25

確率密度 26

確率密度関数 27

仮説 110

仮説の検定 109

片側検定 111

片側危険率 111

加法定理 19

ガンマ関数 71

危険率 93, 110

記述統計学 1

基本事象 17

帰無仮説 110

級心 3

共分散 10

空事象 16

区間推定 92

経験的確率 18

サ行

 3σ の定理 8

試行 15

事象 15

自由度 n の χ^2 -分布 74自由度 n の t -分布 77自由度 m, n の F -分布 80

周辺分布 10

乗法定理 20

信頼区間 93

信頼係数 93

信頼限界 93

信頼度 93

推測統計学 1

推定量 91

数学的確率 17

スターリングの近似式 77

正規分布 34, 55

正規分布曲線 55

正規分布の再生性 57

正規分布表 56

積事象 16

全事象 16

相関係数 10

相関図 10

相関表 8

相対度数 3

タ行

大数の法則 18, 43

第1次検定 122

第2次検定 122

第1種の誤り 110

第2種の誤り 110

単純事象 15

チェビシェフの不等式 7, 41

中心極限定理 64

 t -分布 75 t -分布表 77

適合度の検定 124

データ 2

点推定 92

統計量 59

等分散の検定 120

等平均の検定 117

特性値 1

独立 31

独立事象 20

独立性の検定 127

度数 2

度数多角形 4

度数柱図表 3

度数分布 3

度数分布曲線 4

度数分布表 2

ナ行

二項分布 49

二次元確率密度 28

二次元確率密度関数 29

二次元分布 28

ハ行

排反事象 16

ヒストグラム

標準正規分布 55

標準偏差 5

標本 1, 59

標本空間 15

標本標準偏差 60

標本分散 60

標本分布 80

標本平均 60
比率差の推定 106
比率の検定 114
比率の推定 98
不偏推定値 92
不偏推定量 92
分割表 127
分散 5, 35
分布関数 26, 71
平均 5, 35
ベイズの定理 22
ベータ関数 71
ポアソン分布 52
母集団 1, 59

母分散 59
母分散の検定 116
母分散の推定 107
母平均 59
母平均差の推定 102
母平均の検定 110
母平均の推定 93

マ 行

無限母集団 35
モーメント母関数 45

ヤ 行

有意水準 93, 110

有限母集団 35, 59
余事象 16

ラ 行

ラプラスの定義 17
ラプラスの定理 66
離散的確率変数 25
両側危険率 111
両側検定 111
連続的確率変数 27

ワ 行

和事象 16

著者略歴

いのまた せいじ
猪股 清二

昭和26年 東北大学理学部数学科卒
現在 玉川大学・教授

すずき ひろゆき
鈴木 皓之

昭和45年 東京都立大学大学院理学部修士課程修了
現在 玉川大学・助教授

せき や あきら
関谷 昭

昭和27年 東北大学理学部数学科卒
現在 関東学院大学・教授

ほった かずひさ
堀田 和久

昭和46年 東京都立大学大学院理学部修士課程修了
現在 関東学院大学・助教授

数理統計学

© 1982年3月15日 第1版第1刷発行
1992年12月1日 第1版第6刷発行

[検印省略]

	猪 股 清 二
著 者	鈴 木 皓 之
	関 谷 昭
	堀 田 和 久
発行者	早 川 征四郎
印刷者	田 中 昭 三

発行所 開成出版株式会社

郵便番号 112 東京都文京区関口1-27-10
電話 東京 (03) 3232-7281 振替 東京0178378

印刷 (株)理想社/製本 加瀬製本/取次 鈴木書店