

経 営	7 3
測 定	2 2

度 数 曲 線 と 相 関

W. P. ELDERTON

昭 和 3 5 年 3 月

農 林 省 林 業 試 験 場 経 営 部

序

統計学は最近めざましい発展をしつつ色々な産業界や学問の分野に應用されているが、データを統計的にとりまとめる場合、最も基本的なものは度数分布である。度数分布をうまく表現するかどうかによってその問題の統計的な解決は大きく左右されるものである。本書はこの度数分布とそれに関連した問題を確率論的に多数の数値計算をも含めて極めて要領よくとりまとめた古典的名著 "*Frequency curves and Correlation*" W.P. Elderton, 1953, *Forth Edition* の全訳である。原著の初版は1906年に保険会社等で死亡率表の補正法に K. Pearson 系の度数曲線を使おうとする試みに企図出版された。その後この研究により1927、1938年にそれぞれ改訂発刊され、新しい章がいくつも加えられて1953年に第4版まで出版されている。本書の対象は初等統計学かあるいは極く初等的な数字になじんでいる読者が予想されているが、いくつかの細部にはかなり高い数学的な方法も使われているが、訳山の例題と計算例をみればおのづとそれが理解出来るようになっている。また本書での数値例は保険統計からとられたものであるが、生物統計や社会統計においても全く同じように取扱える事を著者は指摘しているが、本書での方法は林業での諸問題にも充分に應用出来てその解決のために有力な武器となるであろう。

なおこの訳出は当研究室貝川端幸蔵の原稿に手を加えたものである。

昭和35年 / 月

測定研究室長 大友栄松

目 次

第 I 章	序 論	1
第 II 章	度数分布	5
第 III 章	モーメント法	13
第 IV 章	<i>Pearson</i> 系の度数曲線	43
第 V 章	計 算 法	59
第 VI 章	色々な曲線系の比較	137
第 VII 章	相 関	153
第 VIII 章	理論的分布、擬相関	170
第 IX 章	定量的測定不可能な特性値の相関	184
第 X 章	標準誤差	197
第 XI 章	適合度の検定	217
第 XII 章	相関比 — <i>Contingency</i>	229
第 XIII 章	偏 相 関	248
附 録 I	モーメントの修正	254
II	B 函数と Γ 函数	261
III	正規曲面の方程式	265
IV	正規誤差曲線に関する式の積分法	269
V	曲線適合の他の方法	275
VI	保険用語とその記号の手引	280
VII	本書の抄録	285
VIII	参考文献、その他	286
IX	数 表	293
	曲線型の図	300

第 I 章 序 論

§ 1 確率論の普通の取り扱いはある事象が起る可能性がわかって
いる仮定のもとに出発し、事象の組合せや特別な実験の繰り返し
に關して起る問題を解く方向に進められていく。すなわち、ある
結果は他の如何なる結果よりもその実験から起り易いものである
こと、また試行の数が限られた場合の結果は期待された結果とは
大きく異なつたように思われる。また最も起りそうな結果との同
の比例的な偏りは試行の数が増加するにつれて一般に減少して行
くこと等が証明される。

確率をあらかじめ正確に推定出来るときに、理論的方法は實際
に實現出来るような結果に導くということを実験により容易に示
すことができる。例えば錢投げで色々な試行を行えば、その中で
次の様なことが発見出来るであらう。もし5つの錢が一階に投げ
られ、その“表”の出た数を記録するなら、この場合の分布は吾
々が期待する普通の理論の振に二項展開 $(\frac{1}{2} + \frac{1}{2})^5$ に一致す
る。

“表”や“裏”の系列は公比 $\frac{1}{2}$ の幾何級数に近い級数を作りし
カード抜きは理論的計算に示す数によく一致した結果を与える。

§ 2 しかしながら確率が分っていない場合が往々起る。錢投げで
の“表”“裏”の系列或はカード抜きの様な実験を我々が取扱っ
ているかどうかを保証することは不可能である。

要するに分っているものは、あるグループの中に入る“場合の
数”の分析だけである。またこの様な事情では統計量が近似する
理論的系列を探り出す逆の問題も重要な問題になって来る。この
問題は統計量が正確に理論分布を支えないという理由と、理論的
結果と實際的結果の間にある差を見付ける事が不可能であると言
う理由で、その困難性が増大する。上述の關係をより明確にする
ために問題を言い直す方がよからう。即ち統計的実験から得られ

たような系列が近似している理論的系列を発見する事が可能であるかどうかを問題にした方がよからう。

或る与えられた場合に應ずる確率をみつける事は困難であり、恐らく不可能かも知れない。然し実験を無制限に繰返すことが可能ならば、そのとき得られた数の系列の論理的な評価をすることは実行できるであろう。

それに統計量が近づいて来る。この理想的結果を見出す事を推奨する理由を考えてみると、正確な基礎的な確率は余り大切なものではなく、系列を合理的に表わすことの方が遙に實際的な価値を持っていると言う事が分るであろう。

統計学者や統計の仕事に携っている保険数理士にとって、集められた数値から實際的結論が容易に引き出せる様な簡単な形に觀察値を表わすことが、先づ初めの目的のうちの1つである事を注意しておこう。

若し利用出来る統計量が自然に50或は60のグループに属しているとすれば、彼が携っているような問題の重要な特色を導き出すためにその統計量をどんな風に排列したらよいかを決定せねばならない。然るに全体に対して一つの指標として用いうるもとの系列も密接な関係がある二三の数値を見付ける事が出来れば同じ様な統計量の比較にある点を助けになる様な結果を与える事が出来るし、また事実を扱わねばならない他の人にとっても、もとのままの形で評価出来るよりも、よりたやすく全体の分布を評価する事を可能ならしめる。

統計家は、彼の経験から得られたものが二三に過ぎないときには、中間項の値の近似値を補充しなければならないし、また、その一部分だけが分っているときには相続く系列或は完全な系列の値を補わねばならない。

多くの場合、もとの系列と同じ項数を採らねばならないが、"場合の数"の限界のための資料の不調和を取り除くことは研究にとって有効なことである。即ち彼はデータを組分けしなければな

らない。

§ 3 実はいこれらの目的は非常によく似ている。というのは統計数値表が代数的に或は超越函数公式で表わされていれば、若し我々が出逢っている分布を組織的に取扱えば相互の比較も容易に出来るし、足りない項を補う事も出来るような二三の数値(公式の中の定数)で莫大な数の系列を置き換えられる。

一方もとの資料の不調和は出来るだけ近似的にもとの統計量を表わす適当な公式を作る事で取り除くことが出来る。

公式が理論的考察に立脚していれば本章の始めで述べた確率に関する解答も与えられるし、實際的且つ理論的必要条件の両方が同時に取り上げられる事も分る。というのは理論研究者にとって調べられたスムーズな系列は實際の仕事に必要な公式と同じものであるからである。

§ 4 曲線系の利点は公式の簡単化で、満足に取扱う事が出来る觀察値の組の数による。何故ならば複雑な表わし方は統計量のもとのグループを殆ど改善しないからである。

また一般に應用不可能なシステムはそれを分析するときは何時でも統計家を困難にするからである。

もう一つ別の事が必要である。即ち若し公式が適当なものである事がわかったら、ある特別な場合によく一致する算術的常数を見付ける幾つかの方法がなければならない。その様な方法はそれが実際に使えるものであれば簡単且つ確実で、一般に組織的に應用出来るものでなければならない。

この目的の大体の意図を達するためには、あらかじめはっきりした意見をもっていなければならない。この事はこの問題と必然的にむすびついた詳細なことが少いために忘れがちである。また系統的取扱いの利点は往々見落されるから、このことは重要である。そして、簡略法や概略の方法、また平準の方法を採用すると

仕事が無効になる。

また科学的な根拠もなく似た様な場合に、他と適当な関係を持たない公式が統計家等によって、むしろ偶然的なやり方で往々使われている。

その結果は一般化が不可能となつてきてある法則が見出だされる場合にでも活動的計算者が何回も計算してみない限り、比較出来る統計量の価値を考えないで自分自身の結論をうる事は殆どないという事がわかる。

第 II 章 度 数 分 布

§ 1 統計量が回数を表わす様に並んでいるか、或る特定の点で起る事象の頻度数を示すように並んでいれば、そのときの並び方が度数分布である。若し我々の結果の幾つかが広く応用出来る可能性があつても、我々はこれらの度数分布に対する注意は一般に限りがあるであらう。

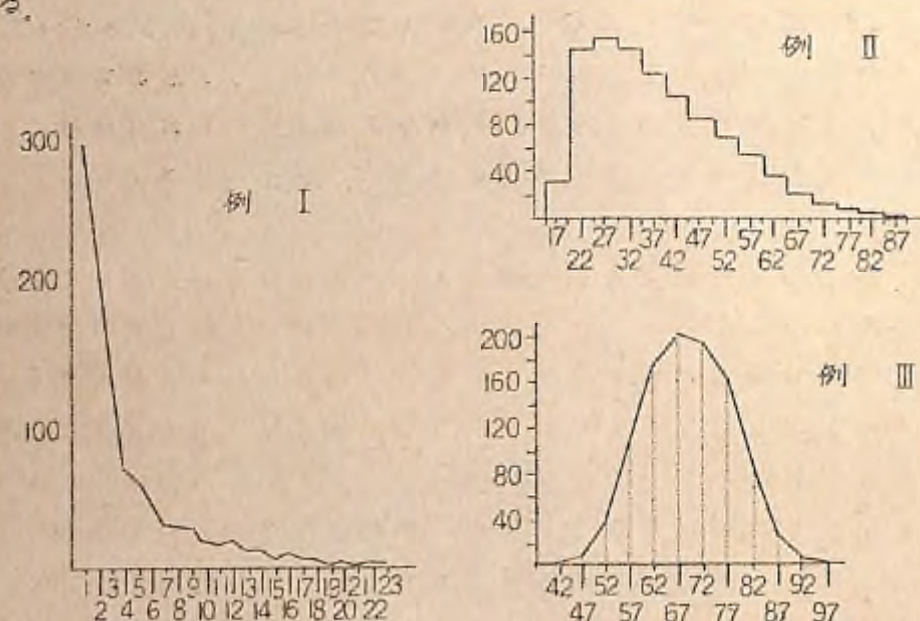
§ 2 このような度数分布を表わすために用いられている公式の名称を知る事が必要である。“度数曲線 (frequency curve)” という語がこの目的に用いられている。錢投げやサイコロ投げのようなある直接の実験での系列の数を表わす幾何級数は極限に於て一つの度数曲線で、その方程式は $y = Na^x$ である。

§ 3 ある分布は独立変数の値をもつ。ある群 (group) に属する場合の数を与える。しかし他方 (表 I の例 V) 独立変数の正確な値に対する場合の数を与える。前者の場合には組と対応する独立変数の正確な値を考えねばならない。例えば“年令 x で死の危険にさらされた”⁽¹⁾ というのは、 $x - \frac{1}{2}$ と $x + \frac{1}{2}$ の間に含まれる死の危険にさらされた数であるが、期間 x での死亡数というのは $x - \frac{1}{2}$ と $x + \frac{1}{2}$ の間に含まれる死亡数である。統計量がグラフ的に表わされているときには、これらの違いに対する効果があたえられまた、より明確に僅かの相異点を引き出すために、 x 頁と y 頁の図がつくられている。これらの図のような分布図は度数多角形またはヒストグラムと呼ばれている。

§ 4 統計量が独立変数の正確な値に対する数を与えているときは、単に図の中に縦座標をプロットし、それらの頂点を連絡する。しかし独立変数の値がグループの場合にはやや複雑になる。何とすれば、我々は区間全体の巾をもつてその (図の例 II) 上に固定さ

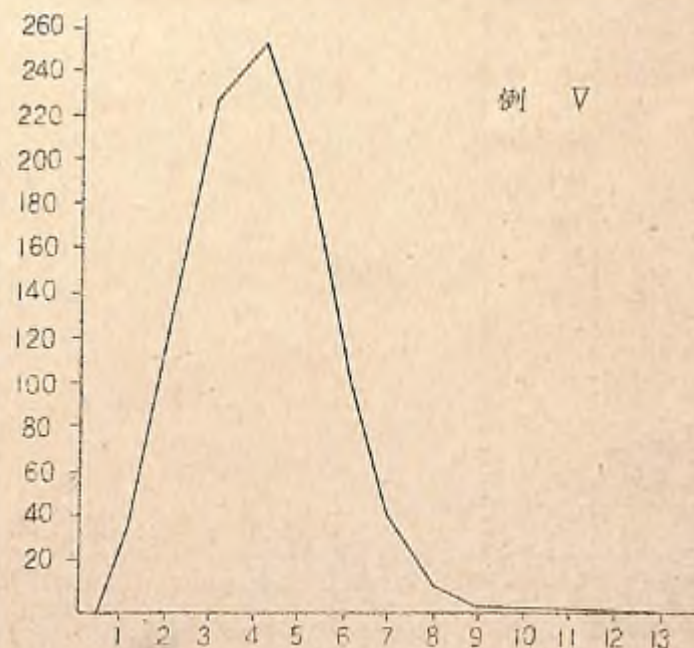
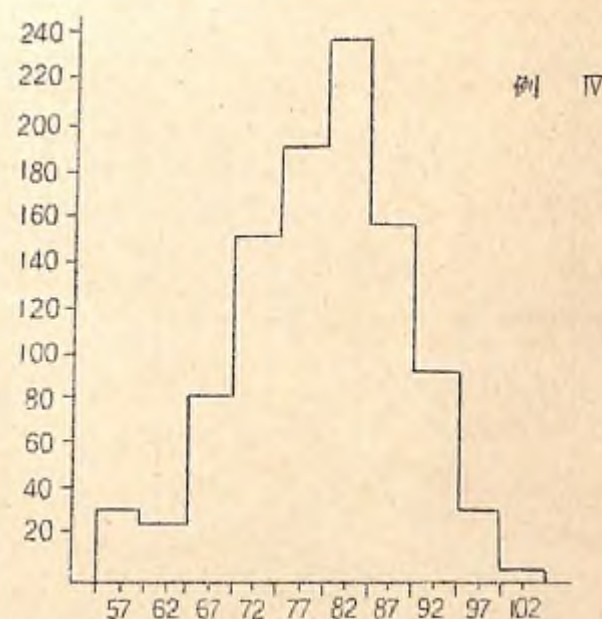
の区間の長さを

れた矩形を作るか、或は各区間の中点で縦線を引き、それからそれらの頂点を連結する(例Ⅲ)のいずれかが出来るからである。前者の方法は統計量によって伝えられる情報量の正確な考え方を与えるが、或る目的(例えば曲線の可能な型をみつけるため)に対しては後者の方がより便利である。けれども、それは技術的な問題がある。例ⅠとⅣのような場合は、与えられている図の種類によって最上に表わされたものである。然し例Ⅲは技術的障害を受け易いが、度数矩形図より実際分布の型をよりよく表示している。



§5 統計量は場合の数の合計が増すにつれてスムーズな系列の方に近づいて行く事實は、読者は既に知っていると思う。この事から、若し有る無限の資料が研究に利用出来るならば、それから得られるスムーズな分布を表わす度数曲線の概念を如何にして通常の実際統計量に導くかを視ることが出来る。

換言すればその様な曲線は調査された特殊の場合が標本であったその全国集団の一つの近似を与えるのである。



I 表

例	I	例	II	例	III	例	IV	例	V
curate duration	Withdrawals with monthly incidence "0" in year of exit (Principles and meth- ods P.92)	Ages	Exposed to risk of sickness (Watson M.U. Tables P.19)	Existing at close of Observations Without Profit "OLD" Assurances	Existing at close of Observations "OLD" Annuities (females)	Terms of the expa- nsion of $1000(\frac{3}{4}+\frac{1}{2})^x$	No. of term		
1	308	- 19	34	...	...	32	1		
2	290	20-24	145	...	...	127	2		
3	118	25-29	156	...	...	232	3		
4	69	30-34	145	...	...	258	4		
5	59	35-39	123	...	...	194	5		
6	44	40-44	103	3	...	103	6		
7	29	45-49	86	9	...	40	7		
8	28	50-54	71	42	...	11	8		
9	26	55-59	55	111	29	2	9		
10	21	60-64	37	176	23	1	10		
11	18	65-69	21	200	81	...	11		
12	18	70-74	13	193	151	...	...		
13	12	75-79	7	160	192	...	...		
14	11	80-84	3	73	239	...	...		
15	5	85-89	1	26	157	...	...		
16	11	90-94	...	6	93	...	...		
17	7	95-99	...	1	29	...	...		
18	6	100-	...	...	6	...	...		
19	1	...	...	...	...	...	...		
20	3	...	...	...	...	...	...		
21	1	...	...	...	...	...	...		
22	3	...	...	...	...	...	...		
23	2	...	...	...	...	...	...		
...	1,000	...	1,000	1,000	1,000	1,000	...		
真合計	1,308	...	2,995.724	2,674	172	...	...		
平均値	4.182	...	37.8750	68.485	79.400	3.998	...		
標準偏差	4.1996	...	276.210	1,771.288	1,774.394	1.46215	...		
曲線の型	I	...	I	II	VII	...	...		

§ 6 度数曲線は分布の全範囲の間で独立変数のすべての値に対応する頻度数を与えると解釈することが出来る事は注意しておこう。また度数曲線は実際の統計量に必要であるというような、多少任意的組合せにのみ限定するものではない。二項級数と幾何級数は我々が非常に多くのグループに分けられるものの取扱いを考えるときには、全く同じものとなる。これは若し二色の砂の大きな量を混合して、その混合物のきまった量を取り出し、各々の中のいづれかの色の砂の粒数を記録するならば、我々は非常に大きい数の試行から連続曲線を得るであろう。

§ 7 さて次に、重要な函数を定義しよう。分布が変数の特性値例えば期間、年齢等の累進値に従って並べられているときには、特性値の平均値(度数の平均ではない)は、分布の平均値(mean)と呼ばれている。それは

$$\frac{f_a \times a + f_b \times b + f_c \times c + \dots + f_n \times n}{f_a + f_b + f_c + \dots + f_n}$$

を与えられる。

ここで、 f_r は r 番目の変数の値に対応する度数である。だから例 I において 200 は 2 に対応する度数である。若し我々が微小な増分を仮定すれば、平均は

$$\frac{\int f_x \times x dx}{\int f_x dx}$$

によって与えられる。ここで積分の両端は全分布を覆うような範囲となる。また平均値は分布の重心(gravity)を通る縦線の位置のようにも言い表わすことができる。この事は或る読者には助けとなるかも知れない。

§ 8 モード(mode)は最も頻繁に起る特性値である。換言すればモードは縦線の最大の位置でもある。縦線が最大であるようなラフな統計量からは、何ら告げることができないし、従ってモードは色々な群に關係した法則。例えば度数曲線が分つて始めて近似的

的に決定できらにすぎない。

§9 さて次に方程式或は曲線はいろいろな分布、例えば年齢によつて与えられる分布、或は期間に従う色々の問題の分布や保険金の合計に従う分布等々に用いてもよいので、我々はそれら自身の分布に基いた関係の標準を持たねばならない。これらの目的には標準偏差 (standard deviation) として知られている函数が用いられる。それは、

$$\sqrt{\frac{f_a a'^2 + f_b b'^2 + \dots + f_n n'^2}{f_a + f_b + \dots + f_n}}$$

によつて与えられる。ここで $a', b', \dots, n'$ は平均値からのへだたりである。積分型による標準偏差は、

$$\sqrt{\frac{\int f_x x^2 dx}{\int f_x dx}}$$

によつて与えられる。ここで x は、平均値から測られる。

標準偏差は測定単位による度数のちらばり方を測定する。平均値から最も離れた度数は、最も大きな x の値が掛けられるから、大きな標準偏差は平均値から外に広げられた度数分布を示す。しかし小さな標準偏差は度数が平均値のまわりに集中していることを示している。標準偏差の相対的大きさを考慮する場合には、測定単位を心にとめる必要がある。何となれば、もし与えられた分布が二つの系列、即ち一つは年齢に従った系列であり、他は5年毎の年齢群に従って纏められていれば、標準偏差は前者よりも後者の方が5倍も大きい。このことは二つの表わし方、

$$\sqrt{\frac{\int f_x x dx}{\int f_x dx}} \quad \text{と} \quad \sqrt{\frac{\int f_x (5x)^2 dx}{\int f_x dx}}$$

の比較によつて直に視ることが出来る。後者は前者の5倍であることは明らかである。標準偏差の値は各々の場合に対して表Iに与えられている。12頁の図は同じ平均値 B とほぼ同じ面積を持つ二つの曲線を示している。然し、点線の曲線は大きな標準偏差

を持っている。何故ならばこれは平均値の両側に実際よりもひろがって出ている。読者は上に与えられた代数式から平均値、モード標準偏差は“場合の数” (例えば曲線の絶対的大きさに因する) に依存するものでなく、これらは単にそのばらつき方にのみ依存する (例えば曲線の型或は割合の数に因する) ということがわかるであろう。標準偏差は平均値から統計量の“広がり” (spread) あるいは“ちらばりかた” (scatter) を測るものである。

§10 度数分布を調べるべく 8頁のI表をみよ) 大抵の場合のから始められ、徐々に最大値に昇りそれから全く異った割合で下降することになる。

若し上昇と下降が同じような割合であるならば、分布は平均のまわりに対称であろう。その平均値はモードに一致しなければならない。平均値とモードの間の差は、対称性からの距離或は“歪度” (skewness) の函数である。満足な尺度を得るために重要な拡張を考慮しなければならない。これは平均値とモードの間の差を標準偏差で割って歪度を測る事ができる。統計量を図で表わしたとき、平均値がモードの左側にあれば、この函数は負でありそして符号に注意すれば次のように書くと便利である。

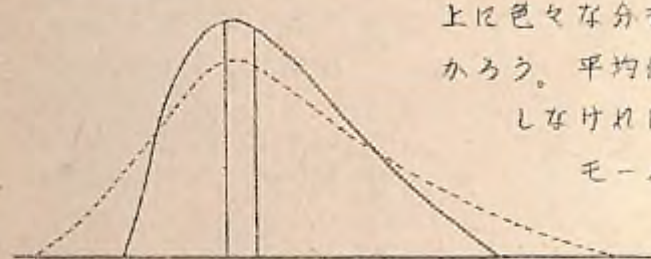
$$\text{歪度} = \frac{\text{平均値} - \text{モード}}{\text{標準偏差}}$$

12頁の図は歪に対する尺度の割合を示す助けになるだろう。この図は同じ平均値と同じモードを持っているが、異なる標準偏差を持っている二つの曲線を与えている。大きな標準偏差を持っている曲線 (点線) は実際の曲線よりも対称的であることは明らかである。

§11 横軸上に平均値及びモードが軸上の曲線の位置を固定することによつて、これらの函数を要約できよう。

標準偏差は資料がどのように平均値の周りに分布しているかを

示す。歪度は資料に現われている対称性からのへだたりの量を示す。これら予備的定義は我々の最初の目的には充分であろうが、この定義された函数は曲線適合の実際的な仕事に關するその実状を學んだときに、もっとよく理解出来るであらう。始めてこの問題を學ぶ學生は分布の性質や外觀になじむために、グラフ用紙の



上に色々な分布を図示してみるがよ
からう。平均値を計算して図に押入
しななければならない。しかし、
モーメント法について一通
りの事を學ぶまでは、
標準偏差の計算を

試みてはならない。

§12 今までの点に關して我々は度数として統計量を定義してきた。すなわち、例Ⅰの意味でのように實際的な数であるか、或は例ⅡⅢⅣのように比較的近いグループの中に入っている統計量を一桶にした“場合の数”として度数を定義したのであった。

しかしながら、我々は経験を表示するときは、個々の觀衆値を取扱ねばならず、それらは次々に群分けされている。この点からもし N 個の觀衆値があるならば、それらを $O_1, O_2, \dots, O_N$ とよぶことにする。ここで O_1 は最初の觀衆値を表わす。そして例Ⅲをみよう。65~69の群において200ヶ存在するものの中の一つであるかもしれない。それは、年令 66.12で生存する場合であつてもよい。また O_2 は年令 72.72で生きているものの数であり、 O_3 は年令 82.26で生きているものの数、 O_4 は 87.37才で生きているものの数である等。そのとき平均値は、

$$\frac{1}{N} \sum_{i=1}^N O_i$$

である。

第Ⅲ章 積 率 法

§1 度数曲線として使うための適当な形を取扱っていく前に、統計的例題に度数曲線を適用する幾つかの方法を見付けることができるかどうかを祝む方がよからう。と言うのは、曲線を示すことは有用でない事は明らかであるし、またそれを用いる慣習もないからである。

それ故に我々は与えられた公式が或る特定の統計的経験に適合され、また我々が補整の基礎として決定してもよいような、如何なる表わし方（例えば死亡率に対する *makeham* の公式）にも応用されうる一般的な方法が必要である。一つの方法を見出す場合に注意しなければならない点の、若し公式の中に n 個の常数があれば、公式と統計量の間に n 個の方程式を作らなければならない。これは、若し $x = 1, 2, 3, \dots$ のときに $y = 20, 40, 88$ であり、それを表わすのに、曲線 $y = a + bx + cx^2$ を用いようと思ふなら、勿論その項が正確に次のような等式によって見出されるように a, b, c の値を求めることができる。

$$a + b + c = 20$$

$$a + 2b + 2^2c = 40$$

$$a + 3b + 3^2c = 88$$

$$a = 28, \quad b = -22$$

$$c = 14$$

しかし $x = 4$ のとき方々番目の項 $y = 96$ を持ち、今与えられた三つの方程式から見出された a, b, c の値を用いるならば、我々は $x = 4$ のときに $y = 164$ であることを発見するであらう。

これは方程式の中にある定数よりも、統計量における項の数が多くあるときには、方程式はそれらからの幾つかを選ぶ事なく凡ての項を用いる事によって形成されていなければならないことを示している。補整曲線は必ずしも觀衆値のどれにも正確に更正されるものではないが、しかし、それらの一般的傾向を表わすように觀衆された事実についての概略を一様に表わすものであらう。

§2 $a_1, a_2, \dots, a_n$ は、補整されるべき項であるとする。そのときもし系列が完全に消らなで、ある既知の法則に従っているならば、各項は正確に、即ち $b_1, b_2, \dots, b_n$ によって再現すること出来る。ここで $a_1 = b_1, a_2 = b_2, \dots, a_n = b_n$ さてもし二つの系列 (a の系列と b の系列) を考えれば、各項は正確に再現されるから

$$\sum_{r=1}^n a_r = \sum_{r=1}^n b_r \quad \text{と} \quad \sum_{r=1}^n C_r a_r = \sum_{r=1}^n C_r b_r$$

ということがわかる。ここで C_r は数係数である。これは各項が正確に再現出来ないときに適用されるべき可能な方法を示唆する。補整された数値の全合計は、補整されない全合計に等しくなければならない。また未知定数を発見するために必要な、これ以外の方程式は異なる項に色々な因子を掛けて、同時に補整されたものの積和と補整されない積和とは等しいとおくことによって作られなければならない。即ち

$$\sum C_r a_r = \sum C_r b_r$$

また C_r を与えるための最良の形を決定することが残っている。

$$\frac{a_1 + 2a_2 + \dots + n a_n}{a_1 + a_2 + \dots + a_n}$$

に等しい平均値は $C_r = r$ が合理的な方程式の一つを与えるであろうことを暗示する。更に我々は補整公式に適用したときに積分可能な形を与える r の或る函数を使用しなければならないから、(さもないと我々は $\sum C_r a_r$ と $\sum C_r b_r$ の間の方程式を立てられない) r の巾は部分積分を試みる時には便利なことがわかる。それ故にもし $C_r = r^t$ とかき、 t に値 $0, 1, 2, \dots$ を順次与えるならば、我々が必要なだけの方程式を得ることが出来る。そして、それらの最初の二つから次々に面積と平均値を発見し、それが補整数値と補整しない場合の数値において同じであることがわかる。

この方法はモーメント法(慣性モーメントと比較せよ)として

知られている。今までの経験から一つの実際の統計実験に曲線適合の万全な方法であることが示されている。理論的側面の確定は提出されたが、然し *Makeham* の死亡率の法則に関して保険数理によってなされたような特別の曲線の適合の他の方法の発明も可能であるが、よりよい一般的方法は提出されていない(附録7をみよ、他の方法に関する注意がある)。

§3 上に与えられた3つの方程式を解くための方法を応用するため、

$$(a+b+c) + (a+2b+2^2c) + (a+3b+3^2c) = 20 + 40 + 88$$

$$(a+b+c) + 2(a+2b+2^2c) + 3(a+3b+3^2c) = 20 + 2 \times 40 + 3 \times 88$$

$$(a+b+c) + 2^2(a+2b+2^2c) + 3^2(a+3b+3^2c) = 20 + 2^2 \times 40 + 3^2 \times 88$$

或は、

$$3a + 6b + 14c = 148$$

$$4a + 14b + 36c = 324$$

$$14a + 36b + 98c = 972$$

をとる。これらの方程式は、それが構成されているもとの方程式と全く同じ結果を与えるであろう。何故ならば、3つの項の各々は正確に補整出来るからである。然しながら、もし我々が4番目の項 $x=4$ 、 $y=96$ を導入するならば、我々は上に与えられた各方程式に4番目の項を加えることによって、モーメント法を修正することができて、

$$4a + 10b + 30c = 244$$

$$10a + 30b + 100c = 748$$

$$30a + 100b + 354c = 2508$$

を得る。これらの方程式の解は

$$a = -23.0$$

$$b = 42.6$$

$$c = -3.0$$

或は、

$x = 1$	$\bar{x} = 16.6$
$x = 2$	$\bar{x} = 50.2$
$x = 3$	$\bar{x} = 77.8$
$x = 4$	$\bar{x} = 99.4$

を与える。これは非常に簡単な例であるが、答えを採す方法を示すために多分助けとなるであろうし、次のための土台になるであろう。

§4 或る度数の n 次のモーメントは、その度数とそれがとられているまわりの重線から度数までの距離の n 乗の積として定義される。或いはモーメントが取られている点から、距離 x の点を通る重線の周りの度数曲線の任意の微線 $\bar{x}$ の n 次のモーメントは $\bar{x}x^n$ である。また微線の系列として扱われた全分布の n 次のモーメントは、 $\bar{x}_1 x_1^n + \bar{x}_2 x_2^n + \dots$ である。ここに $\bar{x}_1 + \bar{x}_2 + \dots$ は全度数である。それ故に、例Ⅶでは $\bar{x}$ を通る重線の周りで項 A に対する度数 40 の 3 次のモーメントは $40 \times (4 + 4^3)$ である。

§5 若し微線が知られているならば、モーメントが求められている所を通る重線と微線の間の距離の何乗かを度数に掛けることで即座にそれら微線のモーメントを計算することができる。そしてその結果を加えるときに距離を与えるためのそれら固有の符号に注意が必要である。若し面積が与えられていれば、それらの面積が立っている各区間上の中点で微線のまわりに度数が集中していることを仮定する事によって、近似値が作られる。表Ⅱの才 3 番目以降の列は、度数が基線の中点に集中しているとの仮定に立って表Ⅰの例Ⅳに対して、年令 A を通る重線の周りのモーメントの計算法を示している。

グループ分けの単位は5年毎に取られている。また往々便利なることもあるが、全度数が 1 であることを仮定すれば、全部の度数は 1000 で割っておかねばならない。我々は 1000 ケの分布とし

Ⅱ 表

年令グループ の中心値 x (1)	度 数 f (2)	$(x-\bar{x})/5$ $= S$ (3)	$f \times S$ (4)	$f \times S^2$ (5)	$f \times S^3$ (6)	$f \times S^4$ (7)
57	29	-4	116	464	1856	7424
62	23	-3	69	207	621	1863
67	81	-2	162	324	648	1296
72	151	-1	151	151	151	151
77	192	0	-498	---	-3276	---
82	239	1	239	239	239	239
87	157	2	314	628	1256	2512
92	93	3	279	837	2511	7533
97	29	4	116	464	1856	7424
102	6	5	30	150	750	3750
計	1,000	---	$\frac{+978}{+280}$	3,464	$\frac{+6,612}{+3,396}$	32,192

モーメントの記号

N = 度数合計

$\bar{v}_n$ = 平均値周りの n 次の未修正統計的モーメント

$\bar{v}'_n$ = 任意原点周りの n 次の未修正統計的モーメント

μ_n = 平均値周りに計算した曲線からの n 次モーメント

= 平均値周りに計算した n 次の修正統計的モーメント

μ'_n = 任意原点周りに計算した曲線からの n 次モーメント

= 任意原点周りに計算した n 次の修正統計的モーメント

注意 $\bar{v}$, $\bar{v}'$, μ はつねに 1 の度数合計に関係をしたものである。

表Ⅰに与えられているように、一般に生起する実際の数を取扱うのが普通である。しかしこの場合には今の方法を用いるのがより良いのである。

(3) 列にある数 -4 , -3 , -2 , $\dots$ はグループ単位によって割った年令 A からのへだたりを表わす。他のグループの中心も殆んど A の場合と同様に処理されている。分布の平均値近くは任意の原点を選ぶ事は便利である。これは平均値く往々求められた

結果)の周りのモーメントの計算を容易にする。また任意の原点のまわりのモーメントとそれらと比較して、これらのモーメントの概略の換算ができる。欄(4)~(7)は、それらの上に書いてある符号で充分説明される。それらは相次いで形成されていて、 z^k を f に束ずることによってチェックされる。 z^k の値は数表から得られる。

算術的計算は、他の方法をチェックされる。例えば z^k の適当な値を各項に束ずることにより、最後の欄のチェックの代りに、我々は新しい欄 $(z+1)^k f$ を作るができる。そしてこれは、

$$z^k f + k z^{k-1} f + \frac{k(k-1)}{2} z^{k-2} f + \frac{k(k-1)(k-2)}{6} z^{k-3} f + \dots$$

と同じものである。この新しい欄の合計は、加法と乗法のチェンクのために用いられる。(表IIの)数値例で、我々は $29 \times (-3)^4$, $23 \times (-2)^4$, 6×6^4 を計算する。この合計は 69240 である。これは次の方法で (2)~(7) 欄の合計に一致している。

$$\begin{array}{r} 32,192 \\ 4 \times 3,336 = 13,344 \\ 6 \times 3,464 = 20,784 \\ 4 \times 480 = 1,920 \\ \hline 1,000 \\ \hline 69,240 \end{array}$$

有用な表(乗束と4次のモーメントの表)は K. Pearson に よって出版された。Tables for Statisticians and Biometricians (Cambridge University Press) にみることができる。

私は以後この書を Tables for Statisticians と呼んでおこう。学生はこれらの書なしにも問題の処理はできるが、いくらか手間が要る。

§6 今までモーメントは、任意の点のまわりについて計算できることが仮定されてきた。然し、そうすることはしばしば不便である。何となれば、もし年令 29.4 のまわりのモーメントを求めた

いならば、我々は $\frac{(52-29.4)}{5}$, $\frac{(62-29.4)}{5}$ のようなものの束束を束束しなければならないであろう。その仕事が大変な労力を要することは全く明らかである。しかしながらそのような場合に我々は、より便利な他の点のまわりのモーメントを作り、それから次のような方法でそれを修正する。

さて A と B との間の距離を $+d$ とおく。ここで A はモーメントが知られている点で、B はモーメントを求めようとする点である。即ち、もし 25 のまわりでのモーメントを見出して 25.7 のまわりのモーメントを求めようと思うならば、 d は 0.7 である。もし 26 のまわりのモーメントを知っているならば d は -0.3 となるのである。そのとき、もし A から任意の距離 x_r までの距離が x_r で B からの距離が x_r であるならば、そのとき

$$x_r = X_r - d$$

$$\text{且つ} \quad x_r^n = (X_r - d)^n$$

今微線の系列として取扱った全分布の n 次のモーメントは、A のまわりで $\sum x_r X_r^n$ また B のまわりでは $\sum x_r x_r^n$ である。だから我々は、

$$\begin{aligned} V_n'' &= \sum x_r x_r^n = \sum x_r (X_r - d)^n \\ &= \sum [x_r \{ X_r^n - n d X_r^{n-1} + \dots + (-1)^n d^n \}] \\ &= V_n' - n d V_{n-1}' + \frac{n(n-1)}{2!} d^2 V_{n-2}' - \frac{n(n-1)(n-2)}{3!} d^3 V_{n-3}' \\ &\quad + \frac{n(n-1)(n-2)(n-3)}{4!} d^4 V_{n-4}' \dots \dots \dots (1) \end{aligned}$$

ここで V_n'' は B のまわりの n 次のモーメントを表わし、 V_n' は A のまわりの n 次のモーメントを表わすのである。

(1) の代りに我々は次のように進めてもよい。

$$\begin{aligned} V_n' &= \sum x_r X_r^n = \sum [x_r (x_r + d)^n] \\ &= V_n'' + n d V_{n-1}'' + \frac{n(n-1)}{2!} d^2 V_{n-2}'' + \dots \dots \dots \end{aligned}$$

$$V_n' = V_n'' + n d V_{n-1}'' + \frac{n(n-1)}{2!} d^2 V_{n-2}'' \dots \dots \dots (2)$$

これら二つの公式の間のどちらを選んでもよい。そして、勿論

二つは同じ結果を与える。

§ 7 我々は表Ⅱの例で、重心を通る重線（例えば平均値を通る重線）のまわりのモーメントを算出するためにこれを応用しよう。任意の点から平均値までの距りは、

$$\frac{\sum (X_r \bar{x}_r)}{\sum \bar{x}_r} = \frac{\sum (X_r \bar{x}_r)}{N}$$

である。ここに N は全度数である。或いは、任意の点から平均値までの距りは、その点を通る重線の周りの分布の一次モーメントであるということができる。重心を通る重線の周りの一次のモーメントは零であるから、もしそのようなモーメントが必要であれば、(2) の V'_1 を含む項は零であることがわかる。

我々が度数曲線を取扱う時に、重心を通る重線の周りのモーメントが一般に必要になり、それを表わす際にダツシユを取って V を用いる。

§ 8 算術的な計算は次のようである。(4) ~ (7) 欄の合計は観衆値の数 (2) 欄の合計) で割られている。その商は 77 の周りのモーメント (V'_1) である。モーメントは全度数が 1 である。即ち実際の度数でなく比率の度数を取扱っている場合に關係したように取扱われている。

$$V'_1 = 0.480$$

$$V'_2 = 3.464$$

$$V'_3 = 3.336$$

$$V'_4 = 32.192$$

V'_1 の値は平均年令 = $77 + 5 \times 0.480 = 79.4$ を与える。公式 (1) あるいは (2) を用いるために d の値が必要である。そのときモーメントの計算は、分布の重心を通る重線のまわりで計算されねばならず、その値は上に見たように V'_1 に同じである。この場合 V'_1 は年令 77 を通る重線のまわりの一次のモーメントである。 d の

巾束対数を使つて、その次に計算される。

d は比較的簡単な数であることが多いので、もしそれが 0.48327 であったら、対数を使うことの優れていることはもつとはっきりわかるであろう。

$$d^2 = 0.2304 \quad d^3 = +0.110592 \quad d^4 = 0.0530842$$

V'_1 は零で、 V'_0 と V_0 は全度数であるから、各々 1 であることに注意すれば、結局次をうる。

$$V_2 = 3.2336$$

$$V_3 = -1.430976$$

$$V_4 = 30.416289$$

小数点以下の桁の大きな数値を計算する方がよい。なぜなら、引き算が含まれている為に、始め 7 桁で始められても、結局 5 桁になってしまうからである。故に、7 桁の対数表 (例えば Chamber の対数表) や逆対数表 (例えば Filipowski の逆対数表) あるいは計算機が使われていることは良く知られている。

次々のモーメントの計算に必要な項は連続的に作ることもできるということを注意しておこう。即ち、公式 (1) で我々は V'_2 の次の束数を計算する必要がある。

$$\begin{array}{ll} \text{2 次モーメントに対しては、} & 1 \\ \text{3 次} & 3d \\ \text{4 次} & \frac{4 \cdot 3}{2} d^2 \end{array}$$

私は他の統計的計算のようにモーメントの計算にも系統だったシステムを採用し、また一定の位置に計算と結果を記入する習慣を作り出すと便利にすることがわかった。そうすると、往々複雑であるような算術的計算でも早く出来、また検算や訂正も容易にできる。

§ 9 上に述べたことはモーメントの計算に通常用いられる方法であるが、他の方法が故 G. E. Hardy 先生によって提案されて、British Offices Tables 1863~93 の補整法に用いられている。

波は統計的数値の和を取ること、 Nx 列が Dx 列から作られているときに、保険数理士達がやったような方法で新しい系列を作ること、これらの結果（ S 列参照）の和を取ることによって、モーメントの方法と同じ結果を与える方程式が作られることを指適している。表Ⅲの配列は、計算の方法と計算過程によって得られた式の両方が示されている。

最初の項に相対した行を考えると、系列の和が与えられており、それから、全度数を 1 とした場合に S_2 と呼ばれるオ二次和が $f(n)$ に対応する点の前の 1 なる距離にある重線同りの全分布の一次モーメントを与えるということがわかる。しかしながら最初の行だけ考えると S_3 は $\frac{n(n+1)}{2!}$ あるいは $\frac{n^2+n}{2!}$ 形の係数倍された各函数を与える。

即ちこれは $\frac{V_2+V_1'}{2}$ を与える。ここで V' はモーメントを表わしている。なぜなら、定義から全分布の n 次のモーメント (V'_n) はすべての n に対して $n^2 f(n)$ の和で与えられているからである。 S_4 と S_5 は各々 $\frac{n^3+3n^2+2n}{6}$ と $\frac{n^4+6n^3+11n^2+6n}{24}$ 倍された各函数を与える。

方程式の結果は次の如くである。

$$S_2 = V_1' \quad S_4 = \frac{1}{6}(V_3 + 3V_2' + 2V_1') \\ S_3 = \frac{1}{2}(V_2 + V_1') \quad S_5 = \frac{1}{24}(V_4 + 6V_3' + 11V_2' + 6V_1')$$

これらの方程式は、選ばれた原点のまわりのモーメントの計算に役立つものであるが、もし平均値のまわりのモーメントを求める必要があるならば、次の関係はより便利である。それらは、公式 (2) に上の値を代入して得られる。そして $S_2 = d$ ということに注意して、

$$V_2 = 2S_3 - d(1+d) \\ V_3 = 6S_4 - 3V_2(1+d) - d(1+d)(2+d) \\ V_4 = 24S_5 - 2V_3\{(2(1+d)+1)\} - V_2\{6(1+d)(2+d)-1\} \\ - d(1+d)(2+d)(3+d)$$

表 Ⅲ

行列	函数	函数の和	函数の 2 次和	函数の 3 次和	函数の 4 次和	函数の 5 次和
1	$f(n)$	$f(n) + f(n) + \dots + f(n)$	$f(n) + 2f(n) + \dots + n f(n)$	$f(n) + 3f(n) + \dots + \frac{n(n+1)}{2} f(n)$	$f(n) + 4f(n) + \dots + \frac{n(n+1)(n+2)}{3!} f(n)$	$f(n) + 5f(n) + \dots + \frac{n(n+1)(n+2)(n+3)}{4!} f(n)$
2	$f(2)$	$f(2) + f(2) + \dots + f(n)$	$f(2) + 2f(2) + \dots + (n-1)f(n)$	$f(2) + 3f(2) + \dots + \frac{(n-1)n}{2} f(n)$	$f(2) + 4f(2) + \dots + \frac{(n-1)n(n+1)}{3!} f(n)$	$f(2) + 5f(2) + \dots + \frac{(n-1)n(n+1)(n+2)}{4!} f(n)$
3	$f(3)$	$f(3) + f(3) + \dots + f(n)$	$f(3) + 2f(3) + \dots + (n-2)f(n)$	$f(3) + 3f(3) + \dots + \frac{(n-2)(n-1)}{2} f(n)$	$f(3) + 4f(3) + \dots + \frac{(n-2)(n-1)n}{3!} f(n)$	$f(3) + 5f(3) + \dots + \frac{(n-2)(n-1)n(n+1)}{4!} f(n)$
4	$f(4)$	$f(4) + \dots + f(n)$	$f(4) + 2f(4) + \dots + (n-3)f(n)$	$f(4) + 3f(4) + \dots + \frac{(n-3)(n-2)}{2} f(n)$	$f(4) + 4f(4) + \dots + \frac{(n-3)(n-2)(n-1)}{3!} f(n)$	$f(4) + 5f(4) + \dots + \frac{(n-3)(n-2)(n-1)n}{4!} f(n)$
$n-2$	$f(n-2)$	$f(n-2) + f(n-2) + \dots + f(n)$	$f(n-2) + 2f(n-2) + \dots + 3f(n)$	$f(n-2) + 3f(n-2) + \dots + 6f(n)$	$f(n-2) + 4f(n-2) + \dots + 10f(n)$	$f(n-2) + 5f(n-2) + \dots + 15f(n)$
$n-1$	$f(n-1)$	$f(n-1) + f(n)$	$f(n-1) + 2f(n)$	$f(n-1) + 3f(n)$	$f(n-1) + 4f(n)$	$f(n-1) + 5f(n)$
n	$f(n)$	$f(n)$	$f(n)$	$f(n)$	$f(n)$	$f(n)$
系列に關する和の一般項		$\sum_{n=1}^n f(n)$	$\sum_{n=1}^n n f(n)$	$\sum_{n=1}^n \frac{n^2+n}{2} f(n)$	$\sum_{n=1}^n \frac{n^3+3n^2+2n}{6} f(n)$	$\sum_{n=1}^n \frac{n^4+6n^3+11n^2+6n}{24} f(n)$
		$= V_0$	$= V_1'$	$= \frac{1}{2}(V_2 + V_1')$	$= \frac{1}{6}(V_3 + 3V_2' + 2V_1')$	$= \frac{1}{24}(V_4 + 6V_3' + 11V_2' + 6V_1')$

§10 表IVは既に直接法で振った数値例の計算方法を示している。
オ5次和が必要でないと言うのは、オ4次和での項の合計だけが
必要な値であるからである。

第 IV 表

頻度数	オI累加度数	オII累加度数	オIII累加度数	オIV累加度数
29	1.000	5.480	19.372	54.508
23	971	4.480	13.892	35.136
81	948	3.509	8.412	21.244
151	867	2.561	5.903	11.832
192	716	1.694	3.342	5.929
239	524	978	1.648	2.587
157	285	454	670	939
93	128	169	216	269
29	35	41	47	53
6	6	6	6	6
合計(検算のため)				
1,000	5,480	19,372	54,508	132,503

各列の合計から次を得る。

$$S_2 = d = 5.48, \quad S_3 = 19.372, \quad S_4 = 54.508, \quad S_5 = 132.503$$

最初の値 S_2 あるいは d は平均値が年齢 $52 + 5.48 \times 5 = 79.4$
であることを示している。年齢 52 は、度数が振られる前のグル
ープの中心であるから用いられている。そして既に注意したよう
に、累加法はこの立場に因してしなければならぬ計算を仮定し
ている。上に与えられた V_2, V_3, V_4 に対する公式を適用すると、

$$V_2 = 3.2336 \quad V_3 = -1.43099 \quad V_4 = 10.4164$$

であることがわかる。

§11 累加法を用いると、二三の方法で算術的計算が楽になる。

もし表IIIに含まれているすべての計算をする代りに、各欄に引

かれた際のすぐ上の和で止めるならば、我々は最後の合計として
次をうる。

$$\sum f(n), \quad \sum n f(n), \quad \sum \frac{n(n-1)}{2!} f(n), \quad \sum \frac{n(n-1)(n-2)}{3!} f(n) \\ \sum \frac{n(n-1)(n-2)(n-3)}{4!} f(n)$$

V_1, V_2 等の項の S_2, S_3, S_4 に対する5つの公式は、+-の符号
を交互に変えることだけ修正する必要がある。この節(§11)に
与えられているモーメントの形は、階乗モーメント(*factorial*
*moment**)と呼ばれている。計算の節約を少しでも助けること
は、表IIIに引かれた際のすぐ下の合計することによって避けれ
る。

これは、もし原点がかたよっていると仮定するとき得られたと
同じような結果を与える。もし我々が数値例に対するオ2の場合

* *semi-invariant* (半不変係数) は *Thiele* によって用いられた⁽¹⁾。
また他の著者はモーメントから求められている。オ2, オ3
semi-invariant はオ2, オ3の平均値のまわりのモーメントと同
じものである。オ4の *semi-invariant* はオ4モーメントよりオ
2のモーメントの平方の3倍だけ小さい。

$$\text{即ち} \quad (\mu_4 - 3\mu_2^2)$$

$$^{(1)} \exp\left[\sum_{j=2}^{\infty} \frac{\lambda_j}{j!} t^j\right] = 1 + \sum_{j=2}^{\infty} \frac{\mu_j}{j!} t^j$$

なるごとく $\lambda_2, \lambda_3, \dots$ は、分布の原点を移動しても不変であっ
て、これを *Thiele* の *semi-invariants* とする。

ここで μ_j は、平均値のまわりの j 次のモーメントである。

$$\mu_2 = \lambda_2$$

$$\mu_3 = \lambda_3$$

$$\mu_4 = \lambda_4 + 3\lambda_2^2$$

ついでながら

$$\exp\left[\sum_{j=1}^{\infty} \frac{K_j}{j!} t^j\right] = \sum_{j=0}^{\infty} \frac{\mu_j}{j!} t^j$$

なる $K_1, K_2, \dots$ を *Cumulant* とする (M. G. Kendall: *The advanced*
theory of statistics (I) 参照)

を取り、且つ表IVからの数値を用いるならば、我々は才2次和に対しては9780、才3次和に対しては8412、才4次和に対して11832を計算する。

また才5次和に対しては、才4欄に記入してある値の6つの和をとり8783を得るだけで充分である。

§12 これらは累加法を用いる直接の方法であるが、しかしモーメント計算の束法におけるように、我々は出発点あるいは任意の原点^{*}としてオ一項の代りに中央の項を用いることにより、計算を簡略化できる。

階乗積率形を使ってこの配列を用いることにしよう。選ばれた点の正の側の和の解釈については、もはや難かしいものはないけれども、負側の項に対するモーメントは計算が負の量の巾によってなされることを仮定するから、僅かの注意が必要である。

表IV(A)は計算しなければならない累加法の例を与えている。数値978は任意原点から正の側の値に対する $\sum n f(n)$ である。また498は符号を無視した負の側に対する同じような和、即ち $\sum n f(-n)$ である。

平均値は、差 $\sum n f(n) - \sum n f(-n)$ を全度数で割って得られる。即ち、 $(978 - 498)/1000 = 0.48$ 今“才3次和”の列の最後の数値を考えると、670は $\sum \frac{n(n-1)}{2!} f(n)$ を表わしていて、324は負の側に対応する数である。他の列も同じようにされている。今正の和にだけ関係している§11の表わし方に変換すれば、次のように書ける。

$$V_2' = 2S_3' + S_2'$$

$$V_3' = 6S_4' + 6S_3' + S_2'$$

$$V_4' = 24S_5' + 36S_4' + 14S_3' + S_2'$$

* 和をとることをみじかくすることの提案に対して Mr. G. J. Lidstone に感謝しなければならない。

ここに S' は異なる和の方法を示すために、 S の代りに用いられている。

表IV(A)において、我々は二つの部分に分布を分け、そして、今与えられた式を用いて各部分別々に、偶数次のモーメントの項に対しては加え、奇数次モーメントの項に対しては引かなければならない。故に、全分布に対して

$$V_2' = (2 \times 0.670 + 0.978) + (2 \times 0.324 + 0.498) = 3.464$$

$$V_3' = (6 \times 0.269 + 6 \times 0.670 + 0.978) - (6 \times 0.139 + 6 \times 0.324 + 0.498) = 3.336$$

$$V_4' = (24 \times 0.059 + 36 \times 0.269 + 14 \times 0.670 + 0.978) + (24 \times 0.029 + 36 \times 0.139 + 14 \times 0.324 + 0.498) = 32.194$$

で、また平均値周りに移して

$$V_2 = 3.2336 \quad V_3 = -1.43098 \quad V_4 = 30.41626$$

次にこれを記号によって表わそう。正の側の合計に対して P とかき、任意にえらばれた原点の負の側に対する合計を N とかくとき、我々は次をうる。

$$V_1' = P_2 - N_2$$

$$V_2' = (2P_3 + P_2) + (2N_3 + N_2) = 2(P_3 + N_3) + (P_2 + N_2)$$

IV (A) 表

度 数	才I累加度数	才II累加度数	才III累加度数	才IV累加度数	才V累加度数
29	29	29	29	29	29
23	52	81	110	139	...
81	133	214	324	...	...
151	284	498	...	...	...
192					
239	524	978	...	...	...
157	285	454	670	...	...
93	128	169	216	269	...
29	35	41	47	53	59
6	6	6	6	6	6
1,000					

また同様に

$$V_3' = 6(P_4 - N_4) + 6(P_3 - N_3) + (P_2 - N_2)$$

$$V_4' = 24(P_5 + N_5) + 36(P_4 + N_4) + 14(P_3 + N_3) + (P_2 + N_2)$$

§13 表IV(A)と表IVを比べれば、数値計算を簡約するには和の出発点として中央点を用いるだけで充分であることがわかる。何となれば、合計は数値的にも小さくなるし、また22頁の公式に入っている S_2 あるいは d は、非常に小さいからである。

項数が非常に多い場合には、いつでも累加法特に表IV(A)に与えられている型がモーメント計算の積の方法その一つの改良であるということが直に評価されよう。加算機を使えば和は100項以上も含むような系列に対してさえも少しの面倒なこともなく、機械的に得ることができる。

§14 オII章§12の平均値は個々の観測値 $o_1, o_2, \dots, o_N$ の項によって表わされている。同じように次のモーメントは、

$$\frac{1}{N} \sum_{i=1}^N o_i^2$$

であり、また t 番目の階乗積率は

$$\frac{1}{N} \sum_{i=1}^N o_i(o_i-1)(o_i-2)\dots(o_i-t+1) = \frac{1}{N} \sum_{i=1}^N o_i^{(t)}$$

である。

§15 さて曲線からのモーメントの計算を考えることが必要である。何故ならば、この計算をしとまらないと定数を見出す方程式を作ることは不可能であるからである。

$y = f(x, a, b, c, \dots)$ とおく。ここで $a, b, c, \dots$ は決定されるべき定数である。14頁～15頁をみると、計算の一つの方法は、

$$f(1, a, b, c, \dots) \times 1^n + f(2, a, b, c, \dots) \times 2^n + \dots$$

すなわち、 $\sum_{x=1}^{\infty} f(x, a, b, \dots) \times x^n$

を見出すことであることがわかった。そしてこれは、もし定数によってこの級数の和に対する代数的表示を見出すことが往々不可能であるという事実がなかったならば、方程式を作る場合に用いられるような一つの結果を与えるであろう。積分の形でこのような表示を見出すことが一般に可能である。また微分係数の定義として $y = x^n$ を定義したように $x = a$ から $x = b$ までの全分布の n 次モーメントは

$$\int_a^b y x^n dx \quad \text{或は} \quad \int_a^b f(x, a, b, c, \dots) x^n dx$$

である。全度数(すなわち調査された場合の全数)は $\int_a^b y dx$ である。平均値は既に注意したように

$$\frac{\int_a^b x y dx}{\int_a^b y dx} \quad \text{である。}$$

§16 方程式から曲線までのモーメントは、この方法で計算され曲線が縦線の系列で成り立っていると仮定して、統計量から計算したモーメントに等しいとした場合には一つの不正確さが導入されるのである。

次の二つの場合を考えよう。

- (1) 統計量は縦座標^{*}あるいは孤立した項の系列であり、それらに非常に近い曲線を引きたいと思うとき
- (2) 統計量が面積の系列であるが、面積が区間の中点に集中していると仮定してモーメントを計算する場合

^{*} 厳密に言えば度数分布ではなくて補正を必要とする値の系列である。分布は度数曲線を描くための面積として一般に取り扱われるのではない。何故ならば、場合の全数はグループ単位で割られており、また x 軸と曲線の間の全面積が用いられなければならないことをこの方法で教えている。

§17 上の(1)の場合では $y_0, y_1, y_2, \dots, y_{n-1}$ の各項は、統計量によって与えられる。そして $\int_{-\frac{1}{2}}^{\frac{1}{2}} y_x dx$ は近似的に y_0 に等しいから、 $\int_{-\frac{1}{2}}^{\frac{1}{2}} y_x dx$ が曲線の方程式で与えられると仮定することがもつとも簡単である*。そして我々は $\int_{-\frac{1}{2}}^{\frac{1}{2}} X y_x dx$ に $\sum_{x=0}^{n-1} X y_x$ が等しいとおくことによって生ずる誤差[†]を打ち消すための補正法を発見しなければならない。

困難を打破する最も実際的な方法は、求積公式(和の近似の公式)によって縦座標 $y_0, y_1, \dots, y_{n-1}$ に対応する真の面積を計算することである。多くの公式は良く知られているが、当面の目的では、面積がその上に値がつけられている区間の中と外の両方に存在する微軸の項で面積の近似値を与えることを表わすのに便利である。

記号的にこれらの公式は $y_{\frac{1}{2}}, y_{\frac{3}{2}}, y_{\frac{5}{2}}, y_{\frac{7}{2}}, \dots$ あるいは $y_0, y_1, y_{-1}, y_2, y_{-2}, \dots$ 等々の項によって $\int_{-\frac{1}{2}}^{\frac{1}{2}} y_x dx$ を表わす。

$$(I) \quad y_x = a + bx + cx^2 + dx^3 + ex^4$$

とおく。そのとき、

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} y_x dx = a + \frac{c}{12} + \frac{e}{80}$$

また $y_0 = a$

$$y_{-1} + y_1 = 2(a + c + e)$$

$$y_{-2} + y_2 = 2(a + 4c + 16e)$$

さて必要な積分が、

$$ay_0 + a(y_{-1} + y_1) + e(y_{-2} + y_2)$$

に等しいと仮定しよう。上に与えられているものを代入して、

* (1) の場合にはこれらの積分範囲を換うことが一般に可能であるが、もし0からnというような積分範囲が取られているときには、これとは異なる求積公式を換わねばならない。

† 保険数理に関する読者は、誤差は $(1 + \frac{1}{2} \alpha_n) = \alpha_n$ の仮定だけで導かれるのに類似していることに注意されたい。

a, c, e の係数をそれぞれ $1, \frac{1}{12}, \frac{1}{80}$ に等しいと置くと、次をうる。

$$a + 2a + 2e = 1$$

$$2a + 8e = \frac{1}{12}$$

$$2a + 32e = \frac{1}{80}$$

これらの方程式の解は、

$$a = \frac{5.128}{5.760}$$

$$c = \frac{308}{5.760}$$

$$e = \frac{-17}{5.760}$$

で与えられ、また

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} y_x dx = \frac{1}{5.760} \{5.128 y_0 + 308(y_{-1} + y_1) - 17(y_{-2} + y_2)\} \quad (I)$$

をうる。

$$(II) \quad y_x = a + bx + cx^2 + dx^3$$

ならば

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} y_x dx = \frac{1}{24} \{y_1 + 22y_0 + y_{-1}\} \quad (II)$$

$$(III) \quad y_x = a + bx + cx^2 + dx^3 + ex^4 \quad \text{ならば}$$

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} y_x dx = \frac{1}{1440} \{802(y_{\frac{1}{2}} + y_{-\frac{1}{2}}) - 93(y_{\frac{3}{2}} + y_{-\frac{3}{2}}) + 11(y_{\frac{5}{2}} + y_{-\frac{5}{2}})\} \quad (III)$$

$$(IV) \quad y_x = a + bx + cx^2 + dx^3 \quad \text{ならば}$$

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} y_x dx = \frac{1}{24} \{27y_0 + 17y_1 + 5y_2 + y_3\} \quad (IV)$$

§18 さて我々はモーメントの計算ができる。そこで $\int_{-\frac{1}{2}}^{\frac{1}{2}} x^k y_x dx$ は $y_0, y_1, \dots, y_{n-1}$ によって、

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} x^k y_x dx = \int_{-\frac{1}{2}}^{\frac{1}{2}} x^k y_x dx + \int_{\frac{1}{2}}^{\frac{3}{2}} x^k y_x dx + \dots + \int_{\frac{n-1}{2}}^{\frac{n-1}{2} + \frac{1}{2}} x^k y_x dx$$

もし公式(I)を応用すれば、それは最初の二項と最後の二項を除いた、この方程式の右辺のすべての積分を用いることができる。これらの値は(IV)によって与えられる。

求められた値を加え合せて分母を 5760 にして (IV) を書くと、
我々は、

$$\int_{-\frac{1}{2}}^{x-\frac{1}{2}} f_x dx = \frac{1}{5760} \{ 4463 f_0 + 4371 f_1 + 6669 f_2 + 5537 f_3 + 5760 (f_4 + f_5 + \dots + f_{n-2} + f_{n-1}) + 5537 f_n + 6669 f_{n+1} + 4371 f_{n+2} + 4463 f_{n+3} \} \dots (V)$$

をうる。これは、最初と最後の隙間に $\frac{4463}{5760}$ ($=1.1220486$) を乗
ずることができる。才二項と最後からの才二項は $\frac{4371}{5760}$ ($=0.7588542$)
倍され、才三項と最後から才三項は $\frac{6669}{5760}$ ($=1.1578125$) 倍され、
才四項と最後から三項を除いた項は $\frac{5537}{5760}$ ($=0.9612847$) 倍され
る。残るすべての項は、そのまま変わらずに残される。そして、
隙座標のこの修正された系列から普通の方法でモーメントを計算す
る。もし隙座標が 8 つ以下であるならば、それに対する他の公式を
展開しなければならない。

§19 次の表には、もとの系列と修正された系列が、最初の二列に
表示してある。そして他の列には直接法による範囲の中心のまわ
りの一次から 4 次のモーメントの計算が表示されている。

V 表

f_x	公式(V)によ り修正した値 f'_x	x	$f'_x \times x$	$f'_x \times x^2$	$f'_x \times x^3$	$f'_x \times x^4$
51.81	58.13	-4	232.52	930.08	3720.32	14,881.28
43.74	33.19	-3	99.57	288.71	868.13	2,598.39
35.42	41.01	-2	82.02	164.04	328.08	656.16
27.80	26.72	-1	26.72	26.72	26.72	26.72
20.42	20.42	0	-440.83	-----	-4941.25	-----
13.79	13.26	1	13.26	13.26	13.26	13.26
8.22	9.52	2	19.04	38.08	76.16	152.32
4.29	3.26	3	9.78	29.34	88.02	264.06
1.69	1.90	4	7.60	30.40	121.60	486.40
207.18	202.41		+69.68 -391.15	1,520.63	+299.04 -4,642.21	19,078.59

202.41 は、以後全度数として取扱う。

また単位度数に対するモーメント (μ_n) は -391.15, 1,520.63
等を 202.18 でなく 202.41 で割って得られる。

そしてこの 202.18 は全度数ではなくて、単にある等距離値の
未修正の和を与える。

§20 計算はかなり簡単にすることができる。何故ならば、もし経
験表の端での値が非常に小さくて、またそれらが最後に 0 になる
前に x 軸に接近している傾向を持っていると (例えばそこに高次
の接触 ⁽¹⁾ high contact) があれば、大抵の保険函数 (actuarial
functions) l_x, a_x, D_x 等は表の左側の端で高位の接触をもつ。
そのとき、隙座標が最初のグループの前と最後のグループの後に存
在すると考えることは合理的であるが、その値は、はつきりし
ない。故に、隙座標の全系列にわたる積分は前に用いた $-\frac{1}{2}$ と
 $n-\frac{1}{2}$ の限界を超えて合法的に拡張することができる。

何とならば、このようにして導入された附加的な面積は無限少
であるからである。

さてそのようにして面積を拡張されるならば、その影響は方程
式(V)で f_0 から f_{n-1} までののはつきりした隙座標はすべて 7 なる
係数を持ち、加重係数を持つ隙座標はすべて 0 になるであろう。

(1) 曲線と曲線、曲線と曲面あるいは曲面と曲面の間の接触は、一
般に 2 つの微分幾何学的曲面 S, S' 間の関係として定義する方が
見通しがよい。すなわち S, S' が 1 点 P を共有し S 上の点 X が S
上を動いて P に近づくとき、 X から S' までの距離が XP の距離よ
りも高位の無限少となると、 S は P において S' に接する (touch)
といい、無限少の位数を接触の係数 (order of contact) という。
 S が S' に接するかまたは S' が S に対するとき S, S' の間に接触
(contact) があるという。

実際の結果は次のごとくである。もし統計量の一方向の端にだけ高次の接触があれば、もう一方の端をだけ調整することが必要である。しかし、両方の端で高次の接触があるならばもはや調整の必要はない。

数学的に高次の接触は次のことを意味する。それは接触の点で最初の二、三の微分係数が0になることである。80頁と71頁の図には曲線の両端点の高次の接触が示されている。また72頁の図には期間のずっと先の方で高次の接触が示されている。

§21 面積の代りに中央縦線が用いられた§16の才(2)の場合を今ここで吟味しよう。

その区間の中央の周りに集中している面積によって、我々は次を仮定する。すなわち、面積 $\int_{-\frac{1}{2}}^{\frac{1}{2}} f_x dx$, $\int_{-\frac{1}{2}}^{\frac{1}{2}} f_x x dx$ 等に必要な距離は、 f_0 , f_1 等からの距離と同じものである。すなわち統計量からの t 次のモーメントは

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} f_x dx X^t + \int_{-\frac{1}{2}}^{\frac{1}{2}} f_x dx (X+1)^t + \dots + \int_{-\frac{1}{2}}^{\frac{1}{2}} f_x dx (X+n-1)^t$$

であり、そして我々は $\int_{-\frac{1}{2}}^{\frac{1}{2}} (X+x)^t f_x dx$ が必要である。ここで X は、モーメントが計算された点の周りの縦線から f_0 までの距離である。

各積分に公式(I) (31頁参照)を用いることにより項を集めて、一般的な係数として次をうる。

$$\frac{1}{5760} \{ \dots + [5178 f_0^t + 308 \{ (f_1 - 1)^t + (f_1 + 1)^t \} - 17 \{ (f_2 - 2)^t + (f_2 + 2)^t \}] f_0 + \dots \}$$

ここで f_0 は簡単のために $X+x$ の代りに書かれている。または

$$\frac{1}{5760} \{ 5760 f_0^t + 240 t(t-1) f_0^{t-2} + 3 t(t-1)(t-2)(t-3) f_0^{t-4} + \dots \}$$

もし $t=1$ ならば上式は f_1 になる。
 $t=2$ ならば $f_2 + \frac{1}{12}$

$t=3$ ならば上式は $f_3 + \frac{1}{2} f_1$

$t=4$ ならば $f_4 + \frac{1}{2} f_2 + \frac{1}{80}$

既に注意したように、もしそこに高次の接触があれば、

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} (X+x)^t f_x dx$$

の値は未修正縦線の使用によって見出すことができる。すなわち二次のモーメントはその一般項が $f_2 x^2$ である系列によって与えられ、3次のモーメントはその一般項が $f_3 x^3$ である系列によって与えられる。等々。故にもし μ が平均値のまわりに正確に修正されたモーメントを表わし、 V が未修正モーメントを表わすとすれば、 μ と V の間の関係は次式で与えられる。

$$\mu_2 + \frac{1}{12} = V_2 \quad \text{或は} \quad \mu_2 = V_2 - \frac{1}{12}$$

$$\mu_4 + \frac{1}{2} \mu_2 + \frac{1}{80} = V_4 \quad \text{或は} \quad \mu_4 = V_4 - \frac{1}{2} V_2 + \frac{7}{240}$$

平均値は修正の必要はない。何故ならもし $t=1$ ならば、一般項は正確な係数 f_1 を持つし、3次のモーメントは一次のモーメントに $\frac{1}{2}$ だけ修正されなければならない。そしてそれは0である。ここでモーメントは平均値まわりにとられている*

上の方法によって n 次のモーメントに対する修正を示すためには少くとも n 次の拋物線が必要である。

表Iの例IVに対して21頁で見出したモーメントにこれらの修正を適用すれば

$$\mu_2 = 3.1503, \quad \mu_3 = -1.430976, \quad \mu_4 = 28.82866$$

をうる。これらの修正はモーメントから得られた定数に評価できる差を生ぜしめることが分っている。項の数が少ないときに特にそうである。換言すれば修正は、群を考慮してそして本書で学ぶべき課題は、可成りの量を群化すれば計算が楽になり、正しい修正の知識に感謝し、上に述べた事情の下では誤差を導かないと言える。

* これらの修正は最初 Proc. Lond. math. Soc. XXIX, 353-80 に W. F. Sheppard によって与えられた。附録Iをみよ。

§22 高次の接触がある場合のモーメントの取り扱いに対して、前の2節の実際的な結論は数値的に検討することができ、各端点で高次の接触を持つ曲線に対する方程式は本節に書き降されているので、我々は等距離の点での縦線或は等間隔での面積を計算することができ、また数字からモーメントを計算することができる。曲線の方程式から全曲線に対応するモーメントと面積が計算でき、そして、それが対応する数字に一致することを発見するであろう。この方法での実験をするのに都合の良い曲線は“誤差の正規曲線”である。何故ならば、縦線と面積は *Tables for Statisticians* に正確に表示してある。しかしチエツクのためにこの種の応用を欲するものは、誰でも度数曲線について少しでも勉強するまでは持たれることを忠告しておこう。

曲線の両端に高次の接触がないときには、修正の評価するのが一層難しくなってくる。

修正を見つけるために作られている提案と、このことに関するより以上の論議は、附録Iに与えられている。しかし初学者はこれらの細部は避けてもよい。

学生は一、この分布に対するモーメントを計算すべきである。そして簡単な修正をやってみよ。そうすれば分布の標準偏差を見つけることもできる。なんとなれば、標準偏差 (S.D.) = $\sqrt{\mu_2}$ 。ここで μ_2 は上の規則によって修正されたものである。表Iの例IIとIVにおいては、明らかに高次の接触がある。そして例Iではモーメントの概数が用いられている。例IIとVではより不確実である。また例IIに対するモーメントの計算において(66頁参照)修正はされていない。

この注意は修正が不必要であるという理由を与えているのではなく、初学者は一般的思想を習得して満足することができ、また二、三の経験を持つまでは精密な部分のいくつかは、残しておいてもよいという理由を与えている。後に、附録Iの方法を検討するときに *Sheppard* の修正だけが分布が急であるときは、通

常概略のモーメントを改良はしないであろうということがわかるであろう。

もっと複雑な曲線の通合を取扱っていく前に、簡単な場合すなわち

$$y = a + bx + cx^2 + \dots$$

のときにモーメント法の応用を考えることが適當である。

今範囲は $2l$ とする。原点は範囲の中点であるとする。また m_0 は面積を表わし、 m_n は範囲の中点のまわりの全分布の n 次のモーメントを表わす。そのとき

$$\begin{aligned} m_{2d} &= \int_{-l}^{+l} (a + bx + cx^2 + \dots) x^{2d} dx \\ &= 2l \times l^{2d} \left(\frac{a}{2d+1} + \frac{cl^2}{2d+3} + \dots \right) \end{aligned}$$

また同じように

$$m_{2d+1} = 2l \times l^{2d+1} \left(\frac{bl}{2d+3} + \frac{dl^3}{2d+5} + \dots \right)$$

これらの方程式は偶数値のモーメントは通常 a, c, e 等を与え奇数次のモーメントは定数 b, d, f 等を与えることを示している。これは勿論範囲の中点のまわりのモーメントを用いた結果であるし、他の方法でするよりも少い労力で方程式が解ける。解は次のように書くことによっていくらか単純化もできる。

$$\frac{1}{2l} \frac{m_{2d}}{l^{2d}} = \frac{a}{2d+1} + \frac{cl^2}{2d+3} + \dots$$

故に

$$\left. \begin{aligned} \frac{1}{2l} m_0 &= a + \frac{cl^2}{3} + \frac{el^4}{5} + \dots \\ \frac{1}{2l} \frac{m_2}{l^2} &= \frac{a}{3} + \frac{cl^2}{5} + \frac{el^4}{7} + \dots \\ \frac{1}{2l} \frac{m_4}{l^4} &= \frac{a}{5} + \frac{cl^2}{7} + \frac{el^4}{9} + \dots \end{aligned} \right\}$$

又同じようにして

$$\left. \begin{aligned} \frac{1}{2l} \frac{m_1}{e} &= \frac{bl}{3} + \frac{dl^2}{5} + \frac{fl^3}{7} + \dots \\ \frac{1}{2l} \frac{m_2}{e^2} &= \frac{bl}{5} + \frac{dl^2}{7} + \frac{fl^3}{9} + \dots \\ \frac{1}{2l} \frac{m_3}{e^3} &= \frac{bl}{7} + \frac{dl^2}{9} + \frac{fl^3}{11} + \dots \end{aligned} \right\}$$

これら方程式の解は必要な定数を与える。例えば

(i) $y = a + bx$ ならば

$$a = \frac{1}{2l} m_0$$

$$b = \frac{3}{l} \cdot \frac{1}{2l} \cdot \frac{m_1}{e}$$

を得。又

(ii) $y = a + bx + cx^2$ ならば

$$a = \frac{3}{4} \left\{ \frac{3}{2l} m_0 - \frac{5}{2l} \frac{m_2}{e^2} \right\}$$

$$b = \frac{3}{l} \cdot \frac{1}{2l} \cdot \frac{m_1}{e}$$

$$c = \frac{15}{4l^2} \left\{ -\frac{1}{2l} m_0 + \frac{3}{2l} \frac{m_2}{e^2} \right\}$$

を得。又

(iii) $y = a + bx + cx^2 + dx^3$ ならば

$$a = \frac{3}{4} \left\{ \frac{3}{2l} m_0 - \frac{5}{2l} \frac{m_2}{e^2} \right\}$$

$$b = \frac{15}{4l} \left\{ \frac{5}{2l} \frac{m_1}{e} - \frac{7}{2l} \frac{m_3}{e^3} \right\}$$

$$c = \frac{15}{4l^2} \left\{ -\frac{1}{2l} m_0 + \frac{3}{2l} \frac{m_2}{e^2} \right\}$$

$$d = \frac{35}{4l^3} \left\{ -\frac{3}{2l} \frac{m_1}{e} - \frac{5}{2l} \frac{m_3}{e^3} \right\}$$

必要であれば、容易に拡張することができ、上の結果に一、二の数値例えの応用を試みよう。

§24 最初の例として §19 の表Ⅴにある統計量を補整してみよう。と言うのは、範囲の中点のまわりのモーメントが既に計算されているからである。曲線 $y = a + bx + cx^2$ をとれば、表Ⅴから

次の値をうる。

$$2l = 9, \quad e = 4.5$$

$$m_0 = 207.41$$

$$m_1 = -391.15$$

$$m_2 = 1520.63$$

故に、

$$a = \frac{3}{4} \left\{ \frac{622.23}{9} - \frac{5}{9} \times \frac{1520.63}{(4.5)^2} \right\} = 20.563$$

$$b = \frac{3}{4.5} \times \frac{1}{9} \times \frac{-391.15}{4.5} = -6.4387$$

$$c = \frac{15}{4(4.5)^2} \left\{ -\frac{207.41}{9} + \frac{3}{9} \times \frac{1520.63}{(4.5)^2} \right\} = 0.36815$$

§25 この補整に対応する縦線を得る最良の方法は、中心項からの最初の差 $b+C$ とオ二の差 $2C$ を計算することである。

それらの値はそれぞれ -6.0706 と 0.7363 である。従ってオ二次階差は定数であるから、計算は下のように連続的になされる。

	Δ	Δ^2
52.208		
43.192	-9.016	
34.913	-8.279	0.736
27.370	-7.543	
20.563	-6.807	
14.492	-6.071	
9.157	-5.335	
4.558	-4.599	
0.696	-3.862	

これらの補整された数値は、表Ⅴの最初の欄に与えられているものとかなりよく一致していることがわかるでしょう。

§26 この外の例として S. H. J. W. Allin (J. Inst. Actu. XXXIX, 350) の論文からとった次の統計量を考えよう。そこには、社員の年金に依って扶助料資金として未亡人に対する年金の額が与え

られている。

年令	年令の値	22頁の公式(V)によつて修正した値 a'	2倍した範囲の中点からの距離 d	$a' \times d$	$a' \times d^2$	$a' \times d^3$
27	21.20	23.29	-7	166.53	1165.71	8157.97
32	19.91	15.11	-5	75.55	377.75	1888.75
37	19.34	22.40	-3	67.20	201.60	604.80
42	18.58	17.86	-1	17.86	17.86	17.86
				-327.14		-10671.38
47	16.74	16.09	+1	16.09	16.09	16.09
52	15.69	18.17	+3	54.51	163.53	490.59
57	14.70	11.15	+5	55.75	278.75	1373.75
62	12.99	14.58	+7	102.06	714.42	5000.94
		137.15		+228.41	2935.71	+6901.37
				-98.73		-3770.01

上のモーメントの計算において、次のことが仮定されている。
すなわち補正される数値は、縦線のシステムを表わすと仮定した。
もしそれらが面積のシステムに表わされているならば、公式(V)による修正は不適当なものとなるであろう。

積分計算を避けて、方程式 $y = f(x)$ から縦線の和と縦線のモーメントを計算することである。

$f(x) = a + bx + cx^2 + \dots$ なる特別の場合には、これは実用向きであるが、それらのモーメントで積分できる多くの表わし方があるが、有限和に適合しない。従つて注意をより一般的な方法に制限した。

ここに偶数個の項があるときに、範囲の中点のまわりのモーメントの計算の困難性はそれらの項が 0.5, 1.5, 2.5, 等々 倍されるべきであるということである。そしてもし補正すべき系列が二、三の項しか含んでいない場合には、上に示した方法で距離 d を取扱うのが最上である。

またそのとき、第一、二、三次のモーメントをそれぞれ求める

ために全体を 2, 4, 8, で割ればよい。この方法によって、我々は次をうる。

$$e = 4$$

$$m_0 = 137.15$$

$$m_1 = -49.36$$

$$m_2 = 733.93$$

$$m_3 = -471.25$$

次に三つの曲線の各々に統計量を適合しよう。それらに対する公式はすでに与えられている。そして補正の結果を比較しよう。

$$(i) \quad y = 17.394 - 1.157x$$

$$(ii) \quad y = 17.633 - 1.157x - 0.0451x^2$$

$$(iii) \quad y = 17.633 - 1.190x - 0.0451x^2 + 0.0035x^3$$

次の表は、補正を示している。

年令	未補正値	(i)	(ii)	(iii)
27	21.20	21.44	21.13	21.13
32	19.91	20.29	20.24	20.28
37	19.34	19.13	19.27	19.31
42	18.58	17.97	18.20	18.22
47	16.74	16.82	17.04	17.02
52	15.69	15.66	15.80	15.76
57	14.70	14.50	14.46	14.43
62	12.99	13.34	13.03	13.05

公式(ii)と(iii)は、実用的に等しくまたこの両方は、(i)より元の数値にかなり似ている。

§27. 今迄に得られた結果を、次のように取りまとめておこう。

- ① モーメント法は特別な統計的例に適した公式の定数を見出だす一般的方法であつて、それは $\sum f(m) \times m^e$ の値を補正公式から得られた同じような表わし方に等しいと置くことから成立っている(ここで $\sum f(m) \times m^e$ は e 次のモーメントと呼ばれていて、そこに起るすべての m に対して加え上げられている)。

これらの後の表わし方は代数的である。連立方程式を算術的定数を見出すために解かねばならない。

- ② 統計量からのモーメントは、度数に g の適当な値を求むるか、あるいは累加法によって計算できる。
- ③ もしモーメントがある一つの重線のまわりに得られていれば本章の §6 の公式で他の点の重線のまわりのモーメントに変換できる。
- ④ 一般に補正公式からのモーメントが積分計算によって見出されねばならないが、一方統計量からのモーメントは、累加法によって見出されるから、後者は正しい定数を得るための方程式を形作ることができるが、その前に補正されなければならない。修正は統計量の擬線系か、もしくは面積系かによるが、前者の場合における修正は方程式 (V) によってなされ、後者の場合の修正はもし曲線の両端に高次の接触があれば、§21 の諸公式によって修正されねばならない。

第 IV 章 Pearson 系の度数曲線

§1 度数分布を表わす曲線系を決定することが実際の仕事において必要になってくるときは、次のことを心に留めておかなければならない。

- (1) 用いられるどんな表わし方も補正公式でなければならぬ。すなわちそれは資料のラフさを取り除かなければならない。
- (2) 非常に多くのモーメントを必要とするほど公式の中に多くの定数があるてはならない。何故ならこのことは正確度か減少されるということを意味するからである。モーメントが高次になればなるほど未修正の観測値から推論されるときには、より一層誤差を増やす。

その極端な場合は、最大数とその巾束が掛けられていることを思いだせば、このことは明確である。

- (3) 度数分布を取扱う方法は組織的でなければならぬ。

§2 さて次に、度数分布のもっとはっきりした性質を考えると、一般に 0 から出発して最大に上り、それからあるときは同じように、またあるときは異った割合で下降することがわかる。分布の最終点において往々高次の接触が存在する。これは数学的には、 x のある値すなわち最大点及び x 軸に接触する曲線の終端点で系列の各方程式で、 $\frac{dy}{dx} = 0$ なるように方程式 $y = f(x)$, $y = \phi(x)$ 等々の系列が決定されねばならないということの意味している。

上のことは $\frac{dy}{dx}$ が $y \times (x+a)/F(x)$ に等しいと置いてもいいことを暗示している。そのときもしも $y = 0$ なら $\frac{dy}{dx} = 0$ 故に、そこで曲線の一方側の端で接触がある。しかしもし $x = -a$ ならば $\frac{dy}{dx} = 0$ で我々は最大値を得る。 $F(x)$ が任意である限り $\frac{dy}{dx}$ に対して仮定した形は、やはり任意であり、 $y = 0$ のときは $\frac{dy}{dx}$ が 0 にならないときの場合も含む。

もし $F(x)$ が Maclaurin の定理によって x の界内に展開でき

るならば、次式をうる、

$$\frac{dy}{dx} = \frac{y(x+a)}{b_0 + b_1x + b_2x^2 + \dots} \quad (1)$$

さてこの方程式(1)にもどって、(1)がどうしたら y の直接の函数として y を表わすように $y = f(x)$ の形に置くことができるかを示そう、

この節の初めにはこのめかしたもので、もっと一般的なものがあることが得られるということがわかるであろう。我々は、色々異なる形をとる曲線を得るであろう。この問題は今日に至るまで、経験的観点から近づいてきたから、方程式(1)を如何にして確率論の初等的定理から分母にある x^2 の項まで得られるかを見ることは興味あることである、

§3 p がある事象が起る確率であり、 q はそれが起らない確率であるとすれば、そのとき n ケの試行の結果、1回、2回、3回……等と、その起る確率は $(p+q)^n$ の展開の各項によって与えられる。あるいは我々が N 回の実験をする場合ならば $N(p+q)^n$ の各項は $n+1$ の群において N 回の度数分布を与える。二項級数は生起のすべての確率を精密には表わすことはできない、

そして生起のすべての確率を表わす系列は超幾何級数である、これは p ケの黒球と q ケの白球の入った袋から、 r ケの球を抜くときに $r, r-1, r-2, \dots, 0$ の得られる機会は、次の系列の相継ぐ項によって与えられる、

$$\frac{p_n(p_n-1)\dots(p_n-r+1)}{n(n-1)\dots(n-r+1)} \times \left\{ 1 + \frac{rqn}{p_n-r+1} + \frac{r(r-1)}{2!} \right. \\ \left. + \frac{qn(qn-1)}{(p_n-r+1)(p_n-r+2)} + \dots \right\}$$

数値例は明らかに生起系列を作る助けとなろう、今 r ケの球の入っている袋がある、

そしてその中の q ケは黒で、 p ケは白である、そのとき、もし r ケの球が引出される確率は、

$$3 \text{ ケとも全部黒である確率} = \frac{4 \cdot 3 \cdot 2}{2 \cdot 6 \cdot 5}$$

$$2 \text{ ケが黒である確率} = \frac{4 \cdot 3 \cdot 3}{2 \cdot 6 \cdot 5} \times p_1$$

$$1 \text{ ケが黒である確率} = \frac{4 \cdot 3 \cdot 2}{2 \cdot 6 \cdot 5} \times p_2$$

$$1 \text{ つも黒くない確率} = \frac{3 \cdot 2 \cdot 1}{2 \cdot 6 \cdot 5}$$

これら4つの表わし方の和は1である、各項は $n=3$ $p_n=4$ $q_n=3$ $r=3$ とおいた系列に一致することがわかる、

他の系列を考えることもできる、しかしこれらは現在の目的には充分であろう、我々はそれらを方程式(1)の形におく方法を考えていこう、

今それらが直面しているような式の不便は、群の大きな数に対して数値計算がされるときにかなり明確になつてくる、そしてこれ以外は連続的でないが、実際の仕事の統計には往々現われるものである、

超幾何級数を考え、方程式(1)に対して必要な函数は $\frac{1}{y} \frac{dy}{dx}$ であり、またこの級数は不連続であるから有限階差を用いねばならないことを思い出せば、次を得る、

$$y_x = \frac{p_n(p_n-1)\dots(p_n-r+1)}{n(n-1)\dots(n-r+1)} \cdot \frac{r(r-1)\dots(r-x+2)}{(x-1)!} \\ \times \frac{q_n(q_n-1)\dots(q_n-x+2)}{(p_n-r+1)(p_n-r+2)\dots(p_n-r+x-1)}$$

$$\Delta y_x = y_{x+1} - y_x$$

$$= y_x \left\{ \frac{r-x+1}{x} - \frac{q_n-x+1}{p_n-r+x} - 1 \right\}$$

$$= y_x \left\{ \frac{(r+1)(q_n+1)-x(n+2)}{x(p_n-r+x)} \right\} \quad p+q=1$$

$$y_{x+\frac{1}{2}} = \frac{1}{2}(y_{x+1} + y_x)$$

$$= \frac{1}{2} y_x \left\{ \frac{(r+1)(q_n+1)-x\{2(r+1)+n(q-p)\}+2x^2}{x(p_n-r+x)} \right\}$$

$$\therefore \frac{\Delta y_x}{y_{x+\frac{1}{2}}} = \frac{2\{(r+1)(q_n+1)-x\{2(r+1)+n(q-p)\}+2x^2\}}{(r+1)(q_n+1)-x\{2(r+1)+n(q-p)\}+2x^2}$$

これは方程式(I)の形におかれたものである

$$\frac{1}{y} \frac{dy}{dx} = \frac{a+x}{b_0+b_1x+b_2x^2}$$

この形に着目すれば、保険数理の仕事に携っている読者は極く普通に死亡率を思い起すであろう。更に死亡率からその荷号を剥いた後で生命表その生存数(l_x)に移行していくことは、今与えられた度数曲線に対する公式から移行していくということと同じことである。

§4 方程式(I)にもどってみれば、次のような形にかかれることかわかる。

$$(b_0+b_1x+b_2x^2+\dots) \frac{dy}{dx} = y(x+a)$$

両辺に x^n を掛け、 x に関して積分すると次をうる。

$$\int x^n (b_0+b_1x+b_2x^2+\dots) \frac{dy}{dx} dx = \int y(x+a) x^n dx$$

$\frac{dy}{dx}$ を一つの項として扱うことによって、左辺は積分可能であり、右辺は二つの函数の和として積分可能である。そして、

$$x^n (b_0+b_1x+b_2x^2+\dots) y - \int \{n b_0 x^{n-1} + (n+1) b_1 x^n + (n+2) b_2 x^{n+1} + \dots\} y dx = \int y x^{n+1} dx + \int y a x^n dx$$

あるいはもし曲線の範囲の最も端で、式 $x^n (b_0+b_1x+b_2x^2+\dots) y$ が0になるならば、我々は

$$-n b_0 \mu_{n-1} - (n+1) b_1 \mu_n - (n+2) b_2 \mu_{n+1} - \dots = \mu_{n+1} + a \mu_n$$

をうる。ここで既に採用している記号を用いている。すなわち、

$$\mu_n = \int y x^n dx$$

もし $n=0, 1, 2, \dots$ とそれぞれおくと、次の方程式によって示されるようなモーメント(μ)によって $a, b_0, b_1, \dots$ 等を見出すために適当な $n+1$ 個の方程式を得る。そしてそれは次の形に方程式を書くことによって得られる。

$$a \mu_n + n b_0 \mu_{n-1} + (n+1) b_1 \mu_n + (n+2) b_2 \mu_{n+1} + \dots = -\mu_{n+1}$$

そこで $n=0, 1, 2, \dots$ 等とおくと、

$$\left. \begin{aligned} a \mu_0 + 0 b_0 + b_1 \mu_0 + 2 b_2 \mu_1 + \dots &= -\mu_1 \\ a \mu_1 + b_0 \mu_0 + 2 b_1 \mu_1 + 3 b_2 \mu_2 + \dots &= -\mu_2 \\ a \mu_2 + 2 b_0 \mu_1 + 3 b_1 \mu_2 + 4 b_2 \mu_3 + \dots &= -\mu_3 \\ a \mu_3 + 3 b_0 \mu_2 + 4 b_1 \mu_3 + 5 b_2 \mu_4 + \dots &= -\mu_4 \end{aligned} \right\} \dots \dots \dots (II)$$

等々をうる。

さて $\mu_1=0$ にして、また才III章で簡単に述べた方法で他のモーメントに変換しよう。なんとなれば $\mu_1=0$ にした結果は、分布の平均値に分布座標系の原点を変えらることであるからである。また μ_0 は1として取扱うことができる。そしてこれらは簡単に次の結果を導くことができる。

① b_0 だけに着目すれば

$$\frac{1}{y} \frac{dy}{dx} = -\frac{x}{\mu_2}$$

をうる。

② b_0 と b_1 に着目すれば、上の式(II)における初めの三つの方程式は、

$$a + b_1 = 0$$

$$b_0 = -\mu_2$$

$$\text{また } a \mu_2 + 3 b_1 \mu_2 = -\mu_3$$

$$\text{あるいは } b_1 = -\frac{\mu_3}{2 \mu_2}$$

$$\text{と } a = \frac{\mu_3}{2 \mu_2}$$

を与えらる。そして、微分方程式は、

$$\frac{1}{y} \frac{dy}{dx} = -\frac{x + \frac{\mu_3}{2 \mu_2}}{\mu_2 + \frac{\mu_3}{2 \mu_2} x}$$

となってくる。

③ b_0, b_1, b_2 に着目すれば、式(II)は

$$a + b_1 = 0$$

$$b_0 + 3 b_2 \mu_2 = -\mu_2$$

$$a \mu_2 + 3 b_1 \mu_2 + 4 b_2 \mu_0 = -\mu_3$$

$$2\mu_2 + 3b_2\mu_1 + 4b_1\mu_2 + 5b_2\mu_3 = -\mu_4$$

を与える。

これらの連立方程式の解はきわめて簡単であり、また次を導く。

$$\frac{1}{y} \frac{dy}{dx} = \frac{x + \frac{\mu_2(\mu_2 + 3\mu_1^2)}{10\mu_2\mu_4 - 18\mu_1^2 - 12\mu_3^2}}{-\frac{\mu_2(4\mu_2\mu_4 - 3\mu_1^2)}{10\mu_2\mu_4 - 18\mu_1^2 - 12\mu_3^2} + \frac{\mu_3(\mu_2 + 3\mu_1^2)}{10\mu_2\mu_4 - 18\mu_1^2 - 12\mu_3^2}x + \frac{2\mu_2\mu_4 - 3\mu_1^2 - 6\mu_3^2}{10\mu_2\mu_4 - 18\mu_1^2 - 12\mu_3^2}x^2}$$

この積換の形において $\beta_1 = \frac{\mu_1^2}{\mu_2}$ と $\beta_2 = \frac{\mu_3}{\mu_2}$ とおくと

$$\frac{1}{y} \frac{dy}{dx} = \frac{x + \frac{\sqrt{\mu_2}\sqrt{\beta_1}(\beta_2 + 3)}{2(5\beta_2 - 6\beta_1 - 9)}}{\frac{\mu_2(4\beta_2 - 3\beta_1) + \sqrt{\mu_2}\sqrt{\beta_1}(\beta_2 + 3)x + (2\beta_2 - 3\beta_1 - 6)x^2}{2(5\beta_2 - 6\beta_1 - 9)}} \quad (III)$$

§5 方程式(I)が最初に得られたこの理由は、 α が原点とモード(mode)の間の距離である。あるいは原点は $\mu'_1 = 0$ とおくことにより平均値に移されているので、 α がモードと平均値の間の距離であることを示している。モーメントによってこの距離を表わせば

$$\frac{\sigma\sqrt{\beta_1}(\beta_2 + 3)}{2(5\beta_2 - 6\beta_1 - 9)}$$

である。ここで σ は標準偏差 $\sqrt{\mu_2}$ である。

従って、歪度は平均値とモードの距離を標準偏差で割ったものである。すなわち

$$\text{歪度} = \frac{\sqrt{\beta_1}(\beta_2 + 3)}{2(5\beta_2 - 6\beta_1 - 9)}$$

§6 沢山の項数を供い $b_2, b_4, \dots$ 等々を考慮に入れると、微分方程式(I)の中の定数を得ることは可能であろうが、そのようなやり方には非常に強い実際的な障害がある。算術的計算の仕事の増す以外は、附加的な定数を導入することの利益は、既に注意したように高次のモーメントが不確定なものになってくるといふことから小さくなる。

Karl Pearson は次のことを示した*。任意標本に關して、一般の母集団では実際にもつていう二分の一あるいは二倍の値をもつ7次あるいは8次のモーメントを得ることは容易であろう。

これらの高次のモーメントに依存する定数は実際的には無駄である。それらは個々の任意標本を厳密に記述することを可能ならしめるが、とにかくこれらの高次のモーメントに依存する定数にもとづく限り、一般の大きな母集団に対して、この個々の標本から勝手に安全な論議を導くことはできない。

幾つかの保険数理統計量において、そこに100,000回位多くがある場合、それはこの級数の次の項まで進むことは価値があるかも知れないが、ここでさえも仕事の価値はなくなってくる。何故なら、同じ目的に關する統計量のどんな他の小さな群でさえも満足に結果と比較できないからである。實際目的には b_2 だけの方程式で満足であろうし、このようにして得られた形に対して我々の注意を限定しよう。

§7 (III)式で与えられる方程式(I)の實際的形にもどると次のことがわかる。統計量から得られたモーメントの値を、その方程式に代入して統計量を表わす公式を求めるといふことは可能であるということがわかるだろうが、これは元のデータに現われたと同じ形で補正を与えないであろう。何故ならば、根着においてはやるが、前者は $\frac{1}{y} \frac{dy}{dx}$ あるいは $\frac{d \log y}{dx}$ を与えるからである。それ故に、元のデータと比較しうる項を得るために求められた式を積分することが必要となるであろう。それは比較のために求めた式の形で方程式を換ふことは、實際の仕事ではより良いものであり、微分方程式を用いその結果を積分するよりもむしろ良い。

* Skew Correlation and non-linear regression. Draper's Company Res. Mem. 1905. P.9. または Chapter X をみよ。

後者の方法は実際の度数ではなく、比率を与えるにすぎない。

§8 次の段階は、方程式

$$\frac{d \log y}{dx} = \frac{x+a}{b_0 + b_1 x + b_2 x^2}$$

を $y = f(x)$ の形の式でおき換えることであり、そうするためには $(x+a)/(b_0 + b_1 x + b_2 x^2)$ を積分しなければならない。

積分の一般の形として方程式(III)を考えよう。そのとき次のことに気がつく。すなわち取られる積分の形は分母にある x の係数の特別の値によるということになる。この問題は実際には、

$$b_0 + b_1 x + b_2 x^2 = b_2 \left[x - \frac{-b_1 + \sqrt{b_1^2 - 4b_0 b_2}}{2b_2} \right] \left[x - \frac{-b_1 - \sqrt{b_1^2 - 4b_0 b_2}}{2b_2} \right]$$

に対する分母の取る形の考察に過ぎない。

特別の場合における固定した形に対する判別式は、明らかに方程式 $b_0 + b_1 x + b_2 x^2 = 0$ の根の性質に対すると同じものである。すなわち $b_1^2/4b_0 b_2$ である。これを公式(III)に代入することによつて

$$\frac{\beta_1(\beta_2 + 3)^2}{4(2\beta_2 - 3\beta_1 - 4)(4\beta_2 - 3\beta_1)} \quad \text{--- (IV)}$$

が得られる。

§9 式(IV)が負であれば根は実根で異符号であり、Karl Pearson によつて Type I と呼ばれた曲線の主要型の一つを得る。曲線のこの分類法については、どれでも Pearson-Type と言われている。

式(IV)が正で 1 より小さいならば、根は複素数である。そして我々は、才二の主要型 (Pearson の Type IV) をうる。また(IV)が正で 1 より大きいなら、根は実数で同符号である。そして才三の主要型 (Pearson Type VI) をうる。

これはすべての場合を含んでいるが、ある Type から他の Type に変える極限の場合において、我々は簡単な推移曲線の形に当着する。従つて基準 (criterion) が大きい (理論上無限大) と

には、一根は ∞ である (Type III)。また基準が 1 に等しいときには二根は等しい (Type V)。またそれが 0 のときには根は、異符号で同じ大きさである (Type II)。最後の場合において、 $b_1 = b_2 = 0$ ならば、一般に誤差の正規曲線と呼ばれるものに到達する。この名前はしばしば (確率曲線、ガウスの曲線等々) と他の名前があるために間違ひ易い。そこで再び $\frac{d \log y}{dx}$ に対する式は、 $\frac{a'}{b_0 + b_1 x}$ 形に変形できる。そして我々は度数曲線 (Type VII, IX, XI) に対して、直線あるいは二項式を得る。しかし、もしこの式が定数に変形できるならば、曲線は度数曲線システムでの特別の場合としての発見に満足した。普通の幾何級数である。何となれば、すでに我々は貨幣投げ等からの度数に関する確率論の立場でそれに熟知しているからである。

我々は先に進むにつれて、曲線のある状況の下では一つの曲線の限界を持つか、二つに分れた限界をもつ J 字型あるいは平 V 字型さえも取ることがわかるであろう。本書の終りにある図は、公式

$$\frac{d \log y}{dx} = \frac{x+a}{b_0 + b_1 x + b_2 x^2}$$

から展開した曲線の取る形の変化のアイデアを読者に与えるであろう。

実際には色々な種類の度数曲線に対する方程式が必要であろう。そして実際の場合に用いられるべき Type を知りたい。我々は何時でも概略データーの様子から Type を見当づけることはできない。数学的検定を必要とする。

§10 先づ度数曲線の方程式を扱おう。すなわち実際の積分法と共に 3 つの重要な Type から始めよう。

第 I 主要型 (Pearson Type I)

$b_0 + b_1 x + b_2 x^2 = 0$ の根が実数で異符号の時、分母の因子は次の形を取る。

$$b_2 \left[x - \frac{-b_1 + \sqrt{\text{正の量}}}{2b_2} \right] \left[x - \frac{-b_1 - \sqrt{\text{正の量}}}{2b_2} \right]$$

そして、積分される式は部分分数によって、次の形となる。

$$\frac{1}{b_2} \frac{x+a}{(x+A_1)(x-A_2)} = \frac{1}{b_2} \frac{A_1-a}{A_1+A_2} \frac{1}{x+A_1} + \frac{1}{b_2} \frac{A_2+a}{A_1+A_2} \frac{1}{x-A_2}$$

この積分は簡単で、次のようになる。

$$\log y = \frac{1}{b_2} \frac{A_1-a}{A_1+A_2} \log(x+A_1) + \frac{1}{b_2} \frac{A_2+a}{A_1+A_2} \log(x-A_2) + C (\text{定数})$$

$$\therefore y = y'(x+A_1)^{\frac{1}{b_2} \frac{A_1-a}{A_1+A_2}} (x-A_2)^{\frac{1}{b_2} \frac{A_2+a}{A_1+A_2}}$$

ここで y' は積分によって導入させられた定数である。

もし今原点をモードに極せば (例えば $x+a$ を x とおく)。

$$y = y_0 \left(1 + \frac{x}{a_1}\right)^{m_1} \left(1 - \frac{x}{a_2}\right)^{m_2}$$

をうる。ここで $\frac{m_1}{a_1} = -\frac{m_2}{a_2}$

第II主要型 (Pearson の IV 型)

もし方程式 $b_0 + b_1 x + b_2 x^2 = 0$ の根が複素数であれば、実変数に分母を減らすことは不可能である。そのとき基本的な微分方程式の右辺の表示を

$$\frac{x+C}{b_2(x^2+A^2)}$$

の形に書くことによって積分しなければならないことがわかる。

ここで

$$X = x + \frac{b_1}{2b_2}, \quad C = a - \frac{b_1}{2b_2}, \quad A^2 = \frac{b_0}{b_2} - \frac{b_1^2}{4b_2^2}$$

そこで

$$\begin{aligned} \log y &= \int \frac{X+C}{b_2(X^2+A^2)} dX \\ &= \int \frac{X}{b_2(X^2+A^2)} dX + \int \frac{C}{b_2(X^2+A^2)} dX \\ &= \frac{1}{2b_2} \log(X^2+A^2) + \frac{C}{Ab_2} \text{Arctan} \frac{X}{A} + C (\text{定数}) \end{aligned}$$

$$\therefore y = y' (x^2+A^2)^{\frac{1}{2b_2}} \exp\left\{\frac{C}{Ab_2} \text{Arctan} \frac{X}{A}\right\}$$

あるいは

$$y = \frac{y_0}{\left(1 + \frac{x^2}{a^2}\right)^m} e^{-\sqrt{A} \text{Arctan} \frac{x}{a}}$$

ここで a は方程式 (I) に含まれているものとは違った意味をもっている。この Type と Type I の間の関係は、微分方程式の右辺の分母を $b_2(x-iA)(x+iA)$ と因数分解してみればわかる。

それから Type I に同じ形を持つ y に対する式を得るが、やはり複素数の記号は含まれている。

第III主要型 (Pearson の Type II)

因数分解することは Type I の場合と同じであるが、方程式の根は同符号であるから分母の因子は $(x+A_1)(x+A_2)$ の形をとる。やり方は今までと同じであるが、原点は Pearson によってモードに置かれなくて、 x で表わされた式 $x+A_1$ あるいは $x+A_2$ の一方に置かれている。その結果次のようになる。

$$y = y_0 (x-a)^{m_1} x^{-m_2}$$

§11 さて、我々は二、三の推移型を述べよう。Pearson の Type II は、 $a_1 = a_2$ のときに Type I に同じものである。

Type III. これは判定基準が ∞ のときに得られる。それは $b_2 = 0$ のときに起る。

$$\begin{aligned} \log y &= \int \frac{x+a}{b_1+b_2x} dx = \int \left(\frac{1}{b_1} + \frac{a-\frac{b_0}{b_1}}{b_1x+b_0} \right) dx \\ &= \frac{x}{b_1} + \left(a - \frac{b_0}{b_1}\right) \frac{1}{b_1} \log(b_1x+b_0) + \text{定数} \\ y &= y' e^{\frac{x}{b_1}} (b_1x+b_0)^{\left(a-\frac{b_0}{b_1}\right) \frac{1}{b_1}} \end{aligned}$$

あるいは原点を換えることによって

$$y = y_0 e^{-sx} \left(1 + \frac{x}{a}\right)^{ra}$$

ここで a は方程式 (I) に含まれている a とは違った意味を持っている。

この Type は a_2 が無限大になるときの Type I の特別な場合であることがわかる。

Type V この場合にはいつでも根は等実根である。

$$\begin{aligned}\log y &= \int \frac{1}{b_2} \frac{x+a}{(x+b_1/2b_2)^2} dx \\ &= \int \frac{1}{b_2} \frac{(x+\frac{b_1}{2b_2})+(a-\frac{b_1}{2b_2})}{(x+b_1/2b_2)^2} dx \\ &= \int \frac{dx}{b_2(x+b_1/2b_2)} + \int \frac{a-b_1/2b_2}{b_2(x+b_1/2b_2)^2} dx \\ &= \frac{1}{b_2} \log(x+\frac{b_1}{2b_2}) - \frac{1}{b_2} \frac{a-b_1/2b_2}{(x+b_1/2b_2)} + \text{定数} \\ \therefore y &= y'(x+b_1/2b_2)^{\frac{1}{b_2}} e^{-\frac{a-b_1/2b_2}{b_2(x+b_1/2b_2)}} \\ &= y_0 x^{-p} e^{-\frac{1}{x}}\end{aligned}$$

誤差の正規曲線

$$b_1 = b_2 = 0 \quad \text{とおくと}$$

$$\begin{aligned}\log y &= \int \frac{x+a}{b_0} dx \\ &= \frac{x^2}{2b_0} + \frac{ax}{b_0} + \text{定数} = \frac{(x+a)^2}{2b_0} + \text{定数} \\ y &= y' e^{\frac{(x+a)^2}{2b_0}}\end{aligned}$$

あるいは原点を換えて、定数をまとめて

$$y = y_0 e^{-\frac{x^2}{c}}$$

同様にして余り重要でない推移曲線が得られる。これらは

$$(1+\frac{x}{a})^{-m}, \quad (1+\frac{x}{a})^m, \quad e^{-\frac{x}{a}}, \quad x^{-m}$$

等で、また Type I において m_1 か m_2 の何れか一方が負であるときは J 型曲線に到達し、また両方が負のときは h 型曲線に達す

る。J 型曲線は Type III から得ることもできる。

§12 折込みの表は Pearson の Type の数種の曲線系及びそれを一紙に用いる場合の原点を与えた表である。この表は、この種の問題に関する他の文献を読む際に通常採用されている Pearson 系の曲線の種類等かわかるから便利である。しかしながら私は原点が平均値にとられた時の各曲線の方程式をもつけ加えておいた。そこには原点に関して共通に言えるものがあるし、またすべての分布は平均値を持っており、モーメントは平均値を通る垂線のまわりについて計算されるという理由から、平均値を原点に持つ場合が便利なのである。

表の (5) 列は個々の場合にどの曲線を用いたらよいかを示すための判定基準を与えている。学生は同じ代数的形く例えば Type VIII, IX, XI は全部 ax^k にかかれる) に表わされる Type に関して時々出会うやや難かしいことをここで取扱おう。

我々は、どうして a から b まで ax^k を適合できないか、またモーメントの式から a, b, c, k を見つけられないのは何故かという問題を探ねよう。その答えは次の通りである。すなわち判定基準は実際に方程式の単純化を生ぜしめ、そして定数の値と曲線の範囲について多くを機械的に教えてくれる。

§13 次章の数値例の論議のときに二、三の技術的な点にもどるであろう。しかし今は方法の要点を繰返すことにする。そして統計量に多数曲線を適合する際にとらなければならない段階を調べてみる。

- (1) 統計量を系列に並べる。
- (2) 便利な垂線の周りについてモーメントを計算する。
- (3) 分布の重心を通るような垂線にモーメントを移す。
- (4) 曲線の両端に高次の接触があるときは、モーメントに対して Sheppard の補正を適用する(例えば2次または4次の

モーメントからそれぞれ $\frac{1}{12}$, $\frac{1}{2}\mu_2 - \frac{\mu_1^2}{12}$ を引く)。高次の接触がないときは付録のIをみよ。

(5) 判定基準を計算する。

(6) 表Ⅶを使ってどの曲線を使うべきかを決定する。

(5)と(6)のいずれか一方の代りに、どの曲線が用いられるべきかが *Tables for Statisticians* にある図からも見つけられる。その図は *Type* を β_1 と β_2 の項で示してある。

Pearsonが通常 用いた曲線型の 番号	Pearson 番号	注	意	定数の計 算はこの 下の頁を みよ
主要 型	I	$y = y_0(1+x)$	有限範囲、歪む、通常は鐘型、しかしU型、J型 あるいはねじれたJ型にもなる。	68~65
	IV	$y = y_0(1-x)$	無限範囲、歪む、鐘型	74~75
	VI	$y = y_0(1+x^2)$	一方向に無限範囲である、歪む、鐘型、しかし J型にもなる。	82
推 察 型	"正規曲線"	$y = y_0 e^{-x^2/2}$	無限範囲、対称、鐘型	88
	II	$y = y_0(1+x^2)^{-1}$	有限範囲、対称、通常鐘型、しかし $\beta_2 < 1.8$ のと きはJ型になる。	92~93
	VII	$y = y_0(1+x^2)^{-1/2}$	無限範囲、対称、鐘型	96
	III	$y = y_0(1+x^2)^{-1/2}$	一方向に無限範囲、通常は鐘型であるが、J型 にもなる。	97~98
	V	$y = y_0 x^2$	一方向に無限範囲、鐘型	101
	VIII	$y = y_0(1+x^2)^{-1/2}$	-aでの無限範囲から0での有限範囲までの範囲であ る(あるいは原点を平均値にもつ方程式では -a(m+1)/(m+2)から0/(m+2)までである。)	107
	IX	$y = y_0(1+x^2)^{-1/2}$	範囲はy=0なるx=-aからy=aなるx=0までである。 (あるいは原点を平均値にもつ方程式では -a(m+1)/(m+2)から0/(m+2)まで)	111
	X	$y = y_0 e^{-x^2/2}$	指数で0での有限範囲(平均値を原点にもつ方程式 では-0から0での無限範囲までの範囲、 J型)	113~114
	XI	$y = y_0(1+x^2)^{-1/2}$	J型、x=a(あるいは平均値を原点にもつ方程式で は-b/(m+2))で出発する。そこでの範囲は有限	116
	XII	$y = y_0(1+x^2)^{-1/2}$	ねじれたJ型、特別の場合にI型になる。	117

$$\beta_1 = \frac{\mu_2^2}{\mu_1^2} \quad \beta_2 = \frac{\mu_3^2}{\mu_2^2}$$

$$\beta_1 - 18\beta_2 - \beta_1(\beta_2 + 35)(8\beta_2 - 9\beta_1 - 12)$$

$$3)^2 + 4(4\beta_2 - 3\beta_1)(3\beta_1 - 2\beta_2 + 6)\}$$

IV 表 反 数 曲 線

Pearsonが通常 用いた曲線型の 番 号	Pearson が通常用いた型の曲線の方程式		平均値を原点にもつ方程式	判 定 基 準	注 意	定数の計 算はこの 下の頁を みよ
	方 程 式	原 点				
主 要 型	I $y = y_0 \left(1 + \frac{x}{a_1}\right)^{\frac{1}{\beta_1}} \left(1 - \frac{x}{a_2}\right)^{\frac{1}{\beta_2}}$	モ - ト	$y = y_0 \left(1 + \frac{x}{A_1}\right)^{\frac{m_1}{\beta_1}} \left(1 - \frac{x}{A_2}\right)^{\frac{m_2}{\beta_2}}$ ここに $\frac{(m_1+1)}{A_1} = \frac{(m_2+1)}{A_2}$	$K < 0$	有限範囲。歪む。通常は鐘型。しかしU型、J型 あるいはねじれたJ型にもなる。	64-65
	IV $y = y_0 \left(1 + \frac{x^2}{a^2}\right)^{-m} e^{-\sqrt{\lambda} \arctan \frac{x}{a}}$	平均値の後 $\sqrt{a^2/2m-2}$	$y = y_0 \left[1 + \left(\frac{x}{a} - \frac{V}{T}\right)^2\right]^{-m} e^{-\sqrt{\lambda} \arctan \left(\frac{x}{a} - \frac{V}{T}\right)}$ ここに $r = 2m - 2$	$1 > K > 0$	無限範囲。歪む。鐘型	74-75
	VI $y = y_0 (x-a)^{\frac{\beta_1}{2}} x^{-\frac{\beta_2}{2}}$	曲線の出発点前 a	$y = y_0 \left(1 + \frac{x}{A_1}\right)^{-\frac{\beta_1}{2}} \left(1 + \frac{x}{A_2}\right)^{\frac{\beta_2}{2}}$ ここに $\frac{(\beta_1-1)}{A_1} = \frac{(\beta_2+1)}{A_2}$	$K > 1$	一方向に無限範囲である。歪む。鐘型。しかし J型にもなる。	82
推 移 型	"正規曲線" $y = y_0 e^{-\frac{x^2}{2\sigma^2}}$	モ - ト (= 平均値)	$y = y_0 e^{-\frac{x^2}{2\sigma^2}}$	$K=0, \beta_1=0, \beta_2=3$	無限範囲。対称。鐘型	88
	II $y = y_0 \left(1 - \frac{x^2}{a^2}\right)^m$	モ - ト (= 平均値)	$y = y_0 \left(1 - \frac{x^2}{a^2}\right)^m$	$K=0, \beta_1=0, \beta_2 < 3$	有限範囲。対称。通常鐘型。しかし $\beta_2 < 1.87$ と なればU型になる。	92-93
	III $y = y_0 \left(1 + \frac{x^2}{a^2}\right)^{-m}$	モ - ト (= 平均値)	$y = y_0 \left(1 + \frac{x^2}{a^2}\right)^{-m}$	$K=0, \beta_1=0, \beta_2 > 3$	無限範囲。対称。鐘型	96
	IV $y = y_0 \left(1 + \frac{x}{a}\right)^p e^{-\sqrt{\lambda} x}$	モ - ト	$y = y_0 \left(1 + \frac{x}{A}\right)^p e^{-\sqrt{\lambda} x}$ ここに $A = \frac{(p+1)}{\lambda}, p = \gamma a$	$2\beta_2 = 6 + 3\beta_1$	一方向に無限範囲。通常は鐘型であるが、J型 にもなる。	97-98
	V $y = y_0 x^p e^{-\sqrt{\lambda} x}$	曲線の出発点	$y = y_0 \left(1 + \frac{x}{A}\right)^p e^{(p+1)\sqrt{\lambda} x}$	$K = 1$	一方向に無限範囲。鐘型	101
	VI $y = y_0 \left(1 + \frac{x}{a}\right)^{-m}$	曲線の終点	$y = y_0 \left(1 + \frac{x}{A}\right)^{-m}$	$K < 0, \lambda = 0$ $5\beta_2 - 6\beta_1 - 9 < 0$	-aでの無限範囲から0での有限範囲までの範囲であ る(あるいは原点を平均値にもつ方程式では -a(m+1)/(m+2)から0/(m+2)までである。)	107
	VII $y = y_0 \left(1 + \frac{x}{a}\right)^m$	曲線の終点	$y = y_0 \left(1 + \frac{x}{A}\right)^m$	$K < 0, \lambda = 0, 5\beta_2 - 6\beta_1 -$ $9 > 0, 2\beta_2 - 3\beta_1 - 6 < 0$	範囲はy=0なるx=-aからy=無限なるx=0までである。 (あるいは原点を平均値にもつ方程式では -a(m+1)/(m+2)から0/(m+2)まで)	111
	VIII $y = y_0 e^{-\frac{x}{a}}$	曲線の出発点	$y = y_0 e^{-\frac{x}{a}}$	$\beta_1 = 4, \beta_2 = 9$	指数で0での有限範囲(平均値を原点にもつ方程式 では-0)から0での微小な無限範囲までの範囲。 J型。	113-114
	IX $y = y_0 x^{-m}$	曲線の出発点前 a	$y = y_0 \left(1 + \frac{x}{A}\right)^{-m}$	$K > 1, \lambda = 0,$ $2\beta_2 - 3\beta_1 - 6 > 0$	J型。x=a(あるいは平均値を原点にもつ方程式で は-0/(m+2))で出発する。その後の無限範囲は有限	116
	X $y = y_0 \left\{ \frac{0(\sqrt{3+\beta_1} + \sqrt{\beta_1}) + x}{0(\sqrt{3+\beta_1} - \sqrt{\beta_1}) - x} \right\}^{\frac{(\beta_2+1)}{2}}$	平 均 値	$y = y_0 \left(1 + \frac{x}{A_1}\right)^{m_1} \left(1 - \frac{x}{A_2}\right)^{m_2}$ これはもう一つの形である。 ここに m_1 と m_2 は数値的に等しく共に < 1 で異符号である。	$5\beta_2 - 6\beta_1 - 9 = 0$	ねじれたJ型。特別の場合にI型になる。	117

$$K = \frac{\beta_1(\beta_2 + 3)^2}{4(4\beta_2 - 3\beta_1)(2\beta_2 - 3\beta_1 - 6)} \quad \beta_1 = \frac{M_1^2}{M_2^2} \quad A_2 = M_2/M_1^2$$

$$\lambda = \frac{(4\beta_2 - 3\beta_1)(10\beta_2 - 12\beta_1 - 18)^2 - \beta_1(\beta_2 + 3)^2(8\beta_2 - 9\beta_1 - 12)}{(3\beta_1 - 2\beta_2 + 6)\{\beta_1(\beta_2 + 3)^2 + 4(4\beta_2 - 3\beta_1)(2\beta_2 - 3\beta_1 - 6)\}}$$

第 V 章 計 算

§ 1 次に考えなければならぬ莫はモーメンが計算されていて、用いられるべき型が決定しているときにある特定の分布に対する定数の計算である。数値計算に対して必要な公式を各型について与えておこう。果敢的に並べられた数値の計算を含めた数値例は公式の証明と共に后節に与えるであらう。

§ 2 定数が発見されたときに曲線の計算に関係するいくつかの一般的な特長をここで考えるのが便利である。定数が知られているから定数曲線の式に又の要測定値を代入する事によって任意の x に対する曲線を計算する事が出来る。またもし面積が要求されれば曲線から面積を導く方法を発見しなければならぬ。最も簡単なのは起らく中央曲線を計算しそれから y の面積公式 (I) 或は (II) によって面積を求める事である。各グループの始めの莫で曲線を計算しそれから公式 (II) を用いるのがしばしば便利である。これらの公式は差の形でもつともよく適用出来る。即ち公式 (II) から次式をうる。

$$\int_{-\frac{\Delta}{2}}^{\frac{\Delta}{2}} y_x dx = y_0 - \frac{\Delta^2}{24} \{ \Delta y_{-1} - \Delta y_1 \}$$

また (I) から

$$\int_{-\frac{\Delta}{2}}^{\frac{\Delta}{2}} y_x dx = y_0 - \frac{\Delta^2}{24} \{ \Delta y_{-1} - \Delta y_1 \} + \frac{\Delta^4}{5760} \{ \Delta y_{-2} - \Delta y_2 \}$$

また公式 (III) から

$$\int_{-\frac{\Delta}{2}}^{\frac{\Delta}{2}} y_x dx = \frac{\Delta}{2} \{ y_{\frac{\Delta}{2}} + y_{-\frac{\Delta}{2}} \} + \frac{\Delta^3}{1440} \{ \Delta y_{1, \frac{\Delta}{2}} - \Delta y_{\frac{\Delta}{2}} \} - \frac{\Delta^5}{14400} \{ \Delta y_{2, \frac{\Delta}{2}} - \Delta y_{1, \frac{\Delta}{2}} \}$$

等々をうる。

公式 (II) は、一般的には充分正確である。しかしその他は普通の場合には Δ の数字に正しい結果を与える事が分るであらう。—— 例外的な場合は后述の数値例を参照されたい。

§3 グラフ的表示は果敢的配列をみるためにしばしば助けとなる。そしてこの事は幾つかの例でも既に済んだ事である。モードを垂直高さを書き入れることは最良の方法である。曲線の端および、曲線の高さが計算されたものである事に注意せよ。これらの高さは曲線上の矢と与える。そしてかなり容易にそれらを通つて曲線を描く事が出来る。実数の計算の場合と同様に曲線を描く際には垂直の符号に注意しなければならない。

何故ならモードの間違った側には歪みを持つ曲線を描き易いからである。またもし分布が対称的に近いなら我々が期待する様にこの誤りを指摘する事はそんなに容易ではない。モードの真での曲線の誤差はJ型曲線の場合や数少いが普通の推移型 (Transition Type) の幾つかを除いて又端に平行である。

§4 明白さを得るためにはむしろ大きな尺度によつて測く方がよい。ここで与えた曲線はそれらのありのまゝの大きさとよりも少し大きく測がいたものである。勿論細少は複写の過程を終て作られる。

区画は高さに比例して充分大きくすべきである。そうすれば曲線は余り急に上る事はない。他方果敢的に配列した曲線とそうでない曲線の間の小さな水平差は曲線が急激に上昇或は下降するときには大きな垂直差を随しがちである。然し重要なのは后者の差である。

§5 読者は以下に考えられている凡ての場合が完全な分布を仮定している事に注目すべきである。またそれは極端な努力を要する相対的近似値によつて分布の一部から曲線を見出す事が一般に可能であるにすぎない。

他の実は付録にもう一度なされる事であるが統計量のグループ分けに関してである。多くのグループを得る事は不可能な場合がしばしばあるが、モーメントを見出す際の正確度に対しては場合

の全数が小さくなければ群の数が大きければ大きい程よい。この関係においてわずかの手加減が必要であるが実際の統計においては或るときは200000個位多くにもとずく事がしばしばありそれにグループ分けを用いるのが正確さのためにもよい。我々の例では計算の師役と紙の余地がないために唯グループ分けした式でそのグループ分けの方法は変らないのである。我々がそれに適当な修正を知っている程の高次の接触があるとすればグループ分けは殆どあるいは全く誤差を生じない。

10年毎にグループ化しそれを単位として用いるときにオス次モーメントに対する右の修正は5年毎のグループにした場合或は年毎のグループを用いるとに比例的に多くの効果を持つ。

グループ化が正確度を壊すと言う様な印象を与える恐れはこの様な場合には何等適当な根拠のない事である。この実に関連してわずかの数値的証明は付録Iにみられる。

§6 ここで取扱う方が適當と思われるもう一つは判定規準Kに関する事である。これは $-\infty$ から $+\infty$ までの任意の値を取り得る。そして次の図からKのあらゆる可能な値はどんな型にもあてはまつており互に重なり合っていない事が分る。



J 値 $K = 0$ になる前から Type I は対称的に近づき、その値が大きくなると、無根能面の非対称曲線となる等々。

各臨界 (Critical) 点で一つ或はそれ以上の推移曲線がある。若し読者が誤つた主要型を用いたとすれば、実数を求める方程式に現われる平方根の一つは虚数となるからすぐにその間違を見出すであろう。然し推移型は判定基準値が理論値に近いときに用

いる事が出来る。実際には、それらは極限をとった場合に正確な値を与える近似値として見る事が出来る。

我々が推移型を使用する際にどの様な限界内でそれを用いる事が妥当であるかと言う正確な限界を与える事は不可能である。その妥当性は右に示す様に理論的な証明に取扱われた函数の標準誤差の大きさによる。然し実際には大部分は経験によつて導かれる。若し場合の数が少ない場合があれば判定標準での大きな偏差は場合の数が多い場合よりもより多く起るであろう。個々の場合にそれらの限界について考えるべきである。しかし諸君が述べている様な場合は推移型の使用をとける事が正確度の点では安全である。読者は凡ての場合に一つの安全な指針、即ち「論より証」である。と言う事を知っているだろう。

推移曲線の応用性に何の望みももたない訳は、三の場合に推移曲線を使つてみてそれを正しい主要型によつて得られたものと其の結果と比較すべきである。そうすれば両方の曲線について多くを知るであろう。

§7 色々の型に対して与えられる公式で平方根の符号選択は、 M_0 の符号による。若し度数が平均値の右よりも前の方により密集していれば、モードは平均値の左側にあり、また M_0 は正である。各型でのある度数の符号はモードと平均値がそれらの正しく相対的な位置にあるために M_0 の符号による。しかし与えられた平方根は正の根を意味している事から式の符号が定められる事は注意を要しない事である。読者は、これらの法則は一つは M_0 に対して正で他は負を与える二つの例をしらべてみるとたやすく理解出来る事が分るであろう。したがつても我々が 1, 3, 7, 13, ... 等反対の順に書かれているものを Type I の例での度数に想定しても、すべての数値計算は同じであろう。しかし $M_0 = 2.7745728$, $M_1 = 0.407833$, $M_2 = 1.352728$, $M_3 = 1.774535$ であろう。また補正値も同じであろうが 70 頁の表の各欄の数値

と反対の順に一致する筈である。

§8 算術的計算は大抵の読者にとつて幾つかの奥に馴染めなく大変なものである。そこには最後の段階で注意し大乗づけ倍取を用いる以外はそれを成功する近路はない。最初の段階で失敗を発見すると再計算しなければならぬ為に算術的に長い仕事の最後ではいらいらさせられる。しかしこれらの失敗は経験と共に重要でなくなつたりまた少なくなつて来る。何故ならば、実際には我々は大きな誤りは直ぐさま間違つた値になつて来た事を感ずくからである。私は大抵 π の対数値を用いて必ずしも最後の数字まででなくとも各段階にわたつてチェックする。この様な方法が本書の算術的計算に用いられている。検算は補正に影響しない誤り或は定数や係数の最後の数値に影響を受けるに過ぎない様な誤りを見えられないかも知れない。したがつて 68 頁にある $\log y_0$ の最後の3つの数値が誤っているならば（その事については誤り理由を持たない）この失敗は残念である。しかし 70 頁にある補正は誤らないであろう。その上多くの操作が行なわれた時に最後の数字の普通の不確実性のために、他の計算者の正確な数値結果の再観を視る事は困難であるかもしれない。長い算術的計算での最後の二つの数字は不確実であるかも知れないし、その上分離を与えるかも知れない。これは π の対数値は π の対数値も同様である事を意味してはいない。何故ならば、若し π の数字が正確に π の数字を与えるならば我々は一様に言われている様に π の数値計算は π 桁までの数値が信頼出来るに過ぎない事を仮定するからである。

積率に対する諸公式

これらの公式は曲線のあらゆる型に適用される。

$$\left. \begin{aligned} V_1' &= d \\ V_2 &= V_1' - d^2 \\ V_3 &= V_2' - 3dV_2 - d^3 \\ V_4 &= V_3' - 4dV_3 - 6d^2V_2 - d^4 \end{aligned} \right\} \text{或は} \left\{ \begin{aligned} V_1' &= d \\ V_2 &= V_1' - d^2 \\ V_3 &= V_2' - 3dV_2 + 2d^3 \\ V_4 &= V_3' - 4dV_3 + 6d^2V_2 - 3d^4 \end{aligned} \right.$$

或は

$$S_2 = d$$

$$V_2 = 2S_2 - d(1+d)$$

$$V_3 = 6S_2 - 3V_2(1+d) - d(1+d)(2+d)$$

$$V_4 = 24S_2 - 2V_3\{2(1+d)+1\} - V_2\{4(1+d)(2+d)-1\} - d(1+d)(2+d)(3+d)$$

$$\mu_2 = V_2 - \frac{1}{12}$$

$$\mu_3 = V_3$$

$$\mu_4 = V_4 - \frac{1}{2}V_2 + \frac{7}{240}$$

曲線の両端に高次の接触があるとき
の Sheppard の補正

$$\sigma \text{ (標準偏差)} = \sqrt{\mu_2}$$

$$B_1 = \frac{\mu_3}{\mu_2^2}$$

$$B_2 = \frac{\mu_4}{\mu_2^3}$$

$$K = \frac{B_1(B_2+3)^2}{4(B_2-3B_1)(2B_1-3B_2-4)}$$

オ 1 主要型 (I 型)

$$y = y_0 \left(1 + \frac{x}{a_1}\right)^{m_1} \left(1 - \frac{x}{a_2}\right)^{m_2}$$

$$\text{但し } m_1/a_1 = m_2/a_2$$

原典はモード

順次計算されねばならない値は次の通りである。

$$r = 4(B_2 - B_1 - 1) / (4 + 3B_1 - 2B_2)$$

$$a_1 + a_2 = \frac{1}{2} \sqrt{\mu_2} \sqrt{B_1(r+2)^2 + 4B_2(r+1)}$$

その値は次の式で与えられる。

$$\frac{1}{2} \{ r - 2 \pm r(r+2) \sqrt{\frac{B_1}{B_1(r+2)^2 + 4B_2(r+1)}} \}$$

また、 μ_2 が正であるときは m_2 は正の根である。

$$y_0 = \frac{N}{a_1 + a_2} \cdot \frac{m_1! m_2!}{(m_1 + m_2)!} \cdot \frac{\Gamma(m_1 + m_2 + 2)}{\Gamma(m_1 + 1) \Gamma(m_2 + 1)}$$

$$\text{モード} = \text{平均値} - \frac{1}{2} \frac{\mu_3}{\mu_2} \cdot \frac{r+2}{r-2}$$

平均値を原典として曲線を表すならば (57 頁の折込みの表 VI の周辺をみよ)

$$A_1 + A_2 = a_1 + a_2$$

$$(m_1 + 1)/A_1 = (m_2 + 1)/A_2$$

$$y_0 = \frac{N}{A_1 + A_2} \cdot \frac{(m_1 + 1)! (m_2 + 1)!}{(m_1 + m_2 + 2)!} \cdot \frac{\Gamma(m_1 + m_2 + 2)}{\Gamma(m_1 + 1) \Gamma(m_2 + 1)}$$

Γ 函数の表は 296~299 頁或は *Tables for Statisticians* をみよ。

注意

曲線の普通の形は次の例にみられる様なものである。 m_1 と m_2 が近似的に等しいなら曲線は対称的に近づき m_1 と m_2 が小さくない限り曲線はその両端で次第に細くなる。また m_1 と m_2 がともに小さいときは曲線は両端で急に上昇する。もし m_1 が負であるなら曲線は J 型である。曲線は無限横軸から出発し急速に下降するを以て個定された典で止まる (数値例は 135 頁をみよ)

もし m_1 と m_2 の両方が負であるならば曲線は U 型である。このとき曲線の出発典と終典はともに無限横軸である。そして普通原典としてモードの代りに逆モード (anti-mode) をもっている (数値例は 138 頁をみよ) J 型または U 型曲線において横軸は無限であるけれども面積は有限である。もく 1 のときに必要である $\Gamma(t)$ に対する Γ -函数の値を求めるときには注意が必要である。そして、 $\log \Gamma(1+x)$ 即ち $\log x + \log \Gamma(x)$ が表にあたえられている。J 型曲線の場合には、平均値を原典にもつ形を用いるか或は曲線の出発典を原典に持つ形 $y' x^{m_1} (a_1 + a_2 - x)^{m_2}$ で曲線を表すのが最良である。そして、

$$y' = \frac{N}{(a_1 + a_2)^{m_1 + m_2 + 1}} \cdot \frac{\Gamma(m_1 + 1) \Gamma(m_2 + 1)}{\Gamma(m_1 + m_2 + 2)}$$

J型曲線の将来ある変型は m_1 と m_2 がともに数値的に1より小さく且つどちらか一方が0である場合に起る。そのときの曲線の型は本書のおわりにある曲線図を $16(11)$ の形と同じ様及ものである。例えばその $16(11)$ は、ねじれたJ型である(117~119頁の例と注意をみよ)

例

この型の例として表Iの数値(例II)を用いよう。モーメントは次の表に示されている様な累加法(Ⅲを参照)によつて定められる。前者は任意の原点として年齢45歳とつてもつと直接的方法でモーメントを再計算する事で結果の検算が出来るこれは私自身が計算にいつも用いている方法である。私は系列が非常に長いとき累加法しか用いない。私は唯例によつてここにそれを示そう。

年 令	中 心 地 の 年 令	表 I の 例 II に お け る 人 数	第 I 累 加 数	第 II 累 加 数	第 III 累 加 数	第 IV 累 加 数
15 ~ 19	17	34	1000	5175	19809	64389
20 ~ 24	22	146	966	4175	14634	44580
25 ~ 29	27	156	821	3209	10457	29946
30 ~ 34	32	146	665	2388	7250	19487
35 ~ 39	37	123	520	1723	4962	12237
40 ~ 44	42	103	377	1203	3139	7375
45 ~ 49	47	86	294	806	1936	4236
50 ~ 54	52	71	208	512	1130	2300
55 ~ 59	57	55	137	304	618	1170
60 ~ 64	62	37	82	167	314	552
65 ~ 69	67	21	45	85	147	238
70 ~ 74	72	13	24	40	62	91
75 ~ 79	77	7	11	16	22	29
80 ~ 84	82	3	4	5	6	7
85 ~ 89	87	1	1	1	1	1
計		1000	5175	19809	64389	186638

$$S_2 = 5175/1000 = 5.175$$

$$S_3 = 19809/1000 = 19.809$$

$$S_4 = 64389/1000 = 64.389$$

$$S_5 = 186638/1000 = 186.638$$

次の段階は、64頁の諸公式を使つて重心軌線、の周りのモーメントを求める事である。この場合にモーメントが修正^{*}されていないので μ と M は同じものである。

$$\mu_1 = 766237^{(1)}$$

$$\mu_2 = 15.1067$$

$$\mu_3 = 172.326$$

$$B_1 = 25472955 \quad B_2 = 2935110$$

B_1 と B_2 の値から判定基準 K が計算出来る。その値-0.2845はI型を用いなければならぬ事を示している(可表をみよ)

$$Y = 5186811 \quad \log Y = 0.7147004$$

$$Y+1 = 5186811 \quad \log(Y+1) = 0.7914669$$

$$Y+2 = 7186811 \quad \log(Y+2) = 0.8565363$$

$$Y-2 = 3186811 \quad \log(Y-2) = 0.5033563$$

$\log(Y+1)$ 等の値はGaussの対数表から引いた

$$a_1 + a_2 = 15.52366 \quad a_1 = 1.99638$$

$$\mu_1 = 9809535 \quad a_2 = 13.52728$$

本 モーメントは曲線が急激に上下するときに適當する方法で修正したものでなければならぬ。これらは既に予習した前例と訂正されているそれは再計算する必要がない。それ以上の事は付録を参照されたい。同じような条件は他の π 。その例にも適用される。

$$(1) \quad V_1 = 2S_2 - S_1(1+S_1) = 2 \times 19.809 - 5.175 \times 6.175 = 766237$$

$$V_2 = 6S_3 - 3V_1(1+S_1) - S_1(1+S_1)(S_1+2) = 6 \times 186.638 - 3 \times 766237 + 5.175 - 5.175 \times 6.175 \times 7.175 = 15.1067$$

$$m_2 = 2776778$$

$$\text{平均値} - \text{モード} = 4.223116$$

$\sqrt{\{B_1(Y+2)^2 + 16(Y+1)\}}$ が $A_1 + A_2$ と m の式の両方に現れる事に注意しておこう。

平均値は $12 + 5/75 \times 5 = 378750$ の年令の所でありモードは $378750 - 223116 \times 5 = 2675742$ の年令の所にある。至差は 28032 である。

$\log y_0$ の計算は次の通りである。

$$\log N = 3.00000$$

$$\text{Colog}(A_1 + A_2) = 2.80901^{(1)}$$

$$m_1 \log m_1 = 7.84123$$

$$m_2 \log m_2 = 1.23177$$

$$\text{Colog}(Y-2)^2 = 2.39570$$

$$\log \Gamma(Y) = 1.50406$$

$$\text{Colog} \Gamma(m+1) = 0.05217$$

$$\text{Colog} \Gamma(m_1+1) = 7.34037$$

$$\log y_0 = 2.17455$$

(前頁下の脚注のつづき)

$$\begin{aligned} V_0 &= 24S_2 - 2V_1 \{2(1+S_2)+1\} - V_2 \{4(1+S_2)(2+S_2)-1\} - S_2(1+S_2) \\ &\quad (2+S_2)(3+S_2) = 24 \times 186.638 - 3 \times 15.1067 \{2 \times 6.175 + 1\} - \\ &\quad 7.66237 \{4 \times 6.175 \times 7.175 - 1\} - (5.175)(6.175)(7.175)(8.175) \\ &= 172.326 \end{aligned}$$

$$B_1 = \frac{H_1^2}{M_1^2} = \frac{(15.1067)^2}{(7.66237)^2} = 0.5072755$$

$$B_2 = \frac{H_2}{M_2} = \frac{172.326}{(7.66237)^2} = 2.935110$$

1) 余対数

$\log \frac{m}{n} = \log m - \log n$ となるが減法は不便でもあり間違いを起し易いのでこれを加法に通ずためには余対数 (Colog) を使用する事がある。ある数の Colog とはその数の逆数の対数である即ち。

$$\text{Colog } n = \log \frac{1}{n} = \log 1 - \log n = -\log n \quad (68 \text{ 頁下へつづく})$$

ここで勿論 $\log \Gamma(m_1+1) = \log(3.776778) = \log 2.776778 + \log 1.776778 + \log \Gamma(1.776778)^{(1)}$ 。最後の $\log \Gamma(1.776778)$ の値は本書の終りにつけてある表から引いたものである。

この表に対しての計算は統計量を補整するための曲線

$$y = 149.47 \left\{ 1 + \frac{x}{1.77638} \right\}^{1.469833} \left\{ 1 - \frac{x}{13.52728} \right\}^{2.776778}$$

として与えられる。こゝで原表は 2675742 年であり単位は5年である。

次の表は今与えられたものからの指標の計算法を示す。

オ(ヌ)欄とオ(ヨ)欄では一定のオ一階差 (first difference) をもっている。即ち $\frac{1}{5}$ 或は 0.500907 と $\frac{1}{6}$ 或は 0.073925 である。どれか一つの奥でその値が計算されチェックされれば、他の項は次々に連続して作られる。

故に $\text{Colog } n = 0 - \log n$ より求められるが仮数を正数に書きかえておく

例 $\text{Colog } 1.634$ を求める

$$\log 1.634 = 0.213252$$

$$\therefore \text{Colog } n = 0 - 0.213252 = -1 + (1 - 0.213252) = 7.786748$$

注 意

ある数の $\log$ と Colog との和は 0 即ち 指標の和は -1 。仮数の和は $+1$ である。

上例では $\log 1.634 + \text{Colog } 1.634 = 0.213252 + 7.786748 = 7 + 1 = 0$ 。

1)

Γ 函数の性質 $\Gamma(m+1) = m\Gamma(m)$ を使って

$$\log \Gamma(m+1) = \log m + \log \Gamma(m)$$

$\Gamma(m)$ について同じようにすれば上の計算が出来る。

年 齢 (1)	$1 + \frac{x}{a_1}$ (2)	$1 - \frac{x}{a_2}$ (3)	$\log(2)$ (4)	$\log(3)$ (5)	$M_x(4)$ 列 (6)	$M_x(5)$ 列 (7)	$\log y_0$ $= \log(8)$ (8)	y_2 (9)	y_0 (10)
17	002228	1144229	334792	405854	73229	01626	16601	457	44
22	052319	107037	71866	002955	78847	00821	21404	1382	137
27	102410	099644	001034	799846	00042	79957	21745	1496	149
32	152501	092252	018327	096498	00751	09027	21525	1421	142
37	202592	084859	030662	092870	01267	08020	21023	1266	127
42	252683	077466	040257	088911	01650	06921	20317	1076	108
47	302774	070074	048111	084566	01972	05711	19429	877	88
52	352865	062681	054760	079714	02244	04367	18357	685	69
57	402956	055287	060526	074264	02481	02853	17080	510	51
62	453047	047896	065615	068030	02689	01122	16657	360	36
67	503136	040504	070169	060750	02876	29100	13722	236	24
72	553229	033111	074291	051977	03046	06670	11461	140	14
77	603320	025719	078065	041025	03199	03623	08568	72	7
82	653411	018326	081519	026307	03341	39635	04622	29	3
87	703502	010934	084726	003878	03472	03307	78525	07	1
92	753593	003541	087714	254913	03695	79769	35050	-	-

(1) 又は年齢がモートド(2670902)を引いて年齢を算出して用いる。

才(4)欄～才(7)欄は表頭に説明してある通りに行える。対数を取るときには5軒以上の数表を用いる方が一般的である事を注意しておこう。特に此或は此が大きいときは尚さらである。此(2409833)と771866のような数に掛ける場合には少しばかりの注意が必要である。

若し計算機が用えるなら此を乗数板の上におき-228134を乗すればよい。その結果の-21153は他の対数にそれを加えられるように79847の形に通しておかねばならない。

才(10)欄は面積を与え、これは59頁の諸公式の一つの応用によつて作られている。最初の群の面積は曲線が年令167775から出発しているから別個に取扱ねばならない。したがつて群の区間は他の場合は5年であるが、ここでは2225の長さである。面積を計算するすぐれた方法は区間の中点と端点に対する数値を計算する事である。Simpsonの公式即ち

$$\int_0^1 y_x dx = \frac{1}{6} \{ y_0 + 4y_1 + y_2 \}$$

の応用は区間の色々違った長さを考慮に入れてその結果に2225/5を掛ける事を注意しておこう。縦線は221であり区間の端点での数値は1155である。出発点での縦線は勿論0である。面積は近似に

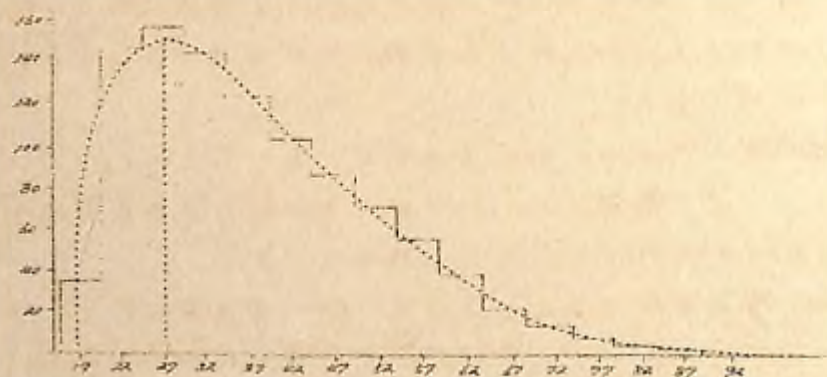
$$\frac{2225}{5} \times \frac{1}{6} \{ 0 + 4 \times 921 + 1155 \} = 44$$

である。縦線を計算する場合には、曲線の出発点と終点である

65頁の注意の外で与えた形を用いる方がよい事が分る。それは a_1 と a_2 の逆数を用いなくともよい。 $\log x$ と $\log(a_1 + a_2 - x)$ の欄はGRANDの対数を使って連続的に作る事が出来る。最初の値は計算されねばならない。そしてチェックとして一つ或は二つの他の値も計算した方がよからう。

* より正確さを与えるためにはもっと多くの縦線を用いるか或は不完全B-函数表を用いる。

I 型



公式の証明*

曲線の方程式は

$$y = y_0 \left(1 + \frac{x}{a_1}\right)^{m_1} \left(1 - \frac{x}{a_2}\right)^{m_2}$$

である。ここは $m_1/a_1 = m_2/a_2$

$$\text{今 } a_1 + a_2 = b \quad z = \frac{a_1 + x}{a_1 + a_2} \quad \text{と置く}$$

又 $x = -a_1$ から $x = +a_2$ までの面積は全度数 N である。

$$\begin{aligned} N &= \int_{-a_1}^{+a_2} y_0 \left(1 + \frac{x}{a_1}\right)^{m_1} \left(1 - \frac{x}{a_2}\right)^{m_2} dx \\ &= \int_{-a_1}^{+a_2} \frac{y_0}{a_1^{m_1} a_2^{m_2}} (a_1 + x)^{m_1} (a_2 - x)^{m_2} dx \\ &= \int_0^1 \frac{y_0}{a_1^{m_1} a_2^{m_2}} [z(a_1 + a_2)]^{m_1} [(1-z)(a_1 + a_2)]^{m_2} (a_1 + a_2) dz \\ &= \int_0^1 \frac{y_0 (a_1 + a_2)^{m_1 + m_2 + 1}}{a_1^{m_1} a_2^{m_2}} z^{m_1} (1-z)^{m_2} dz \\ &= \frac{y_0 (m_1 + m_2)^{m_1 + m_2} (a_1 + a_2)}{a_1^{m_1} a_2^{m_2}} B(m_1 + 1, m_2 + 1) \end{aligned}$$

$$\text{或は } y_0 = \frac{N}{b} \cdot \frac{a_1^{m_1} a_2^{m_2}}{(m_1 + m_2)^{m_1 + m_2}} \cdot \frac{\Gamma(m_1 + m_2 + 2)}{\Gamma(m_1 + 1) \Gamma(m_2 + 1)}$$

* 漸化式とΓ函数又はB函数についての知識のない読者はこの型或は他の型に対する公式の証明を読む前に附録IIを覗ぐべからう。

面積に対して今与えたものと同じモーメント法を用いれば、
 $x = -a_1$ を通るy軸に平行な直線まわりのx次のモーメントは次の様である事がわかる。

$$\begin{aligned} N \mu'_x &= \int_{-a_1}^{+a_2} \frac{y_0}{a_1^{m_1} a_2^{m_2}} (a_1 + x)^{m_1} (a_2 - x)^{m_2} (a_1 + x)^x dx \\ &= \int_0^1 \frac{y_0 (a_1 + a_2)^{m_1 + m_2 + 1}}{a_1^{m_1} a_2^{m_2}} z^{m_1 + x} (1-z)^{m_2} dz \\ &= \frac{y_0 (m_1 + m_2)^{m_1 + m_2} b^{m_1 + 1}}{a_1^{m_1} a_2^{m_2}} \cdot \frac{\Gamma(m_1 + x + 1) \Gamma(m_2 + 1)}{\Gamma(m_1 + m_2 + x + 2)} \end{aligned}$$

ここで $\Gamma(P) = (P-1) \Gamma(P-1)$ だから $x = -a_1$ を通るy軸に平行な直線まわりのモーメントは次の様になる。

$$\begin{aligned} \mu'_x &= \frac{b (m_1 + 1)}{m_1 + m_2 + 2} \\ \mu'_z &= \frac{b^2 (m_1 + 1)(m_1 + 2)}{(m_1 + m_2 + 2)(m_1 + m_2 + 3)} \end{aligned}$$

等々

平均値まわりのモーメントを得るために原点を幾え $m'_1 = m_1 + 1$ $m'_2 = m_2 + 1$ 、 $\gamma = m'_1 + m'_2$ とおけば、我々は次式をうる。

$$\begin{aligned} \mu_2 &= \frac{b^2 (m'_1 \cdot m'_2)}{\gamma^2 (\gamma + 1)} \\ \mu_3 &= \frac{2 b^3 m'_1 m'_2 (m'_1 - m'_2)}{\gamma^3 (\gamma + 1) (\gamma + 2)} \\ \mu_4 &= \frac{3 b^4 m'_1 m'_2 \{m'_1 m'_2 (\gamma - 1) + 2 \gamma\}}{\gamma^4 (\gamma + 1) (\gamma + 2) (\gamma + 3)} \end{aligned}$$

$\beta_1 = \mu'_2 / \mu_1^2$, $\beta_2 = \mu'_3 / \mu_1^3$, $E = m'_1 \cdot m'_2$ とかく事により64-65頁の方程式をうるためにこれらの式を簡単にする事が出来る。

$$\beta_1 = \frac{2(\gamma^2 - 4E)(\gamma + 1)}{E(\gamma + 2)^2} \quad \text{或は} \quad \frac{\beta_1 (\gamma + 2)^2}{4(\gamma + 1)} = \frac{\gamma^2}{E} - 4$$

また

$$\beta_2 = \frac{3(\gamma + 1)\{2\gamma^2 + E(\gamma - 1)\}}{E(\gamma + 2)(\gamma + 3)} \quad \text{或は} \quad \frac{\beta_2 (\gamma + 2)(\gamma + 3)}{3(\gamma + 1)} = \frac{2\gamma^2}{E} + \gamma - 1$$

$\frac{Y^2}{E}$ を消去して

$$\frac{\beta_1 (Y+2)^2}{2(Y+1)} - \frac{\beta_2 (Y+2)(Y+1)}{2(Y+1)} = -\delta - Y + \delta$$

$Y+2$ を割れば

$$Y = \frac{\delta(\beta_1 - \beta_2 - 1)}{3\beta_1 - 2\beta_2 + \delta}$$

方程式 $\frac{\beta_1 (Y+2)^2}{2(Y+1)} = \frac{Y^2}{E} - \delta$ から

$$E = \frac{Y^2}{\delta + \frac{1}{2}\beta_1 \frac{(Y+2)^2}{Y+1}}$$

また μ_2 の方程式から

$$\delta^2 = \frac{\mu_2 (Y+1) Y^2}{E}$$

他の方程式は $Y = m_1' + m_2'$ $E = m_1' m_2'$ から直に導ける。
モードと平均値の間の距離は

$$a_1 - \mu_1 = a_1 - \delta m_1' / (m_1' + m_2')$$

である。これは容易に与えられている形に改形出来る。距離に対する一般値 (Type に無関係な) は才 IV 節 §5 に与えられている。

才 = 主要型 (IV 型)

$$y = y_0 \left(1 + \frac{x^2}{a^2}\right)^{-\pi} e^{-\sqrt{A} \arctan \frac{x}{a}}$$

原典は平均値 + $\frac{\sqrt{a}}{Y}$ である。

計算しなければならぬ値の順序は次の通りである。

$$Y = \frac{\delta(\beta_1 - \beta_2 - 1)}{2\beta_1 - 3\beta_2 - \delta}$$

$$m = \frac{1}{2} (Y+2)$$

$$V = \frac{Y(Y-2)\sqrt{\beta_1}}{\sqrt{\beta_1(Y-1) - \beta_2(Y-2)}}$$

$$a = \sqrt{\frac{\beta_1}{2}} \sqrt{\beta_1(Y-1) - \beta_2(Y-2)}$$

$$y_0 = \frac{N}{a F(Y, V)}$$

$$E - D = \text{平均値} - \frac{1}{2} \frac{\mu_2 (Y-2)}{\mu_2 (Y+2)}$$

注 意

曲線は歪んでいて両方向に無限に広がっている。 μ_2 と V は互に異符号である。例えば μ_2 が正のときには V は負である。曲線を計算する簡単な方法は

$$x = a \tan \theta \quad y = y_0 \cos^{Y+2} \theta e^{-V\theta}$$

と置けばよい。 θ が $10^\circ, 20^\circ, 30^\circ, \dots$ 等の値を取るときに x, y は決まる。これは x と y の値に一つの対応を与える。しかし y の値は等距離 (equidistant) ではない。 $e^{V\theta}$ の計算で θ の値は角制度 (circular measure) を取らねばならない。もし等距離縦線が正確に計算する事が必要でも二重形式 (double form) のためにはとんどうする所ない。また $\log(1+x^2)$ や $\text{Arctan } x$ のすぐれた表を持っていれば、ある特別な縦線の計算も非常に簡単な事になつて来る。 $F(Y, V)$ の計算と意味は、証明のところで取扱われている。この函数の対数は *Tables for Statisticians* に載せてある。

Y が可成り大きい時、 y_0 の正確な近似値は次によつて与えられる。ここで $\tan \phi = \frac{V}{Y}$ である。

$$\frac{N}{a} \sqrt{\frac{Y}{2\pi}} \frac{\exp\left(\frac{\cos^2 \phi}{2} - \frac{1}{2Y} - \phi V\right)}{(\cos \phi)^{Y+1}}$$

我々は 前頁に示されたような原典を持つ最も簡単な目取も短い粒子の方程式を得る事が明らかである。それは一概に用いられていてその結果は既に与えられているが、この原典は如何なる物理的意味も持たないしまた平均値を原典にもつ形は使用に際してよりめんどうだと言われる事が多い。即ち

$$y = y_0 \left\{1 + \left(\frac{x}{a} - \frac{V}{Y}\right)^2\right\}^{-\pi} e^{-\sqrt{A} \arctan \left(\frac{x}{a} - \frac{V}{Y}\right)}$$

57 頁の折込の表 VII を参照せよ。

$x=0$ の時のこの式の値。例えば平均値で観測の値は

$$y_0 \left\{ 1 + \frac{y_0}{r} \right\}^{-x} e^{-\lambda \arctan(\frac{y}{r})} = \frac{N}{\sigma} \frac{1}{H(r, v)}$$

である。ここに $H(r, v)$ は $F(r, v)$ に関係した函数である。これの対数が *Tables for Statisticians* の表にされている。読者はこの曲線に重要な注意が必要である事がすぐわかるでしょう。それはすべての Pearson 型曲線のうちで最も困難なものである。

例

次のほとんど対称的な分布を示す数値は、1 週間の病人の数が異なる正規曲線によって表わされている時に Sutton の *Sickness-Tables* (男性-全期間) から両方の危険にさらされたものを表わす。

中心年令	危険にさらされた数	IV 型によって補整された数値
5	10	6*
10	13	16
15	41	49
20	115	135
25	326	321
30	675	653
35	1113	1108
40	1528	1535
45	1692	1712
50	1530	1522
55	1122	1074
60	610	604
65	255	274
70	86	102
75	26	32
80	8	8
85	2	2
90	1	1
95	1	—

*. この群は曲線の残りの面積として作られている。

次の値が得られた

$$\text{平均値} = 445772339 \quad B_1 = 00053666$$

$$\mu_1 = 4527608$$

$$B_2 = 3.167897$$

$$\mu_2 = -0.705687$$

$$\mu_3 = 6498048^{(1)}$$

$$K = 0.0125$$

場合の数が非常に大きいときには K の標準誤差が小さいから IV 型が用いられた。(X 章参照)

$$Y = 40.12143$$

$$V = 4450398 \quad (\mu_1 \text{ が負であるからこれは正である})$$

$$a = 13.39152$$

$$m = 21.06072$$

$$SK = -0.03313$$

上の単位が 5 年毎であるので 1 年毎の単位に変えて計算すれば a は 66.9576⁽²⁾ $a^2 = 4453325$ になる。

$$\text{原 典} = \text{平均値} + \frac{\sqrt{a}}{Y} = 52.504394$$

曲線を引く場合に必要モードは 4492987 である。

Y が大きいので y_0 の近似が用いられた。

$$\tan \phi = \frac{4450398}{40.12143} \quad \log \tan \phi = \log \tan 6'19'' \frac{5.225}{11537} \quad \text{故に}$$

$\log \cos \phi = 7.9973446$ ここでこれから y_0 は 273.3647 である事がわかる。これらの値は *Tables for Statisticians* の表で照合される。二重過程 (double process) による観測の計算は次のようである。

0	年単位の年令 x	$-4450398 \log e$	$4212143 \log \cos \phi$	$\log y$	y
0'	0	...	...	2.23575	273.37
1'	1.1687	7.96637...	7.99721	2.44133	274.38
2'	2.3382	7.93253...	7.98885	2.35513	228.11

(1) この値は正確には 65042280 である。したがってこれ以下の項にわたって少しづつ修正しなければならぬ。補整への影響は少ない。

(2) $13.39152 \times 5 = 66.9576$

才(2)欄は $\tan \theta$ の表からの値に a をかける事によって直接に計算出来る。また x は年単位にする必要があるから $1337152 \times 5 = 668576$ は a に代つて用いられるべきである。才(4)欄は $\log \cos \theta$ に $\gamma + 2$ を掛けて求められる。才(3)欄は次々に加えたものである。

θ が負の時は、才(3)欄は(4)欄から引かればならない。即ち、それは負から正に終る。各々の場合の才(5)欄は(4)欄+(3)欄+ $\log y_0$ から求められる。

この型の曲線を画く時にモードの高さと位置が書きこまれ、それから対応している点。例えば $x = +11687$ と $y = 25138$ が書き入れられる。

右側の点で曲線が最大値になる事は注意されねばならない。直接に計算を行うならば次の各欄を作らねばならない。

$\frac{x}{a}$ (1)	$1 + \frac{x^2}{a^2}$ (2)	$\log(1 + \frac{x^2}{a^2})$ (3)	Arctan $\frac{x}{a}$ 度単位 (4)	(4) 列の 角測定 (5)	(5) 列 $\times$ (-1/deg. R) (6)	-11687 列 (7)	$\log y_0$ + (6) + (7) (8)	$y = \text{antilog}$ (8) (真数) (9)

(2) 欄は $\Delta(1+x^2) = 2x + 1$ だから階差から求められる。Arctan $\frac{x}{a}$ は逆角の正接の表を用いて求められる。(4) 欄から(5) 欄をうるには Chamber の *Mathematical Tables & Tables for Statisticians* を利用すればよい。

度、分、秒の逆補脚のわずらわしい計算は $\tan \theta$ の表の各項を 0 から順次上の方へ数へ上げる事によって避ける事が出来る。例えば Chamber の表は、次の形で各分毎に対する正接の値を与えている。

	0°	1°	2° 等
0	.0000000	(10) .0174551	(120) .0344208
1	.0002907	(11) .0177460	(121) .0352120
2	.0005818	(12) .0180370	(122) .0355033
3	.0008727	(13) .0183280	(123) .0357945
等	...	...	...

もし見出し 1° の欄で括弧の中に示されている数 10, 11 等を求めずの欄で 120, 121 等を求めれば、分単位逆補脚を作る事が出来る。そのとき角測定での 1 分は 0.0002908882 のごとくであるから換算因子 (conversion factor) を掛ける事について必要とする値が求められる。しかしながら実際上は一つの乗数の中には $(-1/\log_{10} e)$ と一緒に組合せられているだろう。また(6) 欄は我々の例では 0.003517003 である乗数を掛ける事によって才(4) 欄から直接に見出す事が出来る。

印刷された表に“分”数値を挿入する仕事は些細な事である。なぜかと言えば、我々がなすべき必要なすべてのものは各欄の頭の数値の下に“分”の数値を書き、それから補脚が行われたときに欄外“分”を視てそれに加えればよいからである。

$\log(1+x^2)$ と Arctan θ の表は *Treats for Computers No. XXIII* に与えられているしまたこれらの表は計算が簡単である。

証 明

$$y = y_0 \left\{ 1 + \frac{x^2}{a^2} \right\}^{-\pi} e^{-\gamma \text{Arctan } \frac{x}{a}}$$

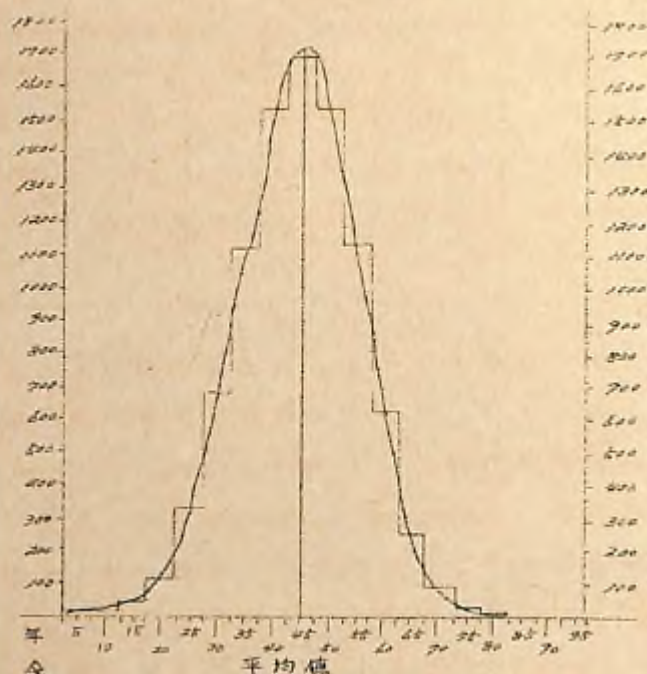
$\theta = \text{Arctan } \frac{x}{a}$

また $\left\{ 1 + \left(\frac{x}{a} \right)^2 \right\}^{-\pi} = \left\{ 1 + \tan^2 \theta \right\}^{-\pi} = (\sec^2 \theta)^{-\pi} = \cos^{2\pi} \theta$

$$\therefore y = y_0 \cos^{2\pi} \theta e^{-\gamma \theta}$$

$$\begin{aligned} \because \sec \theta &= \frac{1}{\cos \theta} \\ \sec^2 \theta &= \frac{1}{\cos^2 \theta} \\ \therefore (\sec^2 \theta)^{-\pi} &= (\cos^2 \theta)^{\pi} \\ &= \cos^{2\pi} \theta \end{aligned}$$

(IV 型)



$$\text{したがって } N = \int_{-\infty}^{\infty} y_0 \left\{ 1 + \frac{x^2}{a^2} \right\}^{-\gamma} e^{-\sqrt{\lambda} \arctan \frac{x}{a}} dx$$

$$\text{変数変換 } \tan \theta = \frac{x}{a} \quad x = a \tan \theta \quad dx = a \sec^2 \theta d\theta = \frac{a}{\cos^2 \theta} d\theta$$

$$\text{によって } = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} y_0 \cos^{2\gamma} \theta e^{-\sqrt{\lambda} \theta} \frac{1}{\cos^2 \theta} d\theta$$

$$= y_0 a \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos^{r-2} \theta e^{-\sqrt{\lambda} \theta} d\theta \quad \text{但し } r = 2\gamma - 2$$

$\cos \theta$ に対して $\sin \phi$ を代入すると $\frac{1}{2}\pi = \theta + \phi$ 、そして上下限が変わるから

$$= y_0 a e^{-\frac{1}{2}\sqrt{\lambda}\pi} \int_0^{\pi} \sin^{r-2} \phi e^{\sqrt{\lambda}\phi} d\phi$$

$$= y_0 a F(r, \sqrt{\lambda})$$

原点のまわりの n 次のモーメントは

$$\mu_n' = \frac{1}{N} \int_{-\infty}^{\infty} y x^n dx$$

$$= \frac{1}{N} \int_{-\infty}^{\infty} y_0 x^n \left\{ 1 + \frac{x^2}{a^2} \right\}^{-\gamma} e^{-\sqrt{\lambda} \arctan \frac{x}{a}} dx$$

前と同じ変数変換によって、

$$= \frac{1}{N} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} y_0 a^{n+1} \cos^{2\gamma-n-2} \theta \tan^n \theta e^{-\sqrt{\lambda} \theta} d\theta$$

$$= \frac{1}{N} y_0 a^{n+1} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos^{r-n} \theta \sin^n \theta e^{-\sqrt{\lambda} \theta} d\theta$$

$$= \frac{y_0 a^{n+1}}{N} \left[-\frac{\cos^{r-n} \theta \sin^{n-1} \theta e^{-\sqrt{\lambda} \theta}}{r-n+1} + \int \left\{ \frac{\cos^{r-n} \theta}{r-n+1} \left[\sin \theta \cos \theta e^{-(n-1)\sqrt{\lambda} \theta} - \sqrt{\lambda} \sin^{n-1} \theta \right] \right\} d\theta \right]_{-\frac{\pi}{2}}^{\frac{\pi}{2}}$$

$\sin^{n-1} \theta e^{-\sqrt{\lambda} \theta}$ を一つの部分としました $\cos^{r-n} \theta \sin \theta$ を他の部分として部分積分により上をうる。そのとき

$$\int \cos^{r-n} \theta \sin \theta d\theta = \frac{-\cos^{r-n+1} \theta}{r-n+1}$$

を用いる。

したがって、 θ が $\frac{\pi}{2}$ 或は $-\frac{\pi}{2}$ のとき $\cos^{r-n+1} \theta \sin^{n-1} \theta e^{-\sqrt{\lambda} \theta} = 0$ であるから結局 μ_n' として次をうる。

$$\mu_n' = \frac{y_0 a^{n+1}}{N(r-n+1)} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \left\{ (n-1) \cos^{r-n+2} \theta \sin^{n-2} \theta e^{-\sqrt{\lambda} \theta} - \sqrt{\lambda} \cos^{r-n+1} \theta \sin^{n-1} \theta e^{-\sqrt{\lambda} \theta} \right\} d\theta$$

$$= \frac{a}{r-n+1} \left\{ (n-1) a \mu_{n-2}' - \sqrt{\lambda} \mu_{n-1}' \right\}$$

この式において $n=1$ とおけば

$$\mu_1' = \frac{y_0 a^2}{N} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos^r \theta \tan \theta e^{-\sqrt{\lambda} \theta} d\theta$$

$$= \frac{y_0 a^2}{N r} \left\{ -\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sqrt{\lambda} \cos^r \theta e^{-\sqrt{\lambda} \theta} d\theta \right\}$$

$$= -\frac{a \sqrt{\lambda}}{r}$$

$$\therefore N = y_0 a \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos^r \theta e^{-\sqrt{\lambda} \theta} d\theta \quad \text{であるから}$$

$M_0 = 1$ と上の式に対する最後の式の結果を用い二つの明確になったモーメント)によつて、

$$M_1 = \frac{a}{r(r-1)} (r+V^2)$$

$$M_2 = -\frac{a^2 V}{r(r-1)(r-2)} (3r-2+V^2)$$

$$M_3 = \frac{a^3}{r(r-1)(r-2)(r-3)} \{3r(r-2)+V^2(4r-3)+V^4\}$$

重心線にこれらのモーメントを関係させるために84頁の公式に於て、

$$d = M_1 = -\frac{aV}{r}$$

とおく事により次式をうる、

$$M_2 = \frac{a^2}{r^2(r-1)} (r^2+V^2)$$

$$M_3 = -\frac{2a^3 V(r^2+V^2)}{r^3(r-1)(r-2)}$$

$$M_4 = \frac{3a^4(r^2+V^2)\{(r+3)(r+V^2)-3r^2\}}{r^4(r-1)(r-2)(r-3)}$$

今とし $Z = r^2 + V^2$ とおき前のように、

$$B_1 = \frac{M_1}{M_0} \quad \text{また} \quad B_2 = \frac{M_2}{M_0}$$

とかくならば

$$-\frac{B_1(r-2)^2}{2(r-1)} = 3\frac{r^2}{Z} - 3$$

$$-\frac{B_2(r-2)(r-3)}{3(r-1)} = r+3 - \frac{3r^2}{Z}$$

をうる。この式を加えて両辺を $r-2$ で割る事によつて、

$$r = \frac{6(B_2 - B_1 - 1)}{2B_2 - 3B_1 - 4}$$

$$Z = \frac{r^2}{1 - \frac{B_1(r-2)^2}{12(r-1)}}$$

明らかに $V^2 = Z - r^2$ であるから74頁の他の諸公式も直に導け

る。

幾座標の最大長での接線は x 軸に平行であるからモードの値直は $\frac{dy}{dx} = 0$ のところである。その長において、例えば、

$$y_0 \left\{ 1 + \frac{x}{a} \right\}^{-N} e^{-\text{Arctan} \frac{x}{a}} \left[-\frac{2\pi x}{a^2} - \frac{V}{a} \right]$$

は0である。それには三つの場合がある。即ち、 $x = -\infty$ 、

$x = +\infty$ 、 $\frac{2\pi x}{a^2} + \frac{V}{a} = 0$ 即ち $x = -\frac{Va}{2\pi}$ のような x の値の場合である。

原長から平均値までの距離は M_1 或は $-\frac{Va}{r}$ である。ゆえに平均値とモードの間の距離は $-\frac{2Va}{r(r+2)}$ でこれは74頁での V と a の値を挿入したときは74頁に与えた方程式になる。

この種の型の曲線に対する y_0 の計算の他の例を与える事は有用であらう。また我々は曲線を作る事が出来る。ここでは $r = 29.590$ 、 $V = 17.886$ 、 $a = 13.650$ 、 $N = 2.162$ 、したがつて $\tan \phi = 0.67205$ 、 $\phi = 33^\circ 54' \frac{8}{22}$ 、 $\cos \phi = 0.82998$ 、 $\log \cos \phi = 7.91707$ 、前測度による ϕ は 0.57172 である。

$$\log N = 3.33486$$

$$\text{Colog } a = 2.86486$$

$$\pm \log r = 0.73557$$

$$\log \frac{1}{\sqrt{2\pi}} = 7.60091$$

$$\frac{\cos^2 \phi}{3r} = 0.00776$$

$$-\frac{1}{12r} = -0.00282$$

$$-\phi V = -11.76700$$

$$-11.762 \times \log_e e = 3.89183$$

$$\text{Colog } (\cos \phi)^{r+1} = 2.47564$$

$$7.90367$$

$$y_0 = 0.80107$$

今考へた式は仮りに V が余り小さくなければ實際目的には充分正確である。もし $V < 2$ なら *Tables for Statisticians* の表に

用いねばならない。

第 三 主 要 型 (Type VI)

$$y = y_0 (x-a)^{p_1} x^{p_2}$$

原点は曲線の出発点の前 a なる点である。

順次に計算されるべき値は、

$$Y = \frac{6(B_2 - B_1 - 1)}{6 + 3B_1 - 2B_2}$$

$$a = \frac{1}{2} \sqrt{M_1 \sqrt{B_1(Y+2)^2 + 12(Y+1)}}$$

$p_2 = -p_1$ は次式によつて与えられる。

$$\frac{Y-2}{2} \pm \frac{Y(Y+2)}{2} \sqrt{\frac{B_1}{B_1(Y+2)^2 + 12(Y+1)}}$$

$$y_0 = \frac{N a^{p_1 - p_2 - 1} \Gamma(p_1)}{\Gamma(p_1 - p_2 - 1) \Gamma(p_2 + 1)}$$

$$\text{原点} = \text{平均値} - \frac{a(p_1 - 1)}{p_1 - p_2 - 2}$$

$$\text{モード} = \text{平均値} - \frac{1}{2} \frac{M_1}{M_2} \frac{Y+2}{Y-2}$$

もし、平均値を原点に持つ曲線の式を表わすならば、(5)頁折込みの VI 表をみよ)

$$A_1 = \frac{a(p_1 - 1)}{(p_1 - 1) - (p_2 + 1)} \quad A_2 = \frac{a(p_2 + 1)}{(p_1 - 1) - (p_2 + 1)}$$

$$y_2 = \frac{N(p_2 + 1)^{p_1} (p_1 - p_2 - 2)^{p_2 - p_1} \Gamma(p_1)}{a(p_1 - 1)^{p_1} \Gamma(p_1 - p_2 - 1) \Gamma(p_2 + 1)}$$

注 意

この曲線の範囲は a から ∞ である。計算の方法は I 型と同じである。 M_2 が負けなら a も負で範囲は $-\infty$ から $-a$ となる。 Y は常に負で p_1 は p_2 よりも大きい。 p_2 が負なら曲線は J 型になる。 y_0 値は曲線の出発点の a に関係しているので、他のどんな頻度分布

にも対応はしない。

読者は平均値を原点にした場合の計算はとつと容易な事がわかるだろう。数値例では両方の形が示されている。

例。

1863 ~ 1873 の間の保険証券の支払制限された保険加入者を 10 年毎の群で加え上げそれを 100 で割った次の系列が得られた。

加入者の数 ÷ 100	VI 型によつて補正された曲線
1	1
56	50
167	168
98	100
34	36
9	10
2	2
1	0.5
368	368

モーメント等は次の通りである。平均値は 167 群の中心后 0.402174 の点である。

$$\begin{aligned} M_1 &= 0.928835 & 1 - p_1 &= -41.03080 \\ M_2 &= 0.893096 & 1 + p_2 &= 7.60950 \\ M_3 &= 4.088800 & p_1 &= 42.03080 \\ B_1 &= 0.9953605 & p_2 &= 6.60950 \\ B_2 &= 4.739347 & a &= 10.37947 \\ K &= 1.895 & \log y_0 &= 46.1821 \\ Y &= -33.42129 \end{aligned}$$

原点は平均値前 12.74270 である。即ち、167 群の中心

これはI型の解の全部を用いる事が出来る。

推移型 Transition Type

誤差の正規曲線

$$y = y_0 e^{-\frac{x^2}{2\sigma^2}}$$

$$C = 2\sigma^2$$

$$y_0 = \frac{K}{\sqrt{2\pi} \mu_2}$$

注 意

この曲線は確率曲線または Gauss の曲線と言うような色々な名称によつて知られて来た。この曲線は Gauss 以前に de Moivre や Laplace によつて研究されている。この曲線は $(p + \frac{q}{x})^x$ において x が無限に大きくなり且つ p, q のどちらもが余り小さくない時の極限である。ここで $p + q = 1$ である。 x が大きくないときでさえも $(\frac{1}{2} + \frac{x}{2})^x$ の正確な表示を与える。

例

次表は(ヌ)欄には親被保険者と割増配当金合計を与えており(ロ)欄には養老被保険事務所の出来た年にしたがつて若干の養老被保険の群分けに起因する予備金を与えている。

誕生の年 グループに 対する中心 年	被保険者と割増配当金 の合計 ÷ 1000		予 備 金 ÷ 1000	
	未 補 整 (1)	補整したもの (2)	不 補 整 (3)	補 整 (4)
17	11	16	2.6	2.6
22	48	45	2.8	2.8
27	124	115	11.5	10.9
32	213	213	27.7	30.1
37	281	287	57.1	58.0
42	295	282	84.7	77.9
47	186	202	74.1	77.7
52	104	105	50.5	52.2
57	40	40	23.2	25.0
62	15	11	12.2	8.0
67	3	3	1.3	2.0
計	1319	1319	347.7	347.7

次表にはモーメントと定数が示されている。

定 数	被保険者と配当金の和	予 備 金
平均年令	39.202426	43.967213
μ_2	3.066840	2.769635
μ_3	0.650127	0.029805
μ_4	27.02516	22.40663
B_1	0.014653	0.0000418
B_2	2.873346	2.920997
K	-0.04	-0.0002
$\sigma (= \sqrt{\mu_2})$	1.751237	1.664222
σ^{-1}	0.5710248	0.6008813
y_0	300.4760	83.34759

正規曲線の規準値は $K=0, B_1=0, B_2=3$ である。

上に与えられた値はこれらからそんなに大きな違いはない。しかし補整した数値と補整しない数値の比較では予備金曲線 (reserve curve) の場合が、被保険者曲線 (sum assured curve) より良く一致する事を示している。何とならばある程度まで B_2 の値は3に近く B_1 は sum assured curve の場合よりやや大きい値をもつからである。

y_0 の計算のために

$$\text{Calc } \sqrt{2\pi} = 760091657$$

の値が必要である。

補整数値とそうでない数値の比較のための面積を算出するのに概線の計算は望まない。なぜならそれには確率積分の計算された表が使えるからである。

Tables for Statisticians に含まれている W.F. Sheppard によつて与えられた表は非常に便利である。そして下の表の各欄は一つの場合 (reserve curve) の面積の計算に用いられる方法を示している。

年令 x	計算単位 (例えば5年) の原資からの 面積	(左欄) $\times 10^{-4}$	(原資から x までの面 積) を用いる事 によって Sheppard の表 から求められた $(1+d)$ の値	左欄の値 $-x$ から $x+5$ の 年令グループに 対する面積	(戻数合計) 347728 けた面積
14.5	5893443	3541258	-	1.00164*	46
19.5	4893443	2746377	0.99836	0.00862	28
24.5	3893443	2339496	0.99634	0.03139	109
29.5	2893443	1738615	0.95895	0.08667	301
34.5	1893443	1137734	0.87238	0.16806	584
39.5	893443	0.536853	0.70432	0.22985*	799
44.5	2106557	0.064028	0.52553	0.22141	770
49.5	1106557	0.164709	0.74674	0.15018	522
54.5	2106557	1.265790	0.89712	0.07190	260
59.5	3106557	1.366671	0.96962	0.02418	84
64.5	4106557	2.467552	0.99320	0.00572	20
69.5	5106557	3.068443	0.99892	0.00108*	44

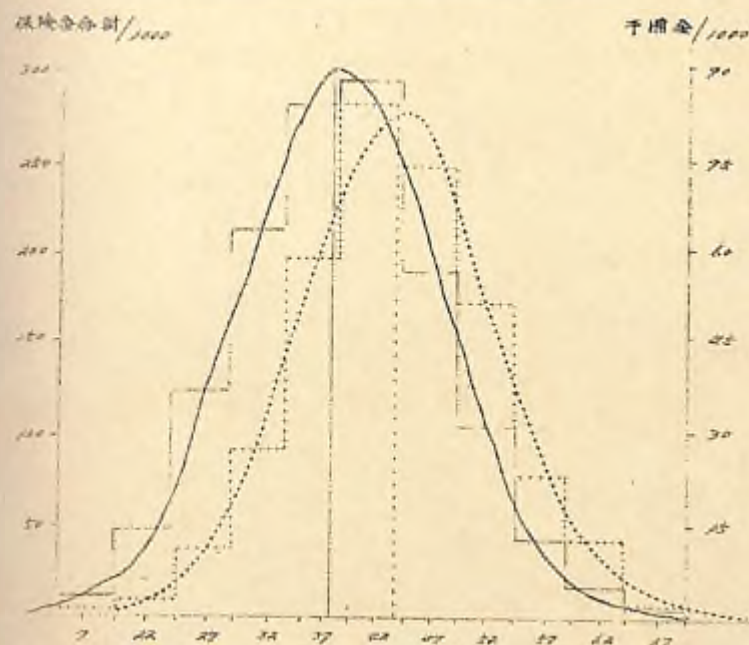
* 19.5以前と69.5以後の残りの面積である。

† $(0.70432 - 0.50000) + (0.52553 - 0.50000)$ 仮定ならば原資を横切つて通っておりその内片はそれぞれ原資の両側にあるからである。

Sheppard の表は標準偏差によつて正規曲線の概線と面積を表わしている。それは次のようである Sheppard は標準偏差は1である事を仮定している。仮の表は h の間隔を用いて読まなければならない。Sheppard の表からの一部抜粋が附表に与えてある。他の大抵の表は $\sqrt{2}$ 倍された標準偏差に基づいている。もし他の表を用いるにはこの区別を心に留めて置かなければならない。

オ(ヌ)欄は計算方法が理解されたときには省略してもよからう。オ(ノ)欄の年令は 17, 22, ... が群の中心年令であると言う仮定を一貫して取られている。

誤差の正規曲線



もし概線が必要なら Sheppard の表の区別を用いねばならない。それを使つて数値に対する曲線を描く事も出来る。reserve 曲線に対する統計量と曲線は実線で示されている。

各群に対する平均予備金は補正数値の平均値によつて得る事が出来る。それはまた任意の未来の価値で得られた予備金の検定に用いる事が出来る。これは決して応用出来る唯一つのラフなチェックではなくて戻数曲線が実際の般所仕事で行なわれるよう反使用法を示しているから興味ある事でもある。

証明

$$\int_0^{\infty} e^{-x^2} dx = \frac{\sqrt{\pi}}{2}$$

を証明するためには

$$\int_0^{\infty} e^{-x^2} dx = K$$

とおく x を ax と変換して

$$\int_0^{\infty} e^{-a^2 x^2} a dx = K$$

$$\therefore \int_0^{\infty} e^{-a^2(1+x^2)} a dx = K e^{-x^2}$$

$$\text{ゆえに} \int_0^{\infty} \int_0^{\infty} e^{-a^2(1+x^2)} a da dx = K \int_0^{\infty} e^{-x^2} dx = K^2$$

$$\int_0^{\infty} e^{-a^2(1+x^2)} a da = \frac{1}{2} \cdot \frac{1}{1+x^2}$$

であるから

$$\frac{1}{2} \int_0^{\infty} \frac{1}{1+x^2} dx = K^2 \quad \because \left(\int \frac{dx}{1+x^2} = \text{Arctan } x \right)$$

$$\therefore K^2 = \frac{\pi}{2} \quad \text{或は} \quad K = \frac{\sqrt{\pi}}{2}$$

したがって

$$\int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi}$$

その他の定数は次のようにしてえられる。部分積分法により

$$\begin{aligned} \int_{-\infty}^{\infty} y_0 e^{-\frac{x^2}{c}} dx &= y_0 \left[x e^{-\frac{x^2}{c}} + \int \frac{2x}{c} e^{-\frac{x^2}{c}} x dx \right]_{-\infty}^{\infty} \\ &= \frac{2y_0}{c} \int_{-\infty}^{\infty} x^2 e^{-\frac{x^2}{c}} dx \end{aligned}$$

或は

$$N = \frac{2K}{c} \mu_2$$

$$\therefore c = 2\mu_2$$

推移型 (Type II)

$$y = y_0 \left(1 - \frac{x^2}{a^2} \right)^{\pi}$$

原典はモードの英 (= 平均値)

$$\pi = \frac{5B_2 - 9}{2(3 - B_2)}$$

$$a^2 = \frac{2\mu_2 B_2}{3 - B_2}$$

$$\begin{aligned} y_0 &= \frac{N \times \Gamma(2\pi+2)}{a \times 2^{2\pi+1} \{ \Gamma(\pi+1) \}^2} \\ &= \frac{N}{a\sqrt{\pi}} \cdot \frac{\Gamma(\pi+\frac{1}{2})}{\Gamma(\pi+1)} \end{aligned}$$

注意と証明

曲線は対称的であるから I 型において $B_2 = 0$ と置くとき $\mu_2 = 0$ となる。同じ理由から $\mu_1 = \mu_2$ も明らかである。 μ が大きければ Γ は近似値を使ってもよい。 μ が正なら曲線は 0 から出発し 最大に達してから再び 0 に下降する。だが、 μ が負なら曲線は無量大のどこから出発して下降し、それから再び無量大に上昇する。

例

G. J. Lidstone の養老保険に関する論文を讀んだ結果の訂論に於いて G. F. Hardy は次のように言っている。“相違く群での誤差は 誤差の正規曲線に非常に良く似た曲線を作る” (J. Inst. Act. XXXIV 987)。また問題になっている系列はむしろ対称的分布の面白い例である。

年単位の未満期期間	平均年令表を用いた事によって合される誤差
0 - 4	11
5 - 9	116
10 - 14	274
15 - 19	451
20 - 24	432
25 - 29	267
30 - 34	116
35	11
等	
	1683

モーメントは15-19群の中心周りに計算されていて0.4985146, 2.161022, 3.104578, 12.60666はそれぞれ1次から4次までのモーメントである。平均値に等し(175+2472573=19992573), Sheppardの補正を用いて次の値をうる。

$$M_1 = 1827172 \quad B_1 = 0.0023706$$

$$M_2 = 0120452 \quad B_2 = 2.548313$$

$$M_3 = 852636 \quad K = -0.007492$$

これはII型を用いられる事を示している。

この型の方程式は

$$y = 4.141766 \cdot a^{-4543077} \cdot y_0 = 462.57$$

を与える。

曲線は対称であるから平均値とモードは一致する。

系列の値を計算するのに次表の配列が便利である。

$\frac{x}{a}$ (1)	$\log(1+\frac{x}{a})$ (2)	$\log(1-\frac{x}{a})$ (3)	(2)+(3) (4)	$\log y_x = M \cdot (4) + \log y_0$ 5

1 - $\frac{x^2}{a^2}$ の値の計算によるよりも上の方法が仕事を早く出来る。実際の例で数値は各群のはじめと中間と終りと計算されている。また Simpson の求積公式 即ち

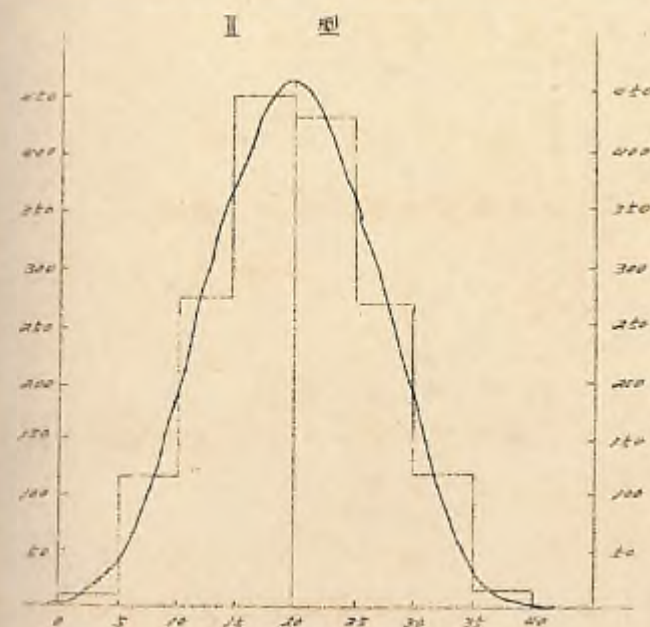
$$\int_0^1 y dx = \frac{1}{3} \{ y_0 + 4y_1 + y_2 \}$$

は面積の系列を得るのに用いられる。

面積と中央縦軸の比較を行えば 中央縦軸を面積の代りに用いた事によって含まれる誤差の概念を与える。そのくいちがい曲線の末尾とモードの近くで最も大きい。

群	未補整値	面積II型	中央縦軸II型	面積正誤差
0-4	11	14	11	22
5-9	116	109	104	95
10-14	274	286	287	270
15-19	451	433	440	455
20-24	432	433	440	455
25-29	267	235	287	269
30-34	116	109	104	95
35	11	14	11	22
	1683	1683	...	1683

曲線は19992573-2271540=-272283から出発して、42.70797に終っている。表の最後の欄は正誤差による補正を与えている。



推移型 (Type VII)

$$y = y_0 \left(1 + \frac{x^2}{a^2}\right)^{-m}$$

原点はモード (= 平均値)

$$\bar{m} = \frac{5\beta_2 - 9}{2(\beta_2 - 3)}$$

$$a^2 = \frac{2M_2\beta_2}{\beta_2 - 3}$$

$$y_0 = \frac{N}{a\sqrt{\pi}} \cdot \frac{\Gamma(m)}{\Gamma(m-\frac{1}{2})}$$

注意と証明

この曲線は IV 型の $\beta = 0$ のときの特別の場合としてとられる。或は II 型で β と a の両方を負にする事から展開することもある。これは $\beta_2 > 3$ のときに生ずる。曲線は対称的で両方向に無限の範囲を持つている。

$$N = \int_{-\infty}^{\infty} y_0 \left(1 + \frac{x^2}{a^2}\right)^{-m} dx$$

$$= 2 \int_0^{\infty} y_0 \left(1 + \frac{x^2}{a^2}\right)^{-m} dx$$

$1 + \frac{x^2}{a^2} = z^{-1}$ と変換する事によって読者は

$$N = \int_0^1 y_0 (a-z)^{-\frac{1}{2}} z^{m-\frac{1}{2}} dz$$

$$= ay_0 B(m-\frac{1}{2}, \frac{1}{2})$$

を示す事が出来よう。或は $\Gamma(\frac{1}{2}) = \sqrt{\pi}$ であるから y_0 が上に示した値を持つ事も示せるでしょう。

例

次は $\beta_2 = 5$, $M_2 = 1$ のときの面積を与え正規曲線での補正が示されている。この例は II 型の例と一緒だ。正規曲線は対称的な

折ですら全く異なる結果を与えない事の注意として扱に立つであらう。

VII 型 $m = 4, \beta^2 = 5$	正規曲線 $\sigma = 1$
1	
1	
2	
4	
7	1
14	5
38	24
93	93
225	278
527	656
1106	1210
1858	1746
2244	1974
1858	1746
等	等

推移型 (Type III)

$$y = y_0 \left(1 + \frac{x}{a}\right)^{ra} e^{-rx}$$

原点はモード

$$Y = \frac{2M_2}{M_2}$$

$$P = ra = \frac{4}{\beta_2} - 1$$

$$a = \frac{2M_2}{M_2} - \frac{M_2}{2M_2}$$

$$y_0 = \frac{N}{a} \cdot \frac{P^{P+1}}{e^P \Gamma(P+1)}$$

$$\text{モード} = \text{平均値} = \frac{M_2}{2M_2}$$

平均値を原点にもつ曲線の表し方は 57 頁の折込みの表 VI をみよ)

$$ye = K \gamma \frac{(P+1)^P}{e^{P+1} \Gamma(P+1)}$$

注 意

曲線は普通鐘型である。P < 0 即ち $\beta > \alpha$ のとき丁型になる。範囲は一方にだけ限界がある。

判定規程は $\alpha\beta = \delta + \gamma\beta$ である。理論的にこれは $K = \infty$ である。しかしながら K が余り大きくなく、また $\alpha\beta$ が近似的に $\delta + \gamma\beta$ に等しければ多くの場合にこの曲線が用いられる⁽¹⁾。 α が正のときは γ と α は正で、そのため範囲はモードの前に距離 α の点で限られる。また α が負のときは γ も α も負で、そのとき範囲はモードよりあとに距離 α の点で限られる。しかしながら $\beta > \alpha$ なら α と γ は異符号を持つ。

例

次の統計量は *Trans Acta Soc. Edinb* IV 44 の論文から取られたもので、母親の年令に対して表示された妻の数を結婚後の年令にしたがつて与えている。特別の系列に対する母の年令は 30 から 34 である。

結婚後の年数	妻の数	Ⅲ型曲線によって相違したものの
1	44	59
2	133	111
3	45	45
4	12	20
5	8	9
6	3	4
7	1	2
8	3	1
合 計	251	251

(1) K が 4 以上ならばかなり実際観測に適合すると言われる。

平均値はオス群の中央よりあと 0.33466ノアの英にある。また中央垂線周りのモーメントは 1.441787, 3.606622, 1.873222ノ、であり、それゆえ $K = -0.44$ である。この K の値が大きいのでⅢ型が用いられる。そして

$$\gamma = 2.7775221 \quad \alpha = -0.078007$$

$$P = -0.0783524 \quad \gamma_0 = 2.148$$

明確さよりもむしろ困難さを示すために本例が上げられたのである。先づ一見して曲線は 0 から出発して最大値に昇りそれから下降する事は期待されてよい。そのかわり我々は期間 0.48192^{*} から曲線が出發する事がわかるからオス群は区間上に長さ 0.31808 のストリップを作る。そのストリップは次の群の値より小さい値を持つている。だがオス群の中にある任意の観測の値を説きとるとそれはオス群の任意の所の観測より大きい。モーメントの観測値に対しての補正は行っていない。



* Ⅲ型の普通の場合のモードは平均値 $-\frac{\alpha}{2.148}$ で与えられる。この場合には $-\frac{\alpha}{2.148} = 1.24075$ だからモードは 0.58391 である。そして観測は {モード - α } = 0.58391 + 0.07801 = 0.66192 から出發する。

証明

この型の式即ち $y = y_0 (1 + \frac{x}{a})^P e^{-rx}$ において $y_0 = P$ と置き $\delta (a+x)$ を x に変数変換すると、 N が全変数であるならば、

$$\begin{aligned} N &= \int_0^\infty y_0 (1 + \frac{x}{a})^P e^{-rx} dx \\ &= \int_0^\infty y_0 x^P a^{-P} e^{-x/a} r^{-P+1} dx \quad \left(\because \frac{dx}{da} = r \right) \\ &= y_0 \frac{e^P}{r P^P} \int_0^\infty x^P e^{-x} dx \\ &= y_0 \frac{a e^P}{P^{P+1}} \Gamma(P+1) \\ \therefore y_0 &= \frac{N P^{P+1}}{a e^P \Gamma(P+1)} \end{aligned}$$

曲線の出発点のまわりの n 次のモーメントは上に与えられた y_0 の値を代入して

$$\begin{aligned} \frac{1}{n!} \int_0^\infty y_0 (1 + \frac{x}{a})^P e^{-rx} (x+a)^n dx \\ = \frac{y_0 e^P}{N P^P r^{P+1}} \int_0^\infty x^{P+n} e^{-x} dx = \frac{\Gamma(P+n+1)}{r^n \Gamma(P+1)} \end{aligned}$$

$\Gamma(P) = (P-1) \Gamma(P-1)$ であるから一次のモーメントは $\frac{P+1}{r}$ である。又二次モーメントは $\frac{(P+1)(P+2)}{r^2}$ である。また三次のモーメントは $\frac{(P+1)(P+2)(P+3)}{r^3}$ である。

統計的止事にこれらの公式を応用するためには重心を通る重線周りのモーメントを求める事が必要である。その(平均値)位置は見出す事が出来る。定義によつて、平均値周りの一次モーメントは0であるから、我々は次をうる。

$$\mu_0 = \frac{P+1}{r}$$

$$\mu_2 = \frac{2(P+1)}{r^2}$$

これらの結果からそれぞれ δ 、 P の値は $2\mu_0/\mu_2$ と $4\mu_0^2/\mu_2 - 1$

となる。

推 移 型 (Type V)

$$y = y_0 x^{-P} e^{-\frac{x}{\delta}}$$

原点は曲線の出発点

$$P = \alpha + \frac{\delta + \sqrt{\delta^2 + \beta}}{\beta}$$

$$\delta = (P+2) \sqrt{\mu_2(P-3)}$$

$$y_0 = \frac{N \Gamma^{P-1}}{\Gamma(P-1)}$$

$$\text{原点} = \text{平均値} - \frac{\delta}{P-2}$$

$$\text{モード} = \text{平均値} - \frac{2\delta}{P(P-2)}$$

δ の符号は μ_2 の符号と同じである。

平均値を原点にもつ曲線に表せば 57 頁の折込の VI 表参照

$$A = \delta / (P-2)$$

$$y_e = \frac{N(P-2)^P}{\delta e^{P-2} \Gamma(P-1)}$$

例

死亡数の次の系列は G. King の論文 "On the rate of mortality amongst female nominees, etc" から取ったものである。

年 令	死 亡 数	Type V による補整値
30-34	1	1
35-39	5	3
40-44	8	6
45-49	12	14
50-54	28	32
55-59	82	68

(ツツツ)

年 令	死 亡 数	Type V による調整値
60-64	128	137
65-69	253	247
70-74	342	381
75-79	525	480
80-84	438	441
85-89	265	261
90-94	53	80
95-99	18	10
100-	4	1
	2,162	2,162

平均値は75.7782605年であり修正されたモーメント等は次の通りである。

$$\mu_2 = 3573346 \quad \beta_1 = 24950379$$

$$\mu_3 = -4752613 \quad \beta_2 = 3,996134$$

$$\mu_4 = 5102583 \quad K = 0.85$$

厳密に言えば IV 型を用いるべきだがこの K の値は 1 から余り離れていないし、またその次の V 型の死数が見出された。

$$P = 37.27145$$

$$Y = -390.6609 \quad (\mu_3 \text{ が負であるから})$$

$$\log y_0 = 56.930518$$

$\log \Gamma(P-1)$ の値には近似値が用いられている。

原典は 131,32606 年でモードは、78.7467 年の筈である。

概算の計算のために用いられ各欄は次の通りである。

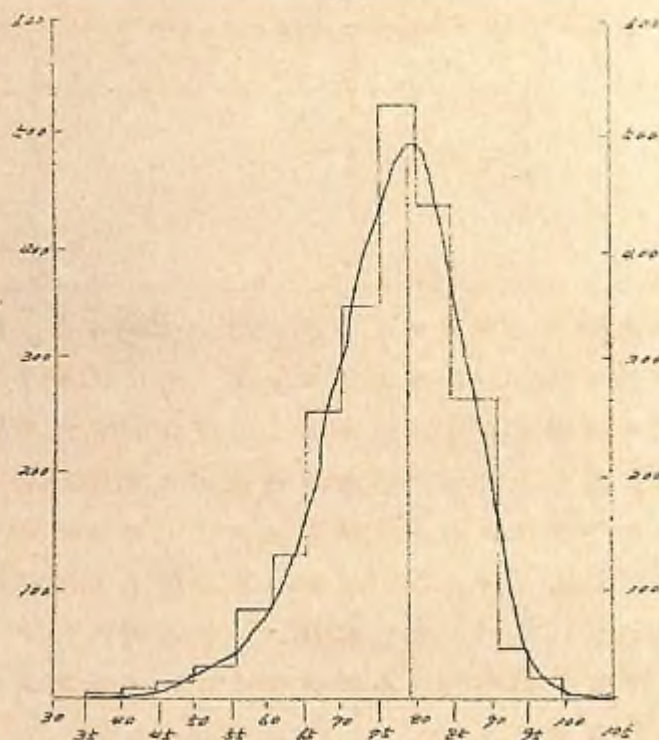
x (1)	$\log x$ (2)	$-P \log x$ (3)	$\frac{1}{x}(-x \log_{10} e)$ (4)	$\log y = \log y_0 + (3) + (4)$ (5)	$y = \text{anti log } (5)$ (6)

この欄は計算機の乗数ダイヤルの上に $x \log_{10} e$ を置き逆数表から得られた値を掛け、結果の読みを負にして計算するのが最もよい。この型の曲線を引くのに注意しなければならない点はモードと原典が同じ場所がないから最大概算が正確な位置とその大きさを与えると言う事に注意が必要である。(IV 型参照)

補整された数値は 90-94 群以下ではもとの統計量に可成り正確に一致している。しかし最後の二つの群にとっては不適当である。この理由は際限ない範囲を持つている IV 型を用いるべきであった事を教えている。

特別の場合には部分的に決定される事もある。なぜなら μ_3 が負である場合の例は μ_3 が正であるときよりむしろ不便でやりにくい。その様な場合には逆の順序(この場合には μ_4 , μ_3 , μ_2 等)に書かれた統計量を考える事がすぐれたチェックである。そうすれば負の符号をさけえられる。

▽ 型



証明

$y = y_0 e^{-\frac{x}{\gamma}} x^{-p}$ において $\frac{x}{\gamma} = z$ において
0 から ∞ まで積分すれば

$$N = y_0 \gamma^{p-1} \Gamma(p-1)$$

或は $y_0 = -\frac{N}{\gamma^{p-1}} \frac{1}{\Gamma(p-1)}$

を得る。上と同じ変換を用いれば原点のまわりの n 次のモーメントは

$$\mu_n' = \frac{y_0}{N} \gamma^{n-p+1} \int_0^{\infty} e^{-z} z^{p-n} dz$$

$$= \frac{y_0}{N} \gamma^{n-p+1} \Gamma(p-n-1)$$

$$= \gamma^n \frac{\Gamma(p-n-1)}{\Gamma(p-1)}$$

これから $\mu_1' = \frac{\gamma}{p-2}$ が与えられる。これは平均値と原点の間の

距離を与える。また

$$\mu_2' = \frac{\gamma^2}{(p-2)(p-3)}$$

$$\mu_3' = \frac{\gamma^3}{(p-2)(p-3)(p-4)}$$

重心を通る垂線にモーメントを移すと

$$\mu_2 = \frac{\gamma^2}{(p-2)^2(p-3)}$$

$$\mu_3 = \frac{4\gamma^3}{(p-2)^3(p-3)(p-4)}$$

ゆえに $B_1 = \frac{\mu_2}{\mu_1^2} = \frac{16(p-3)}{(p-4)^2} = \frac{16}{p-4} + \frac{16}{(p-4)^2}$

また $(p-4)^2 - \frac{16}{B_1}(p-4) - \frac{16}{B_1} = 0$

$p-4$ は方程式の正の根或は γ として取られねばならない。それは上の方程式から

$$(p-2) \sqrt{\mu_2(p-3)}$$

によって与えられ、虚数である。

最大値線の頂上で曲線の接線は x 軸に平行であるからモードの位置は $\frac{dy}{dx} = 0$ のような点である。ここでは例えば

$y_0 x^{p-1} e^{-\frac{x}{\gamma}} \{-p + \frac{x}{\gamma}\}$ は 0 である。 $x=0$ と $x=\infty$ は、曲線が x 軸に接するような場合を与える。他の場合は $p - \frac{x}{\gamma} = 0$

或は $x = \frac{\gamma}{p}$ である事を要求する場合である。即ち、モードは原点から $\frac{\gamma}{p}$ の点である。

普通でない度数分布型

今までは度数曲線の普通のものを取扱つて来た。しかし今市に
関する統計計算の過程ではその代数学的型において普通のものと
は異つたように見える分布型がしばしば見出される。けれどもや
はりこれらの型についても正確に述べる事が出来る。この種の場
合に指示を与える一つの例として、鉄投げやサイコロ投げの實驗

どの系列の数を記録する事から生ずる分布を覚えている。その分布は幾何級数でありまたこれは確率論のよく知られた結果であるがもし $p=0$ であればⅢ型の特別の場合である。なぜならばそのとき我々が欲する指数函数 e^{-rx} を得るからである。

I, II, VII型のある極限の場合は 直線、無限級数で出発して有限級数で終る曲線、度数の二つの別個のブロック及び有限級数で出発し有限度かもしくは無限度のいずれかで0に終る曲線を与える。このうちで最優のものはもちろん我々が既に取扱った指数函数である。

これらの新しい型に対する式を調べる前に保険または他の資料から得られた数々の独特の分布の表を与えるのが有益であろう。

普通でない度数分布型の例

469	45	119	4145	33	48
186	38	100	2028	53	24
166	46	86	982	65	17
134	53	75	480	81	14
122	43	61	266	101	12
112	38	50	132	131	11
- - -	49	39	71	186	10
- - -	41	27	36	350	10
- - -	44	22	17	- - -	10
- - -	52	12	9	- - -	11
- - -	- - -	3	2	- - -	12
- - -	- - -	- - -	1	- - -	20
- - -	- - -	- - -	1	- - -	- - -
- - -	- - -	- - -	1	- - -	- - -
- - -	- - -	- - -	1	- - -	- - -
1189	449	594	8192	1000	219

表にはまれなり型曲線の面積が含まれている。(才6欄) 実際私は保険統計の中にこの型の適当な分布を見付ける事が出来なかつたがそのような分布は10年満期(脱退も含めて)の保険証券に起るようである。或は同じような事は養老保険にも起ると言は

れている。我々はこれらの場合を取扱う事にしよう。しかし我々はもっと重要な型についてもう少し詳しく討論を進めよう。

VII 型

$$y = y_0 \left(1 + \frac{x}{a}\right)^{-m}$$

範囲は $-a$ の点にある無限級数から0の点の有限級数 y_0 までである。 m は次の方程式

$$m^2(4-B_1) + m(9B_1-12) - 24B_1m + 16B_1 = 0$$

の解から求められるが0より小さくても1より大きくてもいけない。

$$a = \pm \frac{1}{2} (2-m) \sqrt{\frac{3-m}{1-m}}$$

$$y_0 = N(1-m)/a$$

$x = -a$ から平均値までの距離は $\frac{a(1-m)}{2-m}$ でありまた $x=0$ から平均値までの距離は $\frac{-a}{2-m}$ である。

m が正のときは a は負である。平均値を原典に持つ形を用いるには(57頁の折込のIV表をみよ)

$$A = \frac{a(1-m)}{2-m} \quad y_0 = \frac{N(1-m)(2-m)^2}{a(1-m)m}$$

この曲線は I 型で $m_2=0$ なる時、即ち、

$$r-2 = r(r+2) \sqrt{\frac{3}{B_1(B_1(r+2)^2+16(r+1))}}$$

のときに I 型の特別な場合である。ここで

$$r = 6(B_1-B_1-1)/(4+3B_1-2B_1)$$

したがって曲線の通否に関する検定は

$$\frac{(4B_1-3B_1)(10B_1-12B_1-18)^2 - B_1(B_1+3)^2(8B_1-9B_1-12)}{(3B_1-2B_1+6)\{B_1(B_1+3)^2+4(4B_1-3B_1)(3B_1-2B_1+6)\}}$$

或は入が0である事である。

VII型に対する判定規準は次の三つの場合にわけらる事が出来る。

- ① I 型の特別の場合 ② $\lambda=0$ ③ $5B_1-6B_1-7<0$

これに $24\beta_2 - 27\beta_1 - 38$ が小さく、また理論的に正である事と付け加えておく。 $\beta_1 = 0$ なら興味ある特別の場合が起る。そのときに $m = 0$ で 曲線は水平の直線になる。これはまた区間Ⅱ型の場合でもある。

これに対する3次方程式の解はやや面倒である。⁽¹⁾ $m = -\frac{2(5\beta_2 - 4\beta_1 - 9)}{3\beta_1 - 2\beta_2 + 6}$ から求められる。

またこれは β_2 を含んでいるけれども理論的には3次方程式のと全く同じ値を与える。

判定規準は実際に得られた二つの結果が必ずしも等しくはないので、次を用いて3次方程式から m を得る方が望ましい。

$$m = \frac{24\beta_2 - \sqrt{24^2\beta_2^2 - 54\beta_1\{m'(4-\beta_1) + 7\beta_2 - 12\}}}{2\{m'(4-\beta_1) + 7\beta_2 - 12\}}$$

ここで m は上には与えられた β_1 と β_2 の式から求められるが何か他の経験的な方法によって得られる。

β_2 の値は判定規準から *Tables for Statisticians* の図から求められる。それは、Ⅷ型の曲線を用いたときに β_2 の特別の値の結果である。そしてこれは実際の統計量によって与えられる β_2 の代りに m の結果にこの理論値を用いる。

例。

度 数	補 整 (1)	補 整 (2)
469	437	436
186	222	207
166	165	161
134	136	141
122	120	127
112	109	115
1189	1189	1189

(1) 3次方程式の取扱い方等に関しては例えば、高木貞治：代数学講義 第一巻を参照。

平均値は186の数の中心よりあとへ距離0.65518ずれている。定数は次の通りである。

$$M_2 = 2786 \quad \beta_1 = 0.408$$

$$M_3 = 3275 \quad \beta_2 = 2.047$$

$$M_4 = 18.252 \quad \lambda = -0.05$$

$$5\beta_1 - 6\beta_2 - 7 < 0$$

したがって Ⅷ型が用いられる

$$m = 0.500$$

$$a = -5.777$$

$$y_0 = 102.6$$

曲線はオ一群の中央より前0.277から最後の群の端点よりあと0.022までに渡つた。この補整は上の(1)に示されている。

-rから-aまでの間にある面積は次の式で計算出来る。

$$\frac{y - r(a - r)}{1 - m}$$

特別の場合にはこの範囲は、養老保険の中の6ヶ月満期に関するデーターの株に6に固定する事が出来る。また平均値が、

$$\frac{a(1-m)}{2-m}$$

と言う事に注意して、次を得る。

$$m = 0.237 \quad y_0 = 111.1$$

面積の結果は補整(2)に与えられている。次表はこの場合での面積の計算法を示している。

曲線の方程式

$$y = 111.1 \left(1 - \frac{x}{2}\right)^{-2.237}$$

は0から6までの範囲を持つている。aが負なのは M_4 が正であるからである。

x (1)	$1 - \frac{x}{a}$ (2)	$\log(1)$ (3)	$\log(2)$ (4)	$\log(1) - \log(2)$ (5)	$\log \frac{y}{1-y}$ (6)	$\log \frac{y}{1-y}$ (7)	観測回数 (8)	$(19) \times (8)$ (9)	求められた 値 (10)
0	...	...	...	...	...	...	6	1189	118
1	0.8333	.0772	.0348	2.0804	2.3318	2.147	5	1070	127
2	0.6667	.1761	.0778	1.1230	2.3748	2.368	4	947	141
3	0.5000	.3010	.1323	2.1779	2.4479	2.687	3	811	161
4	0.3333	.4815	.2118	2.2574	2.5049	3.227	2	645	207
5	0.1667	.7781	.3419	2.3875	2.6387	4.255	1	436	256

幾つかの例は計算を簡便する事も出来る。それらの計算法ははっきりするために詳しく示されている。両方の調整は実際に可成り適合している。

極限の場合の例が次の統計量で分るだろう。

x	度 数	調 整 値	理 論 値
1	45	42	45
2	38	45	45
3	46	45	45
4	53	45	45
5	43	45	45
6	38	45	45
7	49	45	45
8	41	45	45
9	44	45	45
10	52	48	45
	449	450	450

平均値は 5 の級の中より数 0.57 である。モーメント等は次の通りである。

$$\begin{aligned} M_2 &= 8.374 & M_3 &= 124.46 & B_2 &= 1.78 \\ M_4 &= 0.026 & B_1 &= 0.011 \end{aligned}$$

ゆえに

$$y = \frac{449}{2\sqrt{3} \cdot 11.2} = 44.8$$

範囲は 0.57 から 10.43 である。

上の系列は *Carlisle*⁽¹⁾ $3\frac{1}{2}$ パーセントの最後の数値 10 級の合計によつて得られたものである。A の表と平均値は 0.57 の代りに理論値 0.5 であり y の値は 45 である。範囲は 0.5 から 10.5 までである。この例は極端な場合に *Pearson* 系曲線の補整の方法を示している。興味がある。補整値と理論値が示されている。補整値は小数以下が省略されている。

IX 型

$$y = y_0 \left(1 + \frac{x}{a}\right)^m$$

範囲は $x = -a$ から (但し $y = 0$) $x = 0$ (但し $y = y_0$) までである。

$$a = \pm 0.5(m+2) \cdot \sqrt{\frac{m+3}{m+1}}$$

m は次式の解から得られる。

$$m^3(B_2 - 4) + m^2(7B_2 - 12) + 24mB_1 - 16B_1 = 0$$

$$y = N(m+1)/a$$

$x = -a$ から平均値までの距離は $\frac{a(m+1)}{m+2}$ で $x = 0$ から平均値までの距離は $-a/(m+2)$ である。

平均値を原点に持つ型が望めば (VI 表参照)

$$A = \frac{(m+1)a}{m+2}$$

$$y_0 = \frac{N(m+1)^{m+1}}{a(m+2)^m}$$

VI 型のように、 m の値は 3 次方程式を 2 次式に簡単にすることが或は前に示された他の方法によつても得る事が出来る。

調整率は VII 型に対するものと同じ方程式の助けを求められ

(1) 英属 Cumberland 州にある都市 人口 48000 England に近い地方の主要都市である。

の4つの場合に分けられる。(1) I 型の特別の場合。(2) $\lambda = 0$ 。(3) $5B_2 - 6B_1 - 7 > 0$ 。(4) $2B_2 - 3B_1 - 6 < 0$ 。 $B_2 = 2.4$ $B_1 = 8$ $B_2 = 2.4$ $B_1 = 0$ 32 であれば曲線は、横いた直線

$$y = \frac{\sqrt{2}x}{30} \left(1 + \frac{x}{3\sqrt{2}0}\right)$$

になる。また $B_1 = 0$ なら極限として水平線になる。しかし $B_1 = 7$ $B_2 = 4$ ならば、I 型の他方極限であり指数級数 (X 型) とある事が分る。

例

期 間	保険死と年表で年金の危険にさらされた数	I 型	度 数 直 線
0	119	118	108
1	100	98	97
2	86	85	86
3	75	74	76
4	61	63	65
5	50	52	54
6	37	41	43
7	27	30	32
8	22	20	22
9	12	11	11
10	3	2	0
	594	594	594

危険にさらされた数を期間での概観であるか或は $x = 1/2$ から $x = 1$ の間の面積を表わしていると仮定すれば平均値は 2909 である。

$$\begin{aligned} M_1 &= 8.27 & B_1 &= 0.490 \\ M_2 &= 10.99 & B_2 &= 2.606 \\ M_0 &= 102.50 \end{aligned}$$

$$5B_2 - 6B_1 - 7 > 0$$

$$2B_2 - 3B_1 - 6 < 0$$

曲線は I 型とは余りちがっていないしまた B_1 が 0.32 で B_2 が 2.4 であれば、直線

$$y = 111.8 \left(1 - \frac{x}{10.43}\right)$$

さうな。その範囲は -0.6 から 10.0 まででこれから得られた補整値が表に示されている。

-0.6 から 0.5 までの全面積は期間 x に対する度数として取られている。I 型を用いる事によつて次の定数が計算される。

$$M = 1123 \quad a = 10913 \quad y_0 = 115.54$$

この曲線は -0.586 から 10.327 までの間に渡つてゐる。オ I の表にある 118 は -0.586 から +0.5 の間の面積として取られている。理論的には期間 -0.5 以前の危険にさらされる数はありえないが、我々は単に数字の系列に曲線を適合する例を与えているに過ぎないからこの事を気にする必要はない。概観の系列に曲線を適合する事により或は曲線の出発点を仮定する事によつてこの困難性に対処する事が出来る。

もし M が 1 より小さい事があれば曲線の型は幾分異つて来る。例えば $y = 100 \left(1 + \frac{x}{10}\right)^{1.25}$ ならば、我々は次の概観の値さうな。

$$100, 98, 95, 91, 88, 84, 79, 74, 67, 58, 0$$

是はれた死亡年表において実死亡数はこの形を取るだろう。しかし曲線の形は出発点で平坦性は少ないであろう。例えば

American Medico-Actuarial 死亡年表 1913 年で年齢 30 ~ 34 の級でさうである。

X 型

$$y = \frac{x}{0} e^{-\frac{x}{0}}$$

範囲は 0 から ∞ である。

平均値から原点までの距離は σ である。

平均値 (y_0) での誤差は $\sigma/\sqrt{n}$ である。

この曲線は $\gamma_2 = 0$ 即ち $\beta_2 = 4$ なる正型の分布の場合である。

正型に対する条件は $\beta_2 = 6 + 3\beta_4$ で与えられている。ゆえに
 偏度函数型は $\beta_4 = 4$ 、 $\beta_6 = 7$ で与えられる。また曲線は、正型と
 正型の極限ともなっている。

例

	度 数	補 整 値	理 論 値
1	4165	4132	4096
2	2028	2016	2048
3	782	1015	1024
4	480	511	512
5	266	257	256
6	132	130	128
7	71	65	64
8	36	33	32
9	17	17	16
10	9	8	8
11	2	4	2
12	1	2	1
13	1	1	1
14	1	1	1
15	1	1	1
	8192	8192	8192

修正前の平均値は 20097 である。

$$M_1 = 2.04 \quad \beta_1 = 4.629$$

$$M_2 = 6.290 \quad \beta_2 = 9.502$$

$$M_4 = 39.720 \quad \sigma = 1.43$$

曲線式が指数函数であるときにはモーメントと平均値は修正する必要がある。しかし高次の偏度があるときには Sheppard の補正は適当でない。若し曲線がオ一群の始めから出発するときは β_1 が正ならば平均値は近似的に $\frac{1}{\beta_1}$ だけ大きめに見積られ、また真平均値周りのオ二次モーメントは略 $\frac{1}{\beta_1^2}$ だけ大きめに見られると思う。この修正を用いると平均値は 1934、 β_1 は 2.123、 σ は 1.457 である。

上の統計量は鉄板の系列に属したもので理論値も付け加えられている。11 12... 等と書かれている系列のごとき統計量では 2 1 0 1 0 2... であつた。厳密に言えば我々は鉄板系を取扱つてゐる。我々はモーメントの修正を導くために面積の系列として計算を行つた。

曲線の面積補整の計算では α から β までの間の面積が $(y_\alpha - y_\beta)$ の値であるという事を憶えておけば有用である。どんな補整が理論的結果よりも実度数に近いかを注意する事は興味ある事である。

私はオ一群の始めから出発する事が分っているクリケット得点による次の系列をオ二の例として与えよう。

得 点	0-19	20~	40~	60~	80~	100~	120~	140~	160
系 列	64	34	18	9	6	3	3	1	0
補整値	64	34	18	10	5	3	1	1	1

前の項に対する各項の比は 0.54 である。また補整は殆ど正確である。しかしながら 120 の群の 3 のためにモーメントは理論値 $\beta_1 = 4$ 、 $\beta_2 = 7$ から可成り離れた判定基準を与える。もし我々が前例と曲線の出発点を仮定したとすれば我々は殆ど正確に理論的な結果を再現したのであろう。

XI 型

* 詳しくは附録 (I) をみよ

$$y = y_0 x^{-m}$$

この曲線の範囲は $x = b$ (但し $y = y_0 b^{-m}$) から $x = \infty$ (但し $y = 0$) までである。

次の値は

$$m^2(4 - \beta_1) + m^2(7\beta_1 - 12) - 24\beta_1 m + 16\beta_1 = 0$$

から求められて

$$b = \pm \sigma(m-2) \sqrt{\frac{m-3}{m-1}}$$

$$y_0 = \sqrt{b}^{m-1} (m-1)$$

原点から平均値までの距離は $\frac{b(m-1)}{m-2}$ である。

平均値を原点に持つ形に表わしたいならば (57 頁の折込み VII 表参照)

$$A = \frac{b(m-1)}{m-2}$$

$$y_e = \frac{\sqrt{b}}{b} \frac{(m-2)^m}{(m-1)^{m-1}}$$

Ⅷ型での様に m は 3 次方程式を 2 次に簡単化する事によるか或は前に簡単に述べた他の方法によつて得る事が出来る。 m は 5 から ∞ の任意の値を取り得るが 実際には 5 より小さくはない。曲線は $\beta_1 = 0$ のときに VI 型の特別の場合である。

判定規準は、次の 3 つの場合にわけることが出来る。① VI 型の特別の場合 ② $\lambda = 0$ ③ $2\beta_1 - 3\beta_2 - 6 > 0$ 。

期	間	散 回 の 数	XI 型による調整
0		1 6 5	1 8 3
1		6 5	5 3 7
2		2 3	3 2 6
3		3 2	2 0 6
4		1 3	1 2 4
5		8	7 4
6		1	4 7
7		6	3 1
8		3	1 7
9		3	1 2
10		1	0 8
11		3	1 6
		3 2 3	3 2 3

私はこの型で実際によく表わされた分析に出会わした事が無い。しかし私が与える契約撤回の系列に XI 型を応用する試みは失敗を与えるであろう。

定数等は

$$\beta_1 = 4.97$$

$$m = 29.69$$

$$b = 57.14$$

$$\log y_0 = 54.3563$$

原点から平均値までの距離は 57.205 である。

面積計算では a から ∞ までの面積として $y_0 a^{-(m-1)/(m-1)}$ を使う。

ねじれた J 型曲線

I 型の注意で (65 頁) 指摘したように m_1, m_2 が数値的に 1 より小さく且つ一方が負のときに興味ある曲線を得る。これは、

$$\beta_1 > 1.5 + 1.125\beta_2$$

と

$$\beta_2 < 2 + 1.25\beta_1$$

に帰る。と言うのは m の値の合計が 1 と -1 の間になければならない。或は γ が 3 と 1 の間にある事に注意する事によつて見る事が出来る。特別の場合は

$$y = y_0 \left(\frac{\sigma \{ \sqrt{(3+\beta_1)} + \sqrt{\beta_1} \} + x}{\sigma \{ \sqrt{(3+\beta_1)} - \sqrt{\beta_1} \} - x} \right)^{\sqrt{\beta_1/(3+\beta_1)}}$$

のときに推移型 (XII 型) として議論されている。

範囲は $x = \sigma (\sqrt{(3+\beta_1)} - \sqrt{\beta_1})$ から $x = -\sigma (\sqrt{(3+\beta_1)} + \sqrt{\beta_1})$

原点は平均値である。

$$y_0 = \frac{N}{b \Gamma(m+1) \Gamma(1-m)}$$

$$m = \sqrt{\frac{\beta_1}{3+\beta_1}}$$

$$b = 2\sigma \sqrt{3+\beta_1}$$

β_1 が正のときに平方根は負号を取る; $\beta_1 = 0$ のときの曲線の極限値は水平線である。

判定規準は $5\beta_1 - 6\beta_2 - 9 = 0$ である。

例

頻 度	補 整 値	概 率
...	2	185
33	31	31.6
...	...	40.6
53	49	49.3
...	...	57.0
65	65	65.2
...	...	73.2
81	83	82.4
...	...	92.3
101	103	103.3
...	...	114.6
131	134	132.9
...	...	155.5
181	171	186.0
...	...	244.2
350	342	415.4
1,000	1,000	...

平均値は131群の中心より数へ0.051の差にある。度数等は次の様である。

$$\mu_2 = 4.266 \quad \beta_1 = 0.761$$

$$\mu_3 = -7.688 \quad \beta_2 = 2.646$$

$$\mu_4 = 45.154$$

$$5\beta_2 - 6\beta_1 - 7 = -0.368$$

$$y = 872 \left\{ (5.808 + x) / (2.204 - x) \right\}^{0.05}$$

補整値の他に等距離概率の数値が与えられている。それらは曲

線が始めは急にそしてその後はやゝ急に緩みだすとも急に上る事を示している。並べられた表での概回数はいよいよこの形になる。(例えば日本の死亡率表 1910年、52才の女性)。或るときは人口母集団にこれと同じような形が起る事がある。

U 型 曲 線

この型はI型で β_1 と β_2 。(またはII型の β_1)が負のときに起る。統計量にこの曲線を適合するには幾多の困難がある。なぜならばモーメントの概数を修正する事がやりにくいからである。106頁の表のオム欄に与えられている数値は 曲線

$$y = 10 \left(1 + \frac{x}{2} \right)^{-7} \left(1 - \frac{x}{2} \right)^{-0.05}$$

の面積を計算する事によつて得られたものである。

U型曲線の極限値は範囲の終端で度数群を二つに分離してしまふ。この極限は

$$\beta_2 - \beta_1 - 1 = 0$$

のときに起る。

今我々が取扱つた曲線の幾つかはまれなものであり実際にこれらの曲線適合は避けられるものである。なぜならばそれらは β_1 と β_2 の離反値に依存し、またこれらの正確な値に到達するデマンスは極く少ないからである。換言すればもしある特別の場合に曲線適合の目的が与えられた実際の数値と補整された数値の間に完全な一致がえられたならばそのとき主要型(I, IV, VI)はすべて必要である。なぜならば他の型は推移型でありそして β_1 と β_2 の特別の値による他の型は生ずる必要がないからである。しかしながらもし我々の目的が広い意味で確率論の勉強にあるなら推移型も重要であり、それらは勿論 β_1 の値が判別現率によつて示されたものと少ししか違わないような場合に適当に用いられる。少ししか違わない程度とは(数で命するうちに) β_1 の標準誤差で示された限界内にあると言う事を意味する。

追加例

§1 今まではただ度数曲線の色々な型を説明する観念に立つての例を考えて来た。しかし実際計算で適用性の範囲を示しそれらを適用する際に生ずる二、三の困難な事を注意する機会を与えるに助となる一、二の実際例を考える事は賢明であろう。

実際計算で保険数理士が一般に取扱いを望む函数は危険にさらされているものの或は死亡数の系列或は契約脱落の系列等ではなくて、死亡数と危険にさらされたものの間の比である。即ち死亡率と共に新質、結婚、撤回の割合である。①②に保険数理士の研究している度数曲線はこれらの比率のいずれが我々の実験した曲線を補整出来るかどうかと言う問題に自づとなつて来る。しかしもしその補整が出来なければ、それは何か別の方法に置き換えねばならないだろうかと言う事になつて来る。さて考えられる事はこれらの比率が度数分布であるかどうかである。若しそうであれば度数曲線の使用は経験による事になる。死亡率は各年令で死する人間の比率を与えている。もし各整数年令で危険にさらされた10000人を考えれば死亡数は死亡率の10000倍であり、この事は実際の経験にはほとんど釣り合わないようなものではあるが分布として死亡率を考える事が出来ると言う事を示している様である。死亡率や罹病率を単なる度数曲線で表わす事は不可能である。地方結婚率は確に度数曲線に非常に似たものである。また脱落率は年令にしたがつて考えようが期間にしたがつて考えようがⅢ型の例の様な型となる。しかしながらそれが悪いのまゝにする必要があると言うものにおいて、度数分布の非常に誇張された考え方とは離れているけれども、比率に関する直観的な操作に対しては実際的な障害がある。すべての数の端での危険にさらされたものの数は小さくなる。独身者の死亡や結婚は非常に大きな比率であるが一方二、三の年令附近のものでは、未補整データに示された比率は力であるかも知れない。これは非常に取扱い難いものである。と言うのは実際の観察値が事実であるよりももつと応用に向

いて粗雑に比率を取扱う傾向があり、そして出来る限り避けた方がよい様な任意的なやり方を導きそれを用いる前に資料をつとめて粗分けする事が強いられるからである。勿論50や100と言う様な小さな数の死者でも各年令にしたがつて止むを得ず研分けておかなければならないと結論してはならない。しかしそこに2000或は3000とかであつても比率を扱う事のために導かれる粗雑さは結果に可成りの影響を及ぼすのである。補整率とは危険にさらされた数によつて指示された重みとはかけ離れている所の等しい重みを各死亡率に与えると言う事を意味する。

§2 これらの障害をはつきりさせるような場合を考えそれらに打ち勝つ実際的方法を扱う事は有用であろう。考えられるべき統計量は次の論文 *M. Mackenzie Lees "On Ratio of Mortality and Marriage among daughters of Peers and Heirs Apparent, etc."* (*Trans. Fac. Actu.* [276]) から取られている。そしてデータは128頁のようにまとめられている。モーメントは累加法で計算されている。平均値 μ は 77.77 周りのモーメントは次のように求められている。

$$\begin{aligned}\mu_1 &= 632092 & \beta_1 &= 1557153 \\ \mu_2 &= 627101 & \beta_2 &= 4731321 \\ \mu_3 &= 191033\end{aligned}$$

判定規率は $K = -1.5$ である。しかし私はモーメントの計算に於いて71年での結婚率 0.0057 を有略したので私はⅢ型を用いた。その年令での率の算入は曲線を延長し、そして判定規率の算出値は可成り大きくなる。かつ μ は近似的に $6 + 3\beta_1$ に近い。Ⅲ型に対する定数は

$$\begin{aligned}\gamma &= 0.201592 & \alpha &= 77.77187 \\ \rho &= 1.51381 & \text{モード} &= 23.51122\end{aligned}$$

①②に曲線は年令 14.02937 で出発する。

$$\gamma_0 = 370.05$$

この補整から求められているのは表に与えられている。しかし結婚率の分布は度数曲線に直接に結びつくと言う事を示す傾向があるがそれらは満足出来る補整を与えてはいない。この誤差は全くほとんどが上に述べた障害によるものである。若し我々が結婚率でとられた代数学的型を検定したとすれば我々は個々のものの非常に多くの数を取扱う事によつて資料のラフさを避けるような母集団資料にもとづいた計算によつて定められねばならない。しかしながら我々は補整法を探究しているのだからこれらの障害がどうして処理出来るかを知らねばならない。勿論それは補整法にある程度まで応用出来る方法でなければならぬ。

困難の原因は正しくないような重みがい用いられた項目が与えられていると言う事がはつきりしたし、もつとも明確な表示は実際の危険にさらされたものと結婚の数は別々に補整されなければならぬと言う事である。しかしこれは非常に多くの附加的な計算を必要としており、また危険にさらされたものにおける偏差と結婚における偏差は独立ではないと言う事実は見逃しやうい。

二重補整と今指摘した誤りの両方を避ける所の簡略法を用いる事が出来る。この方法は危険にさらされたものに類似の系列を用いてこの系列を結婚の新しい系列が計算出来る仮想的な危険にさらされたものとして取扱う事からなつてゐる。この方法の長所は我々は一つの補整を行うだけでよく、また幾の色々な部分の重みは近似的に与えられていると言う事である。この関係において次の事に注意して置こう。即ち危険にさらされたものは一般に度数曲線によつて表わす事が出来るので仮想的な危険にさらされたものは、そのような曲線(正規曲線)で仮定された簡単な型として取られる事を示すのが自然である。この正規曲線は便利でもある。なぜなら正規曲線に対する縦線は既に表に与えられているからである。

図3 仮想的な危険にさらされたものは試行もしくは危険にさらされたものの値からで決める事が出来る。124頁の表の E_x の値は

Tables for Statisticians の Sheppard の表から取られている。2が306, 3.084, 3.108, 3.132...等の如く取られており10⁴倍されたものが記入してある。 $M'_x = E'_x X$ が計算され補整されている。次の値は M'_x からの系列から得られたものである。

$$\text{平均値} = 24.85779 \quad B_1 = 140775$$

$$M_1 = 275006 \quad B_2 = 501114$$

$$M_2 = 190.112 \quad K = -7102$$

$$M_3 = 4361.12$$

Kが大きいので正規が用いられた。また

$$Y = 0.310350 \quad y_0 = 192.625$$

$$P = 1.841405 \quad \text{モード} = 21.63 \pm 62$$

$$R = 5933325$$

これから曲線が計算されており最良の補の補整された結婚率は M'_x と E'_x で割つて求められている。これらの値は非補整値によく一致している。

未婚婦人の結婚率

年齢 x	結婚の危険に さらされた 数 E_x	結婚の数 M_x	結婚率 M'_x	度数曲線に よつて補整さ れた結婚率	結婚の危険に さらされた仮 想的な数 E'_x	結婚の数 M_x	補整された 結婚率
15	3655	3	.0008	...	3695	3	...
16	3663	8	.0022	.0027	3633	7	.0019
17	35285	49	.0137	.0132	3187	44	.0157
18	32935	114	.0336	.0332	2757	77	.0350
19	3187	176	.0552	.0517	2742	151	.0541
20	2945	217	.0744	.0667	2541	187	.0695
21	26825	192	.0714	.0776	2354	168	.0807
22	2443	211	.0864	.0846	2179	188	.0880
23	2187	212	.0967	.0881	2016	196	.0917
24	1756	194	.0992	.0887	1864	185	.0920
25	1758	146	.0831	.0875	1723	143	.0901
26	15335	137	.0865	.0845	1571	135	.0861
27	1017	121	.0854	.0843	1469	126	.0812

(4477)

年令 x	船舶の危険 にさらされて いる数 E _x	船舶の数 M _x	船舶率 M _x	度数目線に よって補正さ れた船舶年 E _x	船舶の危険 にさらされた 船舶の数 E _x	船舶の数 M _x	補正された 船舶年 E _x
28	12075	105	0.926	0.953	1355	112	0.954
29	11085	70	0.653	0.673	1247	72	0.673
30	1068	60	0.562	0.640	1151	65	0.631
31	784	24	0.650	0.583	1061	67	0.567
32	7005	41	0.053	0.528	776	63	0.508
33	8485	30	0.354	0.075	897	32	0.452
34	802	39	0.086	0.425	825	40	0.400
35	752	20	0.266	0.378	758	20	0.352
36	711	25	0.352	0.335	696	25	0.307
37	6725	18	0.268	0.275	639	17	0.270
38	638	11	0.172	0.260	586	10	0.235
39	6125	14	0.227	0.228	537	12	0.205
40	5865	15	0.256	0.177	492	12	0.176
41	5685	7	0.158	0.173	451	7	0.151
42	5415	6	0.111	0.151	412	5	0.130
43	515	9	0.155	0.131	376	6	0.112
44	4915	2	0.041	0.113	305	1	0.096
45	476	5	0.105	0.077	315	3	0.082
46	454	5	0.110	0.084	283	3	0.070
47	4405	2	0.045	0.072	262	1	0.060
48	416	5	0.120	0.062	237	3	0.051
49	395	2	0.051	0.054	218	1	0.044
50	3785	...	...	0.026	197	...	0.037
51	3635	...	...	0.039	181	...	0.031
52	3485	1	0.027	0.034	165	...	0.026
53	3355	2	0.039	0.027	150	1	0.022
54	3175	...	...	0.024	137	...	0.019
55	304	...	...	0.020	124	...	0.016
56	291	3	0.103	0.018	...	1	...
57	2785	1	0.036	0.015	...	...	...
58	261	...	...	0.013	...	...	...
59	2485	...	...	0.011	...	...	...
60	2345	...	...	0.007	757	...	0.007
61	2175	1	0.046	0.007	...	...	...
62	2075	...	...	0.006	...	...	...
63	2015	...	...	0.005	...	...	...

年令 x	船舶の危険 にさらされ ている数 E _x	船舶の数 M _x	船舶率 M _x	度数目線に よって補正さ れた船舶年 E _x	船舶の危険 にさらされた 船舶の数 E _x	船舶の数 M _x	補正された 船舶年 E _x
64	171	...	...	0.004	...	...	...
65	177	...	...	0.004	457	...	0.003
66	1655	...	...	0.003	...	...	...
67	154	...	...	0.002	...	...	...
68	1475	...	...	0.002	...	...	...
69	1355	...	...	0.001	...	...	...
70	1205	...	...	0.001	272	...	0.001
71	1125	1	0.089	...	...	...	...
72	1055	...	...	...	...	...	...
73	95	...	...	...	...	...	...
74	805	...	...	...	...	...	...
75	77	...	...	...	...	...	...

表 4 (NM⁺) 表にこの方法を適用する数値例を次に与えよう

$O = 10$ と $52\frac{1}{2}$ 年を原年にもつ正規曲線が用いられている。その値は Crelle 表の助けによつて g_x 倍されている。

計算は次表の通りである。

$g \times E \times 10^5$	年令	Sheppard 表の補正-E	年令	$g \times E \times 10^5$
810	52	0.3984437	53	801
877	51	0.3744773	54	850
644	50	0.3866681	55	875

これらの記入値 ($g \times E \times 10^5$) を 5 年単位にまとめて私は次の表を得た。

年令	$g \times E \times 10^5$
20	13
25	70
30	218

(ツツ)

年 令	$\sum XEX 10^4$
35	594
40	1374
45	2410
50	3702
55	4517
60	4385
65	3402
70	2247
75	1177
80	461
85	133
90	31
95	5
100	1
	25,034

Crelle表の採用と群分けの省略が計算の節約のために用いられている。数値は例によつて必要であつたから充分正確なものである。それから次の値が求められた。

$$\text{平均年令} = 59439762$$

$$M_0 = 4.584327$$

$$M_1 = -0.09999871$$

$$M_2 = 61.17014$$

曲線の型 — I 型

$$M_3 = 32.81166$$

$$M_4 = 26.57123$$

$$a_1 = 18.78553$$

$$a_2 = 1521272$$

$$y_0 = 4607.884$$

$$\text{モード年令} = 59736787$$

(単位は5年単位である)

5年毎に概算を計算して、曲線が誤差の正規曲線から遠く掛離れていない事が分つたから他の年令^{*}でのこれらに対する概算の対数の才^{*}を同差で補挿した。

新平廣公式は面積を見出すために用いられている。また $\sum x$ は既に危険にさらされた項に用いられた概算的数値で割る事によつて見出される。期待された死亡数は次の様である。

間 数	年令群の中心に対して補正された $\sum x$	現 実 値	期 待 値	備 差	
				+	-
15~17	---	---	15	15	---
20	.00643	7	37	---	31
25	.00731	69	610	---	541
30	.00850	205	2006	---	1791
35	.00971	367	3807	11.7	---
40	.01177	588	5756	---	124
45	.01452	801	8114	14.4	---
50	.01866	1064	10638	---	22
55	.02405	1399	13366	---	124
60	.03516	1752	17732	21.2	---
65	.05118	2164	21367	---	273
70	.07682	2216	22612	45.2	---
75	.11648	1965	19258	---	392
80	.17462	1237	12419	27	---
85	.24870	494	5104	20.4	---
90	.33286	127	1266	---	30
95	.43289	18	173	---	27
100	---	1	15	25	---
		14,480	144,721	115.8	1037
				2195	

^{*} $\sum x = 45.2$ は正規曲線の方程式だから計数は $Ax^2 + Bx + C$ 。所定現率は曲線が正規に近いがどうかを示している。

85 保険統計士達にしばしば必要であるから考案された方法の特別の場合を試みるのも興味がある。

Makelamの仮説を $\text{Calog } P_x = A + BC^x$ によって定義して危険にさらされたものの数を表すために正規曲線 ($y_0 e^{-(x-A)^2/2\sigma^2}$) を取り、 $\text{Calog } P_x$ の値を挿入する。これは廣が次の様に表す事が出来る事の仮定を意味している。

$$\begin{aligned} y &= (A + BC^x) y_0 e^{-(x-A)^2/2\sigma^2} \\ &= Ay_0 e^{-(x-A)^2/2\sigma^2} + By_0 e^{-(x^2 - 2Ax + A^2 - 2\sigma^2 \log_e C)/2\sigma^2} \\ &= Ay_0 e^{-(x-A)^2/2\sigma^2} + HB y_0 e^{-(x^2 - 2[A + \sigma^2 \log_e C]x + (A + \sigma^2 \log_e C)^2)/2\sigma^2} \end{aligned}$$

$$\begin{aligned} \therefore H &= e^{(A^2 + 2\sigma^2 A \log_e C + \sigma^4 (\log_e C)^2 - A^2)/2\sigma^2} \\ &= e^{A \log_e C + \frac{\sigma^2}{2} (\log_e C)^2} \end{aligned}$$

$$\therefore y_0 = Ay_0 e^{-(x-A)^2/2\sigma^2} + HB y_0 e^{-(x-A)^2/2\sigma^2} \dots\dots\dots (I)$$

即ち、危険にさらされたものの曲線として同じ標準偏差を携へ同じ原典を持つ二つの正規曲線の和になっている。

二つの平均値の間の差は $\sigma^2 \log_e C$ で与えられるから

$$\log_e C = \frac{x-A}{\sigma^2} \log_e e$$

全解答は分っている原典 (年令 A) の周りのモーメントを求めらる事によって極く簡単に作る事が出来る。何とならば $\int_{-\infty}^{\infty} xy dx$ と $\int_{-\infty}^{\infty} x^2 y dx$ (最初の二つのモーメント) は

$$(x-A) M_1^* \text{ と } M_2^* \sigma^2 + M_2^* \{ \sigma^2 + (x-A)^2 \}^*$$

を与えるからである。

$$\therefore M_1 = Ay_0 \sigma \sqrt{2\pi}, \text{ また } M_2 = HB y_0 \sigma \sqrt{2\pi}$$

$M_1 + M_2$ (全人数) で今与えられた値を割れば我々は既知原典周りの一次モーメントとして $\frac{(x-A) M_2}{M_1 + M_2}$ を得る。二次モーメントとしては、

$$\frac{M_2 \sigma^2 + M_2 \sigma^2 + M_2 (x-A)^2}{M_1 + M_2} = \sigma^2 + (x-A) M_1^*$$

或は $x-A = \frac{M_1 - \sigma^2}{M_1^*}$ と $M_2 = \frac{M_1 (M_1 - \sigma^2)}{x-A}$ をうる。ここに M_1^* は周りのモーメントである。

上に示された様に

$$\log_e C = \frac{x-A}{\sigma^2} \log_e e \dots\dots\dots (II)$$

また一般に便利であるような $y_0 = \frac{10}{\sigma \sqrt{2\pi}}$ であれば、その時

$$A = M_1 / 10$$

$$\text{また } B = \frac{M_2}{10 H \sigma H}$$

$$= \frac{M_2}{10 H} \frac{1}{e^{A \log_e C + \frac{\sigma^2}{2} (\log_e C)^2}}$$

$$= \frac{M_2}{10 H C e^{\frac{x-A}{\sigma^2} \log_e C}} \quad (\text{方程式 (II) を代入})$$

$$= \frac{M_2}{10 H C^{\frac{x-A}{\sigma^2}}}$$

y_0 に用いた値に関しては注意が必要である。また A と B に関しても必然的に注意が必要になって来る。

若し *Tables for Statisticians* の *Sheppard* の表の縦線 (2) が 10^5 倍されて、危険にさらされたものの数として採られているならば計算から得られた A と B の値は $\frac{M_1}{10^5 \sigma}$ と $\frac{M_2}{10^5 H \sigma}$ である。

この理由は *Sheppard* の表が標準偏差を用いて作られている事によるものである。

* 正規曲線は対称的である事を想い起せば対称的な曲線の平均は周りの身数値のモーメントは零である。

* この式は二つの積分の和として考える事が出来る。即ち $(x-A)^2$ は (I) 式の最初の正規曲線の二次モーメントを与え $\{ \sigma^2 + (x-A)^2 \}$ は (I) 式の第二項の正規曲線の二次モーメントを与えている。

86 *Hardy* が英国保険会社の *ノーマン・シタ* の経験死亡率の補正に用いたように $\log_e C$ が既知であると仮定すれば我々は $\frac{(x-A) M_2}{M_1 + M_2}$ を与える唯一つのモーメントを計算するだけでよい。

これは方程式(II)を使って 完全に解く事が出来る。若しCが合計表で得られたいれば選はれた表に対してこの結果を用いてもよいはずである。

§7 ONM 表にある数値例は面白いものである。標準偏差1.0と原英 $\frac{5}{2}$ を持つ正規曲線が保られており表の各項は $\text{Colog } P_x$ 値とされている。これらの値は5年毎にグループされている。取柄の二つのモーメントは年令 $\frac{5}{2}$ 周りで計算されている。グループする場合にわずかの事であるが次の事を記述していなければならぬ。即ち 積 ($P_x \times$ 危険にさらされた数) に関する区間の中心は $x + \frac{5}{2}$ であるけれども ($\text{Colog } P_x \times$ 危険にさらされた数) の結果は x の軌線を与える。これは P_x を用いたときには20~24の5年間の中心英は22.5であり、また $\text{Colog } P_x$ を用いれば中心英は22である。

$$\text{数値は } M_1 + M_2 = 136387$$

$$5\text{年単位で } \frac{5}{2}\text{周りの1次モーメントは } = 1416184$$

$$5\text{年単位で } \frac{5}{2}\text{周りの2次モーメントは } = 41727354$$

2次モーメント*から $\frac{1}{2}$ だけの Sheppard の補正値を引き、1次モーメントと今修正された2次モーメントにそれぞれ $\frac{5}{2}$ を掛ける事によって5年毎の単位を一年毎の単位に直す事が出来る。

$$M_1' = 7080920$$

$$M_2' = 164384085$$

をうる。それから

$$\log(t-A) = 0.9586887$$

$$t-A = 9.092617$$

$$\log_{10} C = 0.03948373$$

$$A = 0.00301707$$

$$B = 0.000004518782$$

$$\log_{10} B = 5.6550214$$

P_x は A, B, C の値から得られた補整 $\text{Colog } P_x$ から計算されている。期待死亡数の表が出来上った。 P_x の値は、度数曲線の補整を示している表に与えられている。

年令群	年令群の中 心に対する 補整された P_x	期待死亡数	実死亡数	補	差
				+	-
25年以下	...	13.0	7	4.0	...
25-29	00812	470	57	...	2.0
30	00882	211.6	205	6.6	...
35	00991	3808	367	11.8	...
40	01162	565.9	588	...	21.1
45	01431	7997	801	...	1.3
50	01854	10575	1054	...	6.5
55	02517	13727	1377	...	6.3
60	03551	17702	1752	38.2	...
65	05160	21530	2164	...	11.0
70	07637	22493	2216	33.3	...
75	11415	18887	1945	...	76.3
80	17053	12136	1237	...	23.4
85	23352	519.1	474	25.1	...
90	36484	136.6	127	7.6	...
95	...	20.6	17	1.6	...
		14460.3	14450	128.2	147.7
				2761	

この結果は、故に G, F, Hardy 卿が最近与えたものと非常に良く似ている。しかし試みによつて C を求める事は避けられている。Hardy の期待値と実際の死亡数は上のよりも良い近似である。しかし私は死亡率はいくらか対称的であるとは思われない。即ち、年令75-79群は不一致を説明している。

全偏差は Hardy の偏差よりも少ない

* 我々は各群において面積ではなくて5つの観察の和を取扱つたので厳密に言えば Sheppard の修正は用いなくてよいと見做すことができる。この差は小さく、また誤差は再計算されなかつた。公式は、 $M_1 = V_1 - 0.02$ 、 $M_2 = V_2 - 0.02V_1 + 0.02742$ である。ここで M は修正されたものであり V は未修正のものである。

§ 8 生命保険や死亡年統計量に対して度数曲線の応用出来るもの他のものは極く最近に討論されている。危険にさらされたものや被保険者統計、或は各年令での保険金の合計額等々は度数曲線によつて何時でも補整する事が出来る。

保険数理士が保険会社の負債額を評価する時に実際には価格査定で仮定されている死亡年表にしたがつて相対的年の各年令で生存する人の比率の割合に即して評価する。

若し年令 x で与えられた死亡年表によつて x 年生き残る割合を ${}_x p_x$ とする。また E_x がこの表で言う年令 x での保険金の合計額であるならそのとき x 年後に生き残る保険の掛けられた合計額は $E_x \times {}_x p_x$ である。

種々の死亡年表に対して x の種々の値と度数曲線の可成り広い範囲を E_x によつて決定することが出来るために我々は x によつて $E_x \times {}_x p_x$ の分布の近似として再び度数曲線を取る事になる。教種の統計例は既に与えられている。* また本書に

与えた以外の例を調べたいと思う読者はこれらと参考にするか、*Barometric Heights* † に K. Pearson と A. Lee によつて与えられている非常に歴大な例を参照されたい。

* J. Inst. Act. L. V. I.

† Philos Trans. A. C. C. 423

§ 9 若し我々が曲線の範囲を知つてゐるときは I 型でさえ、次の多くのモーメントを求める必要がなくなる。なぜならば、曲線の出発点周りのモーメントを与える73頁の方程式が簡単な解

を与えるからである。

我々は次をうる。

$$M'_1 = \frac{b(m_1 + 1)}{m_1 + m_2 + 2} \quad \& \quad M'_2 = \frac{b^2(m_1 + 1)(m_1 + 2)}{(m_1 + m_2 + 2)(m_1 + m_2 + 3)}$$

$$Y_1 = \frac{M'_1}{b} \quad Y_2 = \frac{M'_2}{M'_1 b} \quad \& \text{ 同様に}$$

$$m_1 + 1 = \frac{Y_1(Y_1 - 1)}{Y_1 - Y_2}$$

$$m_2 + 1 = \frac{(Y_2 - 1)(1 - Y_2)}{Y_1 - Y_2}$$

なる。

ここに M'_1 は曲線の出発点周りのモーメントである。

§ 10 しかしながら若し一時的な考え方にしたがつて曲線の出発点だけを決める事が出来さえすれば3次のモーメントまでに依存する次の解が用いられる。

$$\lambda_1 = \frac{M'_2}{M'_1} \quad \lambda_2 = \frac{M'_3}{M'_1 M'_1}$$

とかけば、曲線の方程式の定数は次によつて与えられる。

$$m_1 + 1 = \frac{2(\lambda_2 - \lambda_1)}{2\lambda_2 - \lambda_1 - \lambda_1\lambda_2}$$

$$m_2 + 1 = \frac{2(\lambda_2 - \lambda_1)(\lambda_2 - 1)(1 - \lambda_2)}{(2\lambda_2 - \lambda_1 - \lambda_1\lambda_2)(1 + \lambda_2 - 2\lambda_2)}$$

$$b = \frac{M'_1 m_1 + m_2 + 2}{m_1 + 1}$$

$$a/b_1 = m/m_2$$

§ 11 § 9 の方法の例として I 型にもとつてみよう。そこで、曲線は $1/x$ なる出発点にして $1/x$ 単位の範囲を持つてゐる事は既定した。66頁の表で年令 x に対する行を考えると、最初の群を除いてしまえば、 $1/x$ と $1/(x+1)$ は S_1 と S_2 を与えるし、年令 $1/x$ 周りのモーメントは $1/x$ と $1/(x+1)$ となり年令 $1/x$ 周りに移すと $1/x$ と $1/(x+1)$ となる。同一群に対するモ

メント $0.034 \times \frac{1}{5} \times 24,269.36$ と $0.034 \times (\frac{1}{5})^2 \times 24,269.36$ をそれぞれ加える事により $M_1' = 408.18$ 、 $M_2' = 24,269.36$ となる。ゆえに

$$M_1 = 0.3498$$

$$Q_1 = 1735$$

$$M_2 = 2.7758$$

$$Q_2 = 13.765$$

$$y_0 = 154.2$$

$$\text{モードは } 17.5 + 1.735 \times 5 = 26.175$$

これらの値から、初めの4つの群に対する補正された数値はそれぞれ37、140、152、143である。

8.12 この方法は 8 頁の例 I に用いる事が出来る。修正前のモーメントを用いて

$$y = 1.87082x^{1.129685} (25.49727 - x)^{1.122275}$$

となる。これは102877を原典にもつていてそこが曲線の出発点である。*これはオ1項(308)はオ2項(200)の前に単位距離よりも少ない所にあるべきだと仮定している事を示している。

最初の二つの項の間の距離を0.75であると仮定し、初めの二つのモーメントを計算し範囲を25に固定して、また1.1を出発点にすると次の方程式を発見する。

$$y = 0.89055x^{1.1830} (25-x)^{1.12700}$$

この曲線による補正は次表に示されている。

期 間	契約脱退数	Type I の曲線による補正値
1	308	315
2	200	192
3	118	100
4	67	74

* オ3版の1025頁の補正はこの曲線によつては得られない。私は今与えられたと同じ仮定補正を用いたが、それは多様化する事が不可能である。

(ツヅキ)

期 間	契約脱退数	Type I の曲線による補正値
5	57	57
6	44	46
7	29	38
8	28	31
9	26	26
10	21	22
11	18	19
12	18	16
13	12	13
14	11	11
15	5	9
16	11	7
17	7	6
18	6	5
19	1	4
20	3	3
21	1	2
22	3	2
23	2	1
24	...	1
	1000	1000

8.13 オ一群の補正面積の計算は求平腐公式が応用出来ないで当面は難しいかも知れない。次の方法は正しい値となる限度の方とを与えている。

$$\int_0^x y_0 x^{x_1} (b-x)^{x_2} dx$$

$$= \int_0^x y_0' x^{m_1} (e^{m_2 x} m_2 e^{m_1 x} x + \frac{m_1(m_1-1)}{2} e^{m_2 x} x^2 - \dots) dx$$

$$= y_0' x^{m_1+1} e^{m_2 x} \left(\frac{1}{m_1+1} - \frac{m_2 x}{e^{m_2 x} (m_1+2)} + \dots \right)$$

これは x が小さいときには早く収束する級数である。上述の例で $x = 1.5 - 1.0$ ス 37 = 0.57 のときである。オヌ項はわずかに結果に影響する y_0' は次の公式で計算されるはならない。

$$\frac{N}{e^{m_1+m_2+1}} = \frac{\Gamma(1)}{\Gamma(m_1+1) \Gamma(m_2+1)}$$

式に正型とオノ群の面積を見付け出すための公式は次の通りである。

$$\int_0^x y_0' e^{-rx} x^p dx = y_0' x^{p+1} \left(\frac{1}{p+1} - \frac{rx}{p+2} + \dots \right)$$

こに

$$y_0' = N r^{p+1} / \Gamma(p+1)$$

第VI章 曲線系の比較

§1

前章では Pearson 系の度数曲線を取扱ったが度数分布を表わすのに他の方法も用いられている。Pearson の曲線系は偶然に関連した色々な問題について集められた事実を描写する事は既に知っている。システムと言うものは実際に起った分布に近い値を与えなければ意味がない。二項級数は生起した事象の回数を表わすのでこの観点から正しいことが証明されている。何故ならこの二項級数によつて与えられる数は試行によつて与えられる数に真に近似していることを経験から知っている。我々が問題を考えるときには大抵次の事を認めている筈である。即ち確率論の真の正当さは そのような理論の方法で起る事象は期待するような事象をみちびくと言う事であるしまたもし最近発表された度数曲線系を比較しようと思えば、それらがどのような分布のクラスを表わしているかをみたりまた失敗の事実や応用の困難性を注意深く考察したりすることよりむしろそれらの曲線等が導かれた方法を調べる事によつて比較されねばならない。

§2

二項級数は確率論の簡単な模型を實際に表わしている事を経験から知つて居るように、これから出発してこれを取扱うか、あるいは曲線系の一部としてその極限を取扱うのが普通である。實際上二項級数は拡張された他の一般型的の特別の場合でなければならぬ。我々はそれを採用する事が当然であるらしいとの仮定の上で曲線を設定するか、あるいは二項級数よりもつと複雑な級数(例えば超幾何級数)によつて曲線を作るかのいずれかの方向から進める事が出来る。そしてどちらの表示においても二項分布によるよりもより普遍性をもっているものに選しよう。然し展開されたどの公式のもつ最終的正しいさも大抵は偶然分布として合理的に表わす事の出来る統計量に対してその公式の応用の

広さにかかっていることは注意しなければならない。そのような応用性は公式を得るのに採用された基本的仮定の重要な標準である。何故なら、初期の状態におけるもつともらしさは実際の検定に用いられているかどうかを分析する曲線の単なる弁護にすぎない筈であるという事を認めねばならない。

我々が88頁で扱った良く知られている誤差の正規曲線は簡単な度数曲線を見付ける方向へのオースtepである。しかしこれは p と q が近似的に等しいか、あるいは q が充分大きいときには二項定理 $(p+q)^n$ の記述として具合がよいけれども他の場合には不完全である。保険統計の仕事においてこのような場合が屢々起る。或る保険会社で生命保険をかけられるある年に達した年齢での死亡率あるいは一年間に死ぬ人数の確率が小さくまた毎年50人以上の死者を与える度数分布即ち $q=0.02$ 、 $p=0.98$ というような場合は誤差の正規曲線によつて満足に表しえないであろう、これは正規曲線が大数の法則であると言う意味からも真実である。しかしもしその様な一つの根拠にかゝっている場合だけを取扱っていると実際の統計量における行動の大きな領域を持つことはできないであろう。またそれは級数が二項級数より超幾何級数によく似ているときの値を期待することは殆どない。

§ 3

この点が Person, Thiele, Charlier, Edgeworth, Bruns, Képteyn、その他の人々の研究に導かれた正規曲線の欠点である。またこれらの着者達によつて提案された曲線は凡ての数理統計学を勉強する学生にとって少なからぬ興味のあるものである。本章では実際に起る統計量にこれらの曲線の適合がどれ位離れるかを示すであろう。要するにどれ位離れているかを集められた事典から得られた概略数値を累進的に組合けりて分析される。

しかしながらその前にこの提出された型の簡単な訂論が必要である、即ち統計量は厳密に同質的資料から得られていることはめ

つたにないと言うこの種の実際の仕事での古い困難性を注意しよう。この事実は実地においては代表的問題の一つとみなさなければならぬ。またもしも系列が異質的の小さな量にもかかわらず累進的に組合けする事が出来れば、この系列は他の幾つかの観点からも保険数理工や統計家の手近にある多くの仕事において非常に貴重なものとなつて来る。

§ 4

さて我々がA型と呼んでいる式にもどつて考察しよう。すなわち、

$$F(x) = \phi_0(x) - \frac{1}{3!} \frac{\mu_3}{\sigma^3} \phi_3(x) + \frac{1}{4!} \left(\frac{\mu_4}{\sigma^4} - 3 \right) \phi_4(x) - \frac{1}{5!} \left(\frac{\mu_5}{\sigma^5} - \frac{10\mu_3}{\sigma^3} \right) \phi_5(x) + \dots$$

$$\text{即ち } \phi_0(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}, \quad \phi_n(x) = \sigma^n \frac{d^n}{dx^n} \phi_0(x)$$

故にもし $\sigma=1$ なら、すなわち標準偏差によつて測れば

$$\phi_3(x) = (3x - x^3) \phi_0(x)$$

$$\phi_4(x) = (x^4 - 6x^2 + 3) \phi_0(x)$$

$$\phi_5(x) = (-x^5 + 10x^3 - 15x) \phi_0(x)$$

これらの式の応用に於て x は標準偏差によつて平均からの離れを測つたものである。

即ちこの尺度は *Tables for Statisticians* に用いられているものである。上述のA型の方程式のまる括弧の中の係数は3次、4次、5次の半不変係数⁽¹⁾である事を注意しておこう。

(1) 平均値の周りのモーメント μ_n によつて決まるように表わされる。或る $K_1 = \mu_1$, $K_2 = \mu_2$, $K_3 = \mu_3$, $K_4 = \mu_4 - 3\mu_2^2$, $K_5 = \mu_5 - 10\mu_3\mu_2$, $K_6 = \mu_6 - 15\mu_4\mu_2 + 30\mu_3^2$ 等は度数分布の分布関数の特性関数の計数を Maclaurin 展開して得られる級数の各項の係数として定義されるものである。これは変量の原点の取り方に関して不変である。また Thiele の半不変数とも言われる。

Tables for Statisticians (Part II, Tables V-VII) では

$$Z_{(n+1)}(h) = \frac{(-1)^n}{\sqrt{(n+1)!}} \cdot \frac{d^n}{dh^n} \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{h^2}{2}} \right)$$

でありこれらの表を用いる際にFを次の様に書く事が出来る

$$\frac{N}{\sigma} \{ Z_1(h) + 0.81649658\sqrt{\beta_1} Z_2(h) + 0.45643546(\beta_2-3)Z_3(h) + \dots \}$$

この級数は多くの学者特にヨーロッパ大陸^{*}に於て議論つくされているものである。

またこれは正規曲線に用いられる母函数のようにも考えられている。それは当然正規曲線よりも適用性に広い範囲を持っているが極端に歪んだ場合にも役立つのではなく、そのような場合にはC.V.L. Charlier によつて提案されたB型を用うべきである。それは次の通りである。

$$F(x) = B_0 \psi(x) + B_1 \psi^I(x) + B_2 \psi^{II}(x) + \dots$$

$$\text{ここに } \psi(x) = e^{-m} \frac{\sin \pi x}{\pi} \left(\frac{1}{x} - \frac{m}{1!(x-1)} + \frac{m^2}{2!(x-2)} - \dots \right)$$

であり $\psi(x)^I = \psi(x) - \psi(x-1)$ 故に $\Delta \psi(x-1) \propto \psi(x)$ の値は $x < 0$ に対しては0と仮定されている。

全く同様に $\psi(x)^N = \Delta \psi^{N-1}(x-1)$

m が整数のとき極根においてこの $\psi(x)$ は $\frac{e^{-m} m^x}{x!}$ になる。この式は既にPoisson級数として確率論でよく知られている。

^{*} Gram, Thiele, Charlier, Bruns 等

Researches into the Theory of Probability (Meddelanden Lunds Astronomiska Observatorium, 1906) という表題のついた研究報告の中にC.V.L. Charlierは幾つかの数値例と多くの有益な注意を与えている。J.P. GramはOm Røkkeendringlinger, bestemt ved mindste Kvadraters Methode (Copenhagen 1879)の94頁にこの式は大分前にOppermann によつて示されたと言っている。

正規曲線は屢々大数の法則として呼ばれておりPoisson級数は小数の法則としてよばれている。A型で正規曲線を母函数に用いたと同じような方法でB型では $\frac{e^{-m} m^x}{x!}$ が母函数として取られている。

§5

B型を適合する際にどちらの方法が有効かと言うようなある特別な幾つかの困難を表わすが我々は序文でそれに関して指適したように $x=0$ から $x=\infty$ までの凡ての整数値を用いる事によつてモーメント法で

$e^{-m} m^x / x!$ を適合すれば

$$\mu_2 = m, \quad \mu_3 = m, \quad \mu_4 = 3m^2 + m$$

あるいは

$$\beta_1 = \beta_2 - 3 = \frac{1}{m}$$

をうる。

しかしながらこれらは距離単位は別として縦線系を仮定している。また実際の統計の仕事ではこれらの仮定は不当に我々を制限する事も分つてゐる。

けれども次のように書く事が出来る。

$$F(xW + C) = B_0 \psi(x) + B_1 \psi^I(x) + B_2 \psi^{II}(x) + \dots$$

これはWのためにグループ分けの単位が一般化されており、Cのために計算されたxからの点もまた一般化されている事を示している。

この型でCharlierは4つの適合方法を提出し次の事を注意した即ち級数は通常若し我々が $B_1 = B_2 = B_3 = 0$ のように定数をまとめれば一層よく収束する事を注意している。

① $W=1$ $C=0$ を仮定する これはもとの型に帰着する。 B_1 が0になるような m を選び $B_0 = N$ であるから我々は次をうる。即ち

$$2! B_2 = N(\mu_2 - 1)$$

$$3! B_3 = N(-\mu_3 + 3\mu_2 - 2b)$$

$$4! B_4 = N(\mu_4 - 6\mu_3 - 6b\mu_2 + 11\mu_2 + 3b^2 - 6b)$$

ここに b は原点から平均値までの距離である。この方法は m が b と余り大きく違わないと予想出来るときに用いられる。

② $W = 1$ を仮定し C は未知数として計算する。 B_1 と B_2 が 0 になるように m と C を選ぶ事によって

$$C = b - \mu_2 \quad 3! B_3 = N(\mu_2 - \mu_3)$$

$$m = \mu_2 \quad 4! B_4 = N(\mu_4 - 3\mu_2^2 - 6\mu_3 + 5\mu_2)$$

③ $B_1 = B_2 = B_3 = 0$ となるように m と W と C を見付ければ

$$W = \mu_3 / \mu_2 \quad B_0 = N/W$$

$$m = \mu_2^2 / \mu_3^2 \quad B_4 = \frac{N}{24W^5} \left(\mu_4 - 3\mu_2^2 - \frac{\mu_3^2}{\mu_2} \right)$$

$$C = b - \frac{\mu_2^2}{\mu_3}$$

この方法は普通 μ_3 が 0 のときには W は非常に小さな値を与え m は非常に大きな値を与える。それ故に著しくゆがんだ場合にだけ応用される。

④ C を任意に固定し $B_1 = B_2 = 0$ となるように W 、 m を見付ける

$$m = (b - C)^2 / \mu_2$$

$$W = \mu_2 / (b - C)$$

$$B_0 = N/W$$

$$W^3 3! B_3 = B_0 (W\mu_2 - \mu_3)$$

$$W^4 4! B_4 = B_0 (\mu_4 - 3\mu_2^2 + 5W^2\mu_2 - 6W\mu_3)$$

方程式の色々な組を導くための詳しい理算法は必要ないと思う $e^{-m} m^x / x!$ の表は *Tables for Statisticians* にある。

F. Y. Edgeworth* は A 型と同じような級数 即ち

$$e^{-\frac{K_1}{3!} \left(\frac{d}{dx}\right)^3 + \frac{K_2}{4!} \left(\frac{d}{dx}\right)^4 - \dots} \phi_0(x)$$

を用いている。ここに K_1 、 K_2 等は 3 次、4 次等の半不変系数である。

指数展開により次をうる。

$$\frac{N}{\sigma} \left\{ \tau_1(h) + 0.81649658\sqrt{\beta} \tau_4(h) + 0.98601330\beta \tau_7(h) + \dots + 0.45643546(\beta_2 - 3)\tau_5(h) + \dots \right\}$$

この級数と A 型級数との数値差は通常小さなものである。A 型級数には $K_1^2 \frac{1}{2!} \frac{1}{3!} \frac{1}{3!} \frac{d^6}{dx^6}$ から生ずる τ_7 項を含んでいない。それ以後の項も異なるがこの展開式は 4 次以上のモーメントを用いない事として 7 等の項は無視する事が仮定されている。

母函数として正規曲線の代りに例えば III 型のような他の式を使う事もできる。度数函数に対する必要條件はそれが負の度数を作らない事である。そしてこの点に関する問題を追求したい読者は Steffensen 教授の著書¹⁾を参照されたい。そこには初等理論から面白い計算までを与えている。

Edgeworth の級数や A 型級数の一般的定義とそれらの基礎理論に關して読者は既に発表されたこれらの問題に關する論文や、H. Grenier 教授の論文 "On the Compositions of Elemen-

* Edgeworth は彼の方程式はその特性と理論的根拠は一意的なものであると主張している。それは A 型から生ずる負の度数を避けることは理論的に不合理なものであるという事である。この点の誤りは数値例に引き出されるであろう。Trans. Camb. Phil. Soc. 1905 (Law of Error), J. Roy. Statist. Soc. 1906 (Generalised Law of Error).

¹⁾ J. F. Steffensen, Some recent researches in the Theory of Statistics and Actuarial Science (Cambridge University Press, 1930. Third Lecture)

tary Error". *Skandinavisk Aktuarietidskrift* 1928, の13頁と141頁の辺りを研究されたい。

然しながら上述の曲線や級数に為されている定義が論じ盡くされたと断言は出来ないが、これらの提議が最も一般的支持を受けられる方法を説明するのに使われるかも知れない。またその方法が示されるまでは数値例を与えるようなことはしないつもりである。しかし次の事は注意しておこう。既に示されたように補整は $y = e^{-\frac{1}{2}(f(x))^2}$ と書く事によつて為されるべきである。正規曲線を用いているこの方法は並進法 (method of Translation) と呼ばれており、この方法は最も一般的形では任意である。実際には $f(x)$ の形は規定されねばならない。ある特別の場合には研究*されているがその方法は實際的にも理論的にも欠陥を受けやすいようであるしそんな詳しくは討論されてはいない。

§ 8

数 値 例

例 1

(誤差の正規曲線では満足な補整のされない対称的曲線)

観 察 値	Ⅱ 型	A 型	Edgeworth	正規曲線
11	14	15	16	20
116	109	106	106	95
274	286	284	285	270
451	433	437	436	456
432	433	437	436	456
267	285	283	284	270
116	109	106	106	95
16	14	15	16	20

* 次の文献を参照せよ。

Edgeworth, J. Roy. Statist. Soc. Vol LXI; Kapteyn, *Skew Frequency Curves in Biology and Statistics* (Groningen, 1903); あるいは Bowley, F. Y. *Edgeworth's Contributions to Mathematical Statistics* 1928.

この場合には正規曲線以外の他の曲線は優れた補整を与えている。B型を使わなかったのはA型が不成功の時にだけB型を採用する事を *Charlier* が明らかにしていたからである。彼はA型を用いるべきとき或はB型を用いるべきときを示す統計的判定規準を与えてはいないし、そのような判定規準がいかにして展開出来るかを知る事も困難である。B型を用いねばならかつたときにA型での解は例えばI型が適用出来ないときにそれを用いたために生じたような方法で虚の量を等く事はない。A型と *Edgeworth* の補整を得る際に A_4 と K_2 を夫々含んでいる項を用いている。 A_4 はここでは $\phi_4(x)$ の係数である。その他の A_n についても同様である。 A_n は μ_n を含むが他の μ の式をも含む事に注意せよ。

例 Ⅱ

(極端に歪んだい分布)

観 察 値	Ⅱ 型	A 型	Edgeworth
3	4	5	4
20	17	22	17
38	42	47	42
63	59	60	59
51	53	50	53
29	33	27	32
21	15	13	15
4	5	4	6
0	1.4	1	2
1	0.4	---	1

どの場合にも3つのモーメントが使われている。観察値と *Edgeworth* の補整は *Edgeworth* の論文 ("The generalised law of error") から採られたものである。A型の補整が最も悪い。

例 III
(はつきりと歪んだ分布)

観 察 値	Pearson の I 型	A 型	B 型	Edgeworth
-----	-----	-2	-----	-----
-----	-----	1	-----	1
-----	-----	8	-----	9
-----	2	25	12	30
64	67	53	64	64
116	116	90	104	102
140	138	125	129	130
145	139	145	134	135
134	128	143	128	130
106	110	123	116	111
82	89	93	93	92
72	69	65	73	73
49	51	44	53	53
37	35	31	36	36
25	24	23	25	20
13	15	16	14	10
10	9	10	10	4
5	5	5	5	-----
2	2	2	2	-----
0.4	1	1	1	-----

I 型の数値は *Chances of Death* * から取ったものであり
 Edgeworth の数値は前掲論文から引用したものである。これら
 の補整の各々は 4 つのモーメントを使って得られるものである。
 明らかに I 型による補整が最もよく、次いで B 型がよい補整を与

* *Chances of Death* I. (London, 1897)

えている。*Charlier* はこのような場合は A 型を用いたろう。
 B 型を適合する際に級が 4 つの近似的適用の方法を与えている事
 のために多くの困難がある。これはもつと先で切り抜かれる障
 害であるが今は B 型は面倒なものである。これらの補整で注目し
 なければならない他の点は A 型に現われている頁の度数であり
 また Edgeworth の補整にある始めの 40 ケにはデーターの中に
 それらに対応する場合がない事である。
 しかしながら Edgeworth は曲線の主体にのみ目的を持っていて
 末端については余り多く考える事をしなかつた事は認めている。
 もし一方の端で 1000 ケ中から 40 ケの超過が現われ、また他の
 端が 20 だけ不足していれば、主体はひかえめに述べねばならな
 い事を註しも感ぜざるを得ない。

例 IV
(丁 型 曲 線)

観 察 値	Pearson 型	B 型
133	136.9	134.9
55	48.5	51.6
23	22.6	22.5
7	9.6	9.5
2	3.4	2.9
2	0.8	0.6

この B 型曲線は *Charlier* によつて "*Researches into the
 Theory of Probability*" に与えられている。B 型は僅かに
 良い補整を与えているが一致性は両方とも厳密である。この例は
 丁型曲線として決定的ではないが B 型がそれらをうまく補整する
 事を示している。6 つだけの群をもつ二の特別の例は右側に片寄
 った型のような曲線を持つており 3 つの量は良く一致しているよ
 うである。Edgeworth の曲線は不適當である。丁型でな

い曲線の数値への外見上の補正ではあるけれどもA型による例は他の処で支えられている。

例 V
(ほぼ対称的系列)

観 察 値	Pearson IV型	A 型	Edgeworth
10	6	4	3
13	16	14	10
41	49	46	34
115	135	126	110
326	321	306	298
675	653	637	662
1113	1108	1108	1164
1528	1535	1563	1603
1692	1712	1753	1747
1530	1522	1548	1510
1122	1074	1075	1024
611	604	589	571
255	274	256	263
86	102	92	104
26	32	29	37
8	8	7	12
2	2	2	2
1	1	1	1
1	---	---	---

これらの補正は同じ様な結果を与えており説明も不用である。

例 VI
(二つの最大点をもつ分布)

資 料	Pearson II型	A 型	Edgeworth
10	3	26	4
78	96	74	34
193	191	156	135
286	261	262	270
303	304	354	363
291	319	390	390
303	304	354	363
286	261	262	270
193	191	156	135
78	96	74	34
10	3	26	4

これは二つのこぶを持つ分布の仮想例であり A型で $A_3 = 0$ $A_4 = 0.09$ と置いて作られたものである。そのときの級数は -4, -19, -53, -76, +103, +783, +1929, +2855, 等である。負度数は意味を持たないから捨てられており、これらのデータは最後の数字を四捨五入して切りつめてある。このデータが作られたA型の補正が不十分な一致を与えている点は興味ある特徴である。これは負度数のためでありまたモーメントに対する積分が $-\infty$ から $+\infty$ に渡っているためである。負度数はそれ自身にいくらか意義がある。この例に示されているような大きな範囲に渡つての曲線のあてはめに影響するときにはより一層意義がある。

例 VII

我々はB型の解を決する際に困難がある事を注目して来た。しかし他の公式に比較したこの補正力は *Tables for Statisticians*

and から $\frac{e^{-x} x^k}{k!}$ の形を取る他の二三の例を述べることに
 つて示す事が出来る。比較のために Pearson III型の例を付け加
 えた。どちらの型も他のものに厳密な一致を与えざる意味だと考
 えてはいないけれども、これらは極く稀には応用の範囲について
 一つの考え方を与えるために取られたものであると考える事は出
 来る。 $\phi(x)$ をきむ項を導く事によってB型に関する知識の範囲を
 広げる事が出来。また Pearson 系の凡ての曲線を用いること
 によって我々は Pearson IIIの範囲より広い範囲を覆う事が出来る。

B 型			Pearson III 型		
I	II	III	I	II	III
368	111	45	387	63	31
368	244	140	386	279	149
184	268	217	160	285	230
61	197	224	47	189	218
15	108	173	15	102	160
3	48	107	4	49	101
1	18	55	1	21	56
-----	6	25		9	29
-----	2	10		3	14
-----	-----	3		-----	6
-----	-----	1		-----	3
-----	-----	-----		-----	1

§ 9

上述の二三の例は我々が既に取扱つて来た曲線の型の比較を考
 える助けになるだろう。

Pearson 型の曲線は今まで使つて来たどの例にも充分な補整を
 与えている。しかし例VIで扱つた二重こぶのある想像例のデータ

一を再現する事は出来なかつた。Pearson 型の曲線は対称的分布
 僅に歪んだ分布や非常に歪んだ分布或はJ型、U型分布を補整出
 来る。それらは広い事柄に適合されておりまた一致性の点からも
 満足されている。その中に含まれている算術的計算は大変なもの
 であるがこの曲線型は上のような分布に最も有益なものである。

A型は数値的には最も少ない仕事ですむが非常に歪んだ分布や
 J型U型分布には満足な補整を与えない。そのために受入れ具合
 も小さい。しかし、もし Charlier が提案したようにB型と一
 緒に使えばJ型や歪んだ分布をも補整出来る。我々はB型を応用
 する際に二三の困難がある事を知っている。と言うのはCharlier
 は特別の場合に従うべき4つの適合方法のいずれかの決定に際し
 て多くの助けを与えてはいないからである。またこの型の補整力
 は現在の考察でその正しさを認める我々の試みよりも大きいのか
 も知れない事に気がつく。例IVにあるB型の補整に改良を加える
 事は明らかに不可能である。しかし例IIIとCharlierによつて
Researches into the Theory of Probability の中に
 与えられた二つの例はより一層不運である。

Edgeworthの曲線は大雑に言つてA型と同じような分布を
 補整する事が出来る。

§ 10

次に既に夫々について述べられたEdgeworthの曲線とA型に
 関する二つの障害に言及しよう。例VIIでEdgeworthの補整にお
 いて1000のうち40以上もがそれに対応する観察値を持つてお
 らずまたそれが大きすぎるらしい事を注意した。観察値の正確な
 数の再現は実際に必要な丈でなくそれはモーメント法でも仮定さ
 れている。それ故に若し沢山の数値が観察値の外に生じたら全度
 数が再現されなかつたかまたは度数が置き誤られたかに違ひない。
 どちらの場合においても主体はもとのデーターに示された以下の
 量は人工的に修正されねばならない。僅かに歪んだ分布の度数は

は申分なく再現されしそのような資料の多くの調整は使われている。しかしその方法は上述の困難に打ち勝つ方法が分るまでは一般化公式として十分な考究は殆んど出来ない。

A型に関する障害は対称性の少ないものの補正に現われる負度数を処理する事が大部分である。若し負度数が起つたら正度数は誇張されているだろうか？ そのような負度数の防禦は系列の多くの項を適当に直す事である。

しかし大きな変動におち入り易く且信頼出来ない高次のモーメントからえられた定数について多くの議論をする事と正当さを知る事は困難である。ある曲線がそれ自身近似的にさえも再現出来ない事や例Ⅱの結果が案外つまらぬものになった事は不満足でもある。しかしそれは異質的資料に同係した様な場合を考えるためには恐らく充分であろうし、それ故にニツ以上の曲線^{*}を重ね合せて表した方がより適当なものであろう。若しA型或はEdgeworthの曲線とそれらのモーメントが $-\infty$ から $+\infty$ までの代りに $-a$ から b まで積分されていれば、これらの困難はある程度までさけられるだろう。しかしそう言う事が出来なければ余り異常でない分布に適用の範囲を限定する事が必要であらう。 $-a$ から $+\infty$ まで適合させる近似的方法は既に与えられてはいるが、その結果はPearson III型のように全く優れたものではない。

§ 11

読者が補正のための系列にある大きな試みを行えば、次々の項の係数は収束しなないかも知れない級数を意味しているような事を往々見出すであらう。この事は前節に述べた困難に密接に関係した事である。

* もし $-\infty$ から $+\infty$ まで普通の積分が行われていればA型とEdgeworth曲線のもう二つのコブの再生が統計学的に可能であるかどうかは疑わしい。

なほならば、公式の係数から求められた二次と四次のモーメントの相対的値はともありえないように考えられるからである。

※ E.C. Rhodes, J. Roy. Statist. Soc. 1925, PP. 576 以下

§ 1

背の高い男は低い男よりも長い脚を持っていると言つたり、独身者が年をとればとる程結婚したり、子供を持つたりしそうにないと言つたり、晩婚の男の妻は早婚の男の妻よりも年をとっている等と言う。或は例を生命保険業務からとれば、養老保険が未満期々向に従つてグループされているときには、満期時の平均年齢は未満期々向につれて増加する。これらの陳述はどこにある因果関係があるか或は男の背丈と脚の長さの間に、或は夫の年齢と妻の年齢の間に、或は養老保険の満期年齢と未満期々向の間に相関があると言う事実を違った言葉で表わしたものである。然しながらこの陳述は一般的言い方である。即ちこれらは或る関係が他の関係よりもより密接なものであるかどうかを決定する助けにならないし、またそれは相関の如何なる尺度をも与えない。統計的仕事での目的は測定尺度の発見にある。すなわち我々は確率測定に対する尺度を持つたと同じように相関測定の尺度を求めたい。

そこにどんな相関もないなら我々のスケールは尺度0になるべきであり、確率性が1の確率で示されたように、その関係が出来るだけ密接なものであるときには相関1と呼ぶのがよからうと思いつくであらう。然しながら確率と相関の間の類似性のくずれる点が一つある。即ち負の確率のようなものはなかつたが負の相関を考えられる事は容易に分るであらう。何とならば我々は夫と妻の年齢の如く一緒に大きくなる二つの事柄AとBを考える事も出来るし、或は独身者の年齢と彼が結婚して生まれた子供の数の如く一方が増加するにつれて他方は減少するような二つの事柄CとDを考える事も出来るからである。

§ 2

この諸論をも一緒に考えて次の言葉で相関の定義を設けよう。測る事の出来る二つの特性AとBは次の時に相関があると云う。

すなわちAの異なる値Xには等しく関連しやうなるの同じ値を見付ける事は出来ない。換言すればその或る値は他の値よりそのXの値に従つてより起り易い。若しそうでなければ相関は存在しない。或は特別の例をとると20才 30才、60才或は何才でも、結婚している人々が常に40才の婦人と結婚していればそこにはもはや相関はない。他方もし凡ての人々が正順に彼より9才年下の婦人と結婚していれば相関は完全である。

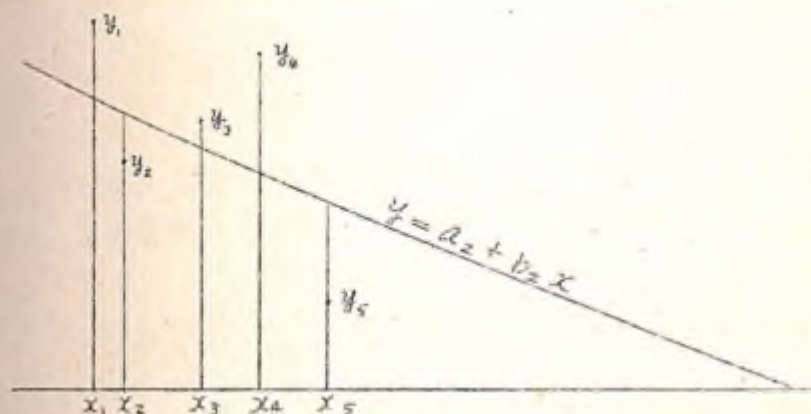
最老保 険の未 満期々 間 (5期分の 平均の 中心値)	満 期 時 の 中 心 年 令										合 計	行に對 する平 均満期 年 令	
	30	35	40	45	50	55	60	65	70	75			
2	2 24	20	16	2 12	26 8	6 4	14 0	6 4				56	53.75
7	1 18	1 15	2 12	6 9	62 6	36 3	40 0	22 3	2 6			172	55.03
12		2 10	9 8	17 6	117 4	99 2	127 0	52 2	8 4	1 6		432	55.85
17	3 6	5 6	4 4	24 3	145 2	155 1	237 0	84 1	11 2	3		665	56.59
22	0 1	0 0	0 0	3 0	133 0	167 0	271 0	73 0	20 0	1 0		674	57.58
27				9 3	90 2	123 1	231 0	71 1	11 2	3 3		538	57.88
32				1 6	11 4	49 2	127 0	49 2	8 4	2 6		247	59.94
37						6 3	49 0	22 3				77	61.04
42						2 4	2 0	3 4	8	1 12		8	62.05
47							0	1 5				1	65.00
合計	6	4	17	62	584	643	1098	388	60	8		2870	
列に對 する平均 満期々間	10.3	13.2	13.2	16.1	17.2	20.1	21.9	21.7	21.5	22.6			

(注) 細字の説明は510をみよ

行や列は array と呼ばれる。行に関連した変量の中央値はその型と呼ばれる。だから3列(即ち見出し40と書かれている)は型40のy array と呼ばれ4行は型17のx array と呼ばれる。ときには型と言う語を除く事もある。

53

この問題の統計的見地は未満期々間に従つてグループした養老保険加入2870例を享受する上の二重記入の表に示されている。表をみるとそこに相関がある事が分る。何故なら各行の平均満期々間を与えている列の数値は53.75から65.00まで絶えず増加している。一方期間はずから47に増加する事に注目する。同じように満期時年令が30から75に増加するにつれて平均未満期々間は10.3から22.6まで増加する。二つの系列は156頁の図に示されている。今相関がなかつたと考へれば列の平均値は他の函数に独立である筈であるから各列の平均値は同じになる筈である。平均値を図上にプロットすれば水平線になる筈である。この事は恐らく相関は平均値を通る直線の勾配によつて測られても良い事を示している。そして相関表に直線($y = a_2 + b_2 x$)を適合し結果から推量出来るものが何であるかを見る事によつてこの考へを追求出来るのである。

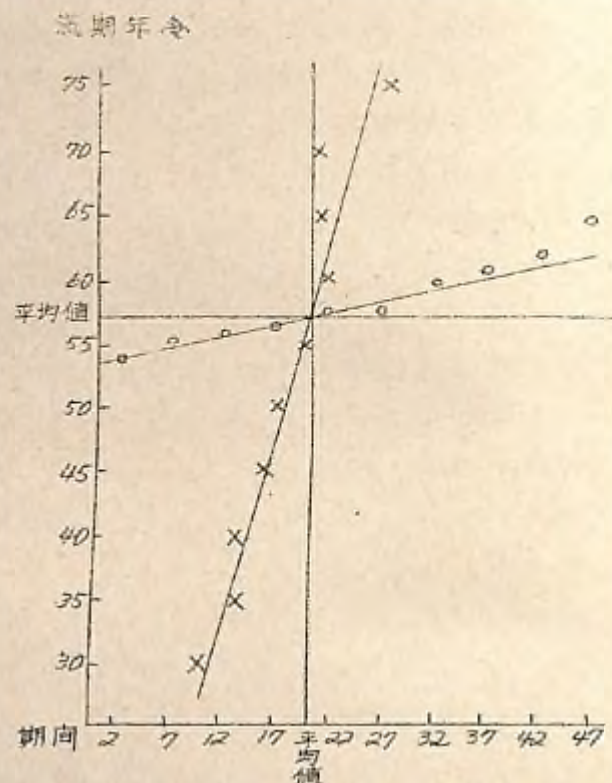


54

枚々が度数分布に曲線を適合するときには普通モーメント法を用いた。そこで次に証明してあるのと同じような理論を採り用いている。

$x_1 y_1, x_2 y_2, x_3 y_3, \dots$ 等は関連した偏差とおく。

$y = a_2 + b_2 x$ は補整に用いられる直線とおく。従つてそのとき x_i に対応する補整値は $a_2 + b_2 x_i$ である。



次にモーメント法を用いての度数曲線の適合の所ぞなりたように進めて行けば我々は補整面積、平均値と未補整面積、平均値は等しいとおくか、或は次のように置く。

$$(a_2 + b_2 x_1) + (a_2 + b_2 x_2) + \dots = y_1 + y_2 + \dots$$

或は

$$Na_2 + b_2 S'(x) = S'(y)$$

また

$$(a_2 + b_2 x_1)x_1 + (a_2 + b_2 x_2)x_2 + \dots = y_1 x_1 + y_2 x_2 + \dots$$

或は

$$a_2 S'(x) + b_2 S'(x^2) = S'(xy)$$

ここで $S'(x)$ は x の凡ての和であるから x の一次モーメントを与

え $S'(y)$ は y のすべての和であるから y の一次モーメントを与える、 $S'(x^2)$ は x の二次モーメントを与え $S'(xy)$ はそのモーメントの中で任意の度数が x 方向と y 方向に距離倍されているモーメントである。

もしこれらのモーメントを度数曲線の適合のときにやつたように平均値の周りに変換すれば、次をうる。

$$Na_2 = 0 \quad \text{或は} \quad a_2 = 0$$

そして

$$b_2 S(x^2) = S(xy) \quad \text{或は} \quad b_2 = \frac{S(xy)}{S(x^2)}$$

しかし全度数 (N) の二次モーメントは $N\sigma_1^2$ であるから

$$b_2 = \frac{S(xy)}{N\sigma_1^2}$$

また

$$\bar{y} = \frac{S(xy)}{N\sigma_1^2} x$$

である。

もし $S(xy) = N\sigma_1\sigma_2 r$ と書けば

$$\left. \begin{aligned} \bar{y} &= r \frac{\sigma_2}{\sigma_1} x \\ \bar{x} &= r \frac{\sigma_1}{\sigma_2} y \end{aligned} \right\}$$

ここで r は x と y の間の相関の統計的尺度 (相関係数) を表わすものである。オ2の方程式も全く同じようにして展開出来る。

35

上に与えた x と y の間の関係を示している二つの方程式はちよつとみた所一致していないようにみえる。然し初めの式 $\bar{y} = r \frac{\sigma_2}{\sigma_1} x$ は x の特別な値に対応する y の平均値 (y の上に棒を書いて表わされている) を与え、一方オ2の方程式は y の特定の値に対応す

え 例えはノフ頁の表IIで度数29は特定の値 (-4) 倍されている。これは我々が29の場合の各々に距離 (-4) を与え (-4) を29に加へ上げる事と同じである。

る X の平均値を与える事は注目しなければならない。例として簡単な場合をとるために $\sigma_1 = \sigma_2 = 1$, $r = 0.1$ を仮定する。そのとき、もし $X = 0$ なら対応する Y の平均値は0である。また $X = 20$ なら Y の平均値は2である。しかしながらこの事情を一変したときに $Y = 2$ に対応する X の平均値は20であると主張する事は勿論出来ずそれは0.2である。

§ 6

この前置きをした后で前の二つの方程式にもどつて、 r はどうして相関の尺度を示すのか またそれはいつでも満足な尺度として取扱う事が出来るのかどうかを考えよう。方程式 $\bar{y} = r \frac{\sigma_2}{\sigma_1} X$ を $\frac{\bar{y}}{\sigma_2} = r \frac{X}{\sigma_1}$ 或は $\bar{Y} = rX$ の型に書き直してみると r が相関の尺度である事が分る。また平均値が原点であり（この事は検算では平均値に対するモーメントに帰着させるべきである）測定単位が各々の場合に標準偏差である場合に一つの特性値は他の特性値で表わすことができるに解釈してもよい。この型に於て次の事が分る。即ち一つの特性（ X ）の増加につれて他の特性の対応する系列の平均値（ $\bar{Y}$ ）は Y の値に依存する範囲に従つて増加する。従つてももし r が負なら $\bar{Y}$ は減少する。 $r = 1$ のときに限り X と $\bar{Y}$ の増加は等しくなり 絶対相関になつて来る。もし $\bar{Y}$ が X の増加につれても定数であれば本章の始めでの定義で述べたようにそこに相関はない。またこの場合に r は方程式 $\bar{Y} = rX$ の形から容易に分るように0である。

我々は相関を測るスケールとして-1 から+1 迄に渡るべきであることを期待して来た。然し負の大きな値は二つの特性が一緒に変るばかりでなく一方が減少するにつれて他方は増加する事を意味していると強調しておこう。

r の数値は対応する二つの特性の変動の広がりを示す。この表示はもし平均値を156頁に示したような図にプロットしたときに

ほぼ直線になる（即ち回帰*は直線である）ならば充分なものである。直線性からの偏差の判断は想像される程易しいものではないが、それらがどの場合にも余り目立たなければ r は相関の満足な尺度で終る事が出来よう。

§ 7

ここでしばしば出会う別の困難を取除くためにこの概念を利用しよう。ある疑問を持つ学生には次の質問によつて示すのが最も良からう。即ちある事柄がいつも他よりも小さいときに完全な相関があるかどうかと云う質問。例として人間の右手の長さ（ X ）と左手の長さ（ Y ）の間の相関を考えよう。相関係数は殆ど1である。そして各特性はそれ自身平均から測られており、また、標準偏差によつて測られているから、もし凡ての左手が右手より向インナが短いか、あるいは右手に対して $\frac{99}{100}$ と云う、長さに固定された比率を持てば相関係数は減少しないであろう。

§ 8

次に数値計算の討論が必要となつて来る。§4の最後の諸公式をみかえせば二つの標準偏差と $S(XY)$ の値が必要な事が分る。既に標準偏差はどうして求められるかは知つており、また平均値周りのモーメントが必要である。けれども分数をさけたり、后の結果を修正するために、はある任意の固定点周りのモーメントを計算するのが実際的には最も良い方法であることをモーメントの計算の時に討論された事を思い出されたい。

勿論 σ_1 と σ_2 の値は64頁の諸公式すなわち $V_2 = V_2' - d^2$ を使つて求められる。二次モーメントから $\frac{1}{12}$ を引く事は度数曲線の適合のときと同じ場合に同じ理由のために行なわれねばならない。

* 回帰と云う語は F. Galton によつて遺伝の研究に際して用いられた。それは特別の親の子供は普通の母集団平均にもどつていく傾向を示すものである。

乗積モーメントに關して次をうる。

$$S(x'y') = S(x+d_1)(y+d_2) \\ = S(xy) + d_1 S(y) + d_2 S(x) + Nd_1 d_2$$

或は $S(x) = S(y) = 0$ だから

$$S(xy) = S(x'y') - Nd_1 d_2$$

ここに $S(x'y')$ は x の平均値から距離 d_1 と y の平均値から距離 d_2 の点の周りに計算されたものである。

§ 9

さて、154頁の例題に進む事が出来る。それは証明をやりそして上に与えた方法をより簡単につかむために役立つ事が分るでしょう。

モーメントが計算された点即ち満期時年齢60と未満期々間22に対応する群の中点が先づ固定されている。そして今の計算はこの点の周りになされている。次の表は平均値の計算と y -array の合計(すなわち表の下の方の合計)の2次モーメントの計算を示している。何故なら列は y -array 而行は x -array であるからである。

度 数	x'	度数 $\times x'$	度数 $\times (x')^2$
6	-6	36	216
4	-5	20	100
17	-4	68	272
62	-3	186	558
584	-2	1,168	2,336
643	-1	643	643
1098	0	-2,121	----
388	1	388	388
60	2	120	240
8	3	24	72
2,870 = N		+ 532 - 1,589	4825

$$d_1 = \frac{-1,589}{2,870} = -0.55366$$

故に 平均年齢 = $60 - 2.7683 = 57.2317$ である。
何故ならグループした単位が5年であるからである。

$$\sigma_1^2 = \frac{4825}{2870} - d_1^2 - \frac{1}{12} \text{ (Sheppard の修正)}$$

$$= 1.37465 - 0.083$$

$$= 1.29132$$

$$\therefore \sigma_1 = 1.13637$$

同じように行を取扱えば次表が作れる。

度 数	y'	度数 $\times y'$	度数 $\times (y')^2$
56	-4	224	896
172	-3	516	1,548
432	-2	864	1,728
665	-1	665	665
674	0	-2,269	----
538	1	538	538
247	2	494	988
77	3	231	693
8	4	32	128
1	5	5	25
2,870		+ 1,300	7,209
= N		- 969	

$$d_2 = \frac{-969}{2,870} = -0.33763$$

$$\therefore \text{平均未満期々間} = 22 - 1.68815 = 20.31185$$

$$\sigma_2^2 = \frac{7,209}{2,870} - d_2^2 - \frac{1}{12}$$

$$= 2.31453$$

$$\sigma_2 = 1.52135$$

§ 10

$S(xy)$ の値は相関表で度数の下に非常に小さな活字の数値を使つて計算される。例えば 50 の列の度数 62 は任意原点から横方向に 2 つと上方方向へ 3 つの間隔だけ離れている。だから $x'y'$ の値はこの間隔を掛ける事によつて求められそれは小さな活字で示されているように $2 \times 3 = 6$ である。他の数値も同じような方法で得られるが符号には注意しなければならない。表の左上の部分或は右下の部分のどの値も正である。何故なら度数は同符号の x と y の積が掛けられている筈だからである。しかし他の部分からのどの値も負である。何故なら異符号の x と y の積が度数にかけられているからである。

乗積モーメントの計算は次のようである。

度 数	$x'y'$	度数の合計(f)	$f \times x'y'$
155 + 71 - 84 - 123	1	+19	+19
145 + 99 + 11 + 49 - 11 - 52 - 49 - 90	2	102	204
24 + 36 + 3 + 22 - 22 - 6 - 9	3	48	144
6 + 6 + 8 + 3 - 6 - 8 - 11 - 2 + 117	4	113	452
1	5	1	5
3 + 17 + 62 + 2 - 1 - 1 - 2	6	80	480
9 + 26	8	35	280
6	9	6	54
2	10	2	20
2 + 2 + 1	12	5	60
1	15	1	15
1	18	1	18
2	24	2	48
			1,799

$$S(xy) = S(x'y') - Nd_1d_2$$

$$= 1,799 - Nd_1d_2$$

$$= 1262.51$$

$$r = \frac{S(xy)}{Nd_1d_2} = \frac{1262.51}{2870 \times 1.13637 \times 1.52135}$$

$$= 0.254$$

養老保険の満期時の年令と未満期々間の間の相関係数は 0.254 である。

この関係を表すもう一つの方程式によれば

$$\bar{x} = r \frac{\sigma_1}{\sigma_2} y$$

$$= 0.190 y$$

ここに凡ての測定値は平均値からなされていて単位は 5 年である。図の中に引かれた直線はこの結果を与えている。

§ 11

モーメント計算のために III 章 § 9 で与えた累加法と同じような方法で相関表に附して用いても便利である。

同じ例をとれば我々は与えられた表からその各列を相続して加える事により y -合計が求められ、この表を横に加える事によつてオ三番目の同じ形のもう一つの表を得る。

相関表のY-合計の表

養老保険 の未満 期年数	満期での中心年数										合計
	30	35	40	45	50	55	60	65	70	75	
2	6	4	17	62	584	643	1098	388	60	8	2870
7	4	4	17	60	558	637	1084	382	60	8	2814
12	3	3	15	54	496	601	1044	360	58	8	2642
17	3	1	6	37	379	502	917	308	50	7	2210
22	0	1	0	13	234	347	680	224	39	7	1545
27	0	0	0	10	101	180	409	146	19	6	871
32	0	0	0	1	11	57	178	75	8	3	333
37	0	0	0	0	0	8	51	26	0	1	86
42	0	0	0	0	0	2	2	4	0	1	9
47	0	0	0	0	0	0	0	1	0	0	1
合計	16	13	55	237	2363	2977	5463	1914	294	49	13381

上表のX-合計の表、此はXYグループと相関表に關して凡ての場合を与えている表である。

養老保 險の未 満期年	満期での中心年数										合計
	30	35	40	45	50	55	60	65	70	75	
2	2870	2864	2860	2843	2781	2197	1554	456	68	8	18501
7	2814	2810	2806	2789	2729	2171	1534	450	68	8	18179
12	2642	2639	2636	2621	2567	2071	1470	426	66	8	17146
17	2210	2207	2206	2200	2163	1784	1282	365	57	7	14481
22	1545	1545	1544	1544	1531	1297	950	270	46	7	10279
27	871	871	871	871	861	760	580	171	25	6	5887
32	333	333	333	333	332	321	264	86	11	3	2349
37	86	86	86	86	86	86	78	27	1	1	623
42	9	9	9	9	9	9	7	5	1	1	68
47	1	1	1	1	1	1	1	1	0	0	8
合計	13381	13365	13352	13297	13060	10697	7720	2257	343	49	87521

上側の表の右側の列にある合計は相関表の右側の列の合計のオ-次合計である、これは下側の表でX=30のランの和と同じものである。

Yの和の合計、あるいはXY表のオ-ランの合計はYの平均値 ($\frac{13381}{2870}$) を与えている、また同じようにオ-行の合計はXの平均値 ($\frac{18501}{2870}$) を与えている。

下側の表の合計はXYモーメント (87521) を与えている。その標準偏差は第一行から作られる系列 18501, 15631, 12767, 9907, 7064, 4283, 2086, 532, 76, 8 とその合計 70855 から求める事が出来る。それから平均値周りの2次モーメントが求められる。数値計算は次のようにして行なえる。

$$X \text{ の平均値} = \frac{18501}{2870} = 6.4463$$

$$V_2 = 2S_2 - d(1+d)$$

$$= \frac{2 \times 70855}{2870} - 6.4463 \times 7.4463$$

$$= 1.3747$$

Yのモーメントも同じようにして求められる。

$$Y \text{ の平均値} = \frac{13381}{2870} = 4.6624$$

$$V_2 = \frac{2(13381 + 10511 + 7697 + 5055 + 2845 + 1300 + 429 + 96 + 10 + 1)}{2870}$$

$$- 4.6624 \times 5.6624 = 2.2312$$

$$XY \text{ モーメント} = \frac{87521}{2870} - 6.4463 \times 4.6624 = 0.4399$$

$V_2 - \frac{1}{12} = 0^2$ ($-\frac{1}{12}$ は Sheppard の修正) と上の計算で平均値は $X=25$, $Y=-3$ 年のグループの中心から測られている事に注意すれば、今求められた値は直接方法で得られたこれらの分つて

いる値に一致している事が分るでしょう。XとYのモーメント (04399) は $1262.5/2870$ すなわち $S(XY)/N$ と同じものである。

§12

我々が相関係数の測定に用いた方法は行、列の平均値は夫々直線上に在ると仮定していた事は(§6で)注意している。その結果この関係が保たれているかどうかを試すための表を作らねばならない。前節で与えた方法の利点は各列、各行の平均値が非常に簡単に求められることであつた。

- ① 未満期々内のグループ間の長さは5年である事。
- ② 満期時の中心年齢間の長さは5年である事。
- ③ 任意原点は満期時の中心年齢 = 25 と未満期々間 = -3 の表す点である事。

を思い出せば各列の平均値は4-合計の表の各々の合計をとりそれに5を乗じ全数で割つて3を引けば得られる、すなわち中心年齢50年の列に対しては

$$\frac{2363}{584} \times 5 - 3 = 17.23$$

行の平均値はその次の表の右にある合計の差から求められる、すなわち未満期々間2に対しては

$$(18501 - 18179) \times 5/56 + 25 = 53.75$$

また未満期々間17に対しては

$$(14481 - 10279) \times 5/465 + 25 = 56.59$$

§13

相関係数を求める算術的なやり方には更にオ三番の方法がある。しかもこれは簡単であり平均値の系列の一つに依存するので非常に讃称されている。

計算は次のように行う。

任意原点の列(未満期々間22)までの列の(全度数の)一次モーメント	任意原点の行(年齢60)から列までの距離	(1) × (2)
(1)	(2)	(3)
-14	-6	+84
-7	-5	+35
-30	-4	+120
-73	-3	+219
-557	-2	+1114
-238	-1	+238
-----	0	-----
-26	1	-26
-6	2	-12
+9	3	+27
		1799

$$r = \left(\frac{(3)の計}{N} - d_1 d_2 \right) / (\sigma_1 \sigma_2)$$

$$= 0.254 \quad (\text{これは§10のものと同じである})$$

単位は終始5年である。(1)ランの計算をするための簡単な方法は168頁の表に示されている。

60年のラン或は22期の行はそれらが0倍されているから記入の要はないが完全性と読者が独りで行う算術上のチェックに使うために時には計算される事もある。

読者がこの一覧表の任意の項例えは見出し40の列の18を考えればそれは154頁の表での9を(-2)倍したものである事に気がつく筈である。

またとりわけ上表の(1)ランにとつたときにはやはり(-4)倍する必要がある。故に9を(-2) × (-4) 即ち8倍したらよい事が分る。この8は154頁の表に9の下に小さな活字で書いてある。

満期 年令	30	35	40	45	50	55	65	70	75
任意原点 からの 距離	負の積								
-4	8	---	---	8	104	24	24	---	---
-3	3	3	6	18	186	108	66	6	---
-2	---	4	18	34	234	198	104	16	2
-1	3	---	6	24	145	155	84	11	---
負の 合計	14	7	30	84	669	485	278	33	2
	正の積								
+1	---	---	---	9	90	123	71	11	3
+2	---	---	---	2	22	98	98	16	4
+3	---	---	---	---	---	18	66	---	---
+4	---	---	---	---	---	8	12	---	4
+5	---	---	---	---	---	---	5	---	---
正の 合計	---	---	---	11	112	247	252	27	11
(1)列の 数値	-14	-7	-30	-73	-557	-238	-26	-6	+9

他の例題や方法を取扱う前に特別の例についてその使い方を指
 導する方がよからう。方程式型の結果は各未満期々間に対応する
 平均年令を与える。次に Lidstone の Σ^* があるいは仮の年金を
 使って各々の記入事項に重みを付けそれから各々の場合の方程式
 を計算して新しい平均値の系列が作られる。評価に用いたこの結
 果は3つの方法の相対的正確さを与えるであろう。私は Σ 重み付
 * 私が使ったこの方法は近似的でありまた改良される可能性もある。
 すなわち、この結果は研究のための可能性の一つの指針として与えた
 にすぎない。

けの公式 (HM 表) を作っている。また

$$\text{満期時の年令} = 57.595 + 0.1200 \times (\text{未満期々間})$$

と言う関係も発表している。

この結果は評価する時、平均年令に関して概略のチェックとし
 ても用いられるしまた我々が今まで使っていた方法の助けを借り
 て養老保険の簡単な "モデル事務" を作るために何らかの役に立
 つ可能性も充分ある。

第 VII 章 理論 分布
擬 相 関

§ 1

前章では当然相関測定のためのスケールが必要なことを知った。また統計量に直線を適合する事*によって相関表を簡単にすれば、そのときその直線の傾きが相関測定の尺度として取られても良い事を示した。

しかしこの事は証明に適當のようではあるが決定的なものではない。むしろ直線の傾き自身よりも直線の傾きのある函数を用いた方がもつとよいかも知れない。且、直線はわれわれが単純化に用いるための最良のものではない事も経験から知っている。それ故に我々はこれらの疑念が取除かれるかどうかを考へねばならないし、それを遂行するための方法は一般的考察から相関の量の推定出来る表を作る事により理論的立場から相関を考える事である。

§ 2

色々の相関表を算出する事は出来るが、ここでは 10 ケの貨幣を投げそのうち 8 ケはテーブルの上に残し他の 2 ケを再び投げるような場合から始めよう。そのとき我々は 10 ケのうち 8 ケはどの組にも共通な投げの組をもつ事になる。何回も実験を繰返しそれで相関表を作れば 10 ケのうち 8 ケを固定しているので相関は 0.8 か或は何か 8 のある函数の比率によって測られるべきだと期待してもよい。同様にして若し 5 ケの貨幣を残せば係数は 0.5 でなければならぬ。理論的に作られた表は 179 頁へ 183 頁に示されているようなものになる筈である。これらの表は対称的である。

二つの標準偏差は同じであり、また平均値(最後の列と下の行をみよ)は直線上を走りまた水平線とこの直線が作る角の正接に、我々はどの相関表にも二本の直線を得る、即ち一つは列平均に対応するものであり、もう一つは行平均に対応するものとする。

よって判定出来る直線の傾きは 0.8 か 0.5 か或は 0.2 である。

§ 3

我々はどの表が 8 ケの貨幣を残した時の数を与える例によって作られた表であるかを説明しよう。

これらの場合のうち、最初の投げで 10 ケが“表”である場合を考へる。そのとき残された 8 ケの貨幣は“表”でなければならぬ。他の 2 ツはもう一度投げられるが次の比率で起る筈である。

2 ツが“表” 1 回

1 ツが“表” 1 ツが“裏” 2 回

2 ツが“裏” 1 回

従つて残した 8 ケと一緒にして、初の投げで 10 ケとも“表”を出しているので 2 回目の投げで 10 ケの“表”が出るのが 1 回、9 ケの“表”が出るのが 2 回、8 の“表”が出るのが 1 回の 4 つの場合を結論する。さて次に最初の投げで 9 ケが“表”で 1 ケが“裏”である場合を考へよう。そのとき我々は 8 ケとも“表”か 7 ケの“表”と 1 ケが“裏”であるどちらかを残す事になる。その仕方の数を我々は 9C_8 または 9C_7 とする事が出来て、夫々 9 と 36 である。あるいは割合としてだけ言うなら 1:4 である。再び投げる二つの貨幣は 1 H H : 2 H T : 1 T T の割合で投げられる。従つて我々はそれから第 2 列を作る事が出来る。読者は各列の合計は 2 項式 $(\frac{1}{2} + \frac{1}{2})^{10}$ の各項の係数である事に気がつくでしよう。

相関係数は第 VII 章の方法で計算出来て 0.8, 0.5, 0.2 各る値を与える。例えば“0.8 の表”の $S(xy)$ は 8192 であり

$\sigma_1 = \sigma_2 = \sqrt{2.5}$ である事が分る。故に

$$r = \frac{S(xy)}{N\sigma_1\sigma_2} = \frac{8192}{4096 \times 2.5} = 0.8$$

§ 4

さて次に相関係数或は相関の尺度として0.8を用うべきかそれとも0.5の函数を使うべきかを決めるのにこれらの表が使えるかどうかを考察しよう。そのため易しい実験は二つの表が同じ合計の分布を与えるように0.8の表を増した言0.8の表と0.2の表を加える事である。そうして得られた結果行の平均値(列の平均値)はそうして合成された表を作ったこれら二つの表に示されている平均値の中間にあるだろう。これらの平均値は0.5表の平均値に等しいけれども場合の数の分布は異っている。これは明白である。すなわち0.8, 0.5, 0.2は相関の特殊な係数であると仮定出来るし、またこれらの数値の函数についての検測する必要がないからである。もしこの結果を一般化するならば、

$$f(x_1) + f(x_2) = f\left(\frac{x_1 + x_2}{2}\right)$$

のような r の函数の発見が欲せられると言つてもよい。またこれは $f(x) = r$ と書いても満される。

§ 5

しかしながら我々が選んだ例は分布が対称的=項型に依存したものであり、それは他の場合にも容易に説明されないものである事に注意しよう。しかし我々はサイコロで同じような表を作る事が出来る。そこではサイコロの6つの面のうち2つを"表"とし他の4つの面を"裏"として考える。そのときは $(\frac{1}{3} + \frac{2}{3})^{10}$ 型の分布を得る。列(行)の平均値は直線上にある。また0.8と0.2の相関を与える等しい大きさの表を一層に加える事によって0.5表の平均値をうる。 $r=0.8$ の表は、0.8と0.2表に与えられている。明らかに我々は二項級数のような度数分布に対応し二重記入の表を取扱ったに過ぎないし、また実際に起る凡ての分布がそのような簡単なものであると期待する事は出来ない。凡ての場合に平均値は直線に従うと仮定してはならないし、また分布がこれらの対

論からよほど相異しているならば直線の傾きは相関の正しい尺度を与えると言う事も出来ない。しかし相関表の大部分はすでに示した型に近似的に一致する。

§ 6

読者は唯今説明された考察で我々が取扱ったのは二項級数に類似の点の系列であつて度数曲線に類似の面の系列でなかつた事に注意しなければならない。前章で取扱った正規曲線は或る条件下では二項級数の極限であり、また正規曲線に対応する度数曲面^{*}は次の通りである。

$$Z = \frac{N}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} e^{-\frac{1}{2}\left\{\frac{x^2}{\sigma_1^2(1-r^2)} - \frac{2xyr}{\sigma_1\sigma_2(1-r^2)} + \frac{y^2}{\sigma_2^2(1-r^2)}\right\}}$$

さて $r=0$ ならこの式は二つの正規曲線の積に書き換えられる。また $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \Sigma xy dx dy$ が求められれば既にV章で考えられたように $N\sigma_1\sigma_2$ あるいは $r = \frac{(\Sigma xy) - \Sigma x \Sigma y}{N\sigma_1\sigma_2}$ となる。

§ 7

正規曲面は特別な注意を引くいくつかの性質をもっている。その型の配列の分布を考えれば次のようである。

$$Z = Z_0 e^{-\{g_1 x^2 + 2h_1 x + g_2 x^2\}}$$

ここに g_1, h_1, g_2 は σ_1, σ_2, r の長い式に替つて書かれたものである。

指数を完全平方の形に作る

$$\begin{aligned} Z &= Z_0 e^{-g_1 \left\{x - \frac{h_1}{g_1}\right\}^2} e^{-g_2 x^2 + \frac{h_2^2 x^2}{g_1}} \\ &= Z_0' e^{-g_1 \left\{x - \frac{h_1}{g_1}\right\}^2} \end{aligned}$$

をうる。これは全曲面の標準偏差と同じ標準偏差を持っている正規分布である。がその平均値は h_1/g_1 だけ全曲面の平均値から異

* 附録 III をみよ。

っている。それは次のためである。

- ① 各配列の平均値の偏差は直接型に比例する。換言すれば配列の平均値は算術級数で増減し且直線上にある。
- ② 凡ての平行配列の標準偏差は等しく且それらの型に独立である。

これらの結論のうち前者に關する限り我々は貨幣投げに見られまた前章で仮定したと同じような性質を持つてゐる。その他の性質は貨幣投げの表にはみられなかつた。この事から正規曲面は實際の使用に堪えられない程の能力がないものだとして結論してはならない。おそらく類似の正規曲線が度数曲線の場合におけるよりもより広い範囲をもつてゐる。読者が次の点に注意するならこの曲面を具体化する助けになるかも知れない。

- (1) X軸、あるいはY軸に平行な垂直切断面は正規曲線である。
- (2) 等高線切断面は楕円である。またこれらの楕円が $z=0$ の面に正射影されるならば、それらは共心的であり相似で且つ相似の位置にある。

その外觀は広い平面の上にある孤立した小山の観がある。これは、はつきりと、丘に近づくにつれて非常にゆっくり上昇し、それから丘陵の両面は頂上近くでは急ではなく全く平坦になつて来る。丘陵は南東あるいは北面からみるとときには最も狭く、また北東、あるいは南西からみれば最も広くなる。

§ 8

さて相関に關する統計的工作を用心しなければならぬ障害を討論しよう。その障害は統計的資料の分類系列でそこに相関がないときに相関を示し、或はそれがあるときに誇張される危険である。我々はこのみせかけの相関（擬相関）の導入される二つの理由を考える事にする。先づ貨幣投げの表から考えよう。等しい割合の二つの表 0.2 表と 0.8 表を加える事によつて 0.5 の相関を与える表を得る事を前にみた。しかし $r=0.5$ の一方の平均値が

れている二つの表を一緒にすれば何が起るかをみよう。これは二人で同じ対称物を記録するとき一人はいつもその結果がある定数だけ誇張されている事以外は正確に測られているかどうかというような実験的仕事の面に起る。混合された結果と作られた表は、183頁の表と同じようなものであろう。相関係数は 0.78 と計算される。

§ 9

さてこの種の事が起つた場合を見つける方法を考えよう。多くの行の平均値は特別の方法で言い表わされる。それらは 0.5 表のように始められて完成されるが、混合出来るときには直線の方角は曲線と一緒に二つの終りの部分を纏いだしたようなものである。更に全合計は二項型や他の一つの度数曲線も構成しない。特別の場合にはこの二つの点は充分警戒しなければならぬ。しかし實際にそれらの応用は困難である。何故なら経験の結果が比較的小さな数に依存しているからでありまた回帰線の真の型は分らず曲線が合計で作られるからである。二人だけの代りに多くの観測者があつてもよい。これらの観測者は端点を曲線に変える事も出来るし、また平べつたい S のような回帰線を与える事も出来る。そのような場合のほんとの救済策は一緒にグループした多くの経験表がそれらの平均値とそれらの分布に關して同じものである事をみればよい。そして混合の正当性が証明された時にだけ混合数値を用いる事である。

§ 10

擬相関が導入されるかも知れないもう一つの原因は指数を使うために起る。

例として保険業務の大きな会社の帳簿にある制限払込み金額の養老保険を参照しよう。そして初期の保証期間 (x_1) と将来保険金を支払うべき数 (x_2) と実施中の保険証券の年数 (x_3) を考へる。

もこの比 x_2/x_1 , x_3/x_1 を依り相関係数を計算しても将来に保険金を支払う数と実査中の証券の年数との間の相関の尺度は求められない。何故かと言へば各々に同じ分母の分数を用いている結果誇張相関になる筈である。一即ち擬相関を導く。

今注意した結果はそれの特別の場合であるが擬相関の一般的定理は次のようである。

I. 平均値、標準偏差 と二つの絶対的測定値の間の相関係数によつて指数の平均値をみつける事

ある四つの互に相関する対象の絶対的大きさを x_1, x_2, x_3, x_4 とおき m_1, m_2, m_3, m_4 をそれらの値の平均値とし、 $\sigma_1, \sigma_2, \sigma_3, \sigma_4$ を標準偏差とおく。また $r_{12}, r_{23}, r_{34}, r_{41}, r_{24}, r_{13}$ を相関係数 $\epsilon_1, \epsilon_2, \epsilon_3, \epsilon_4$ は平均値からの偏差即ち $x_i = m_i + \epsilon_i$ ($i = 1, 2, 3, 4$) とおく。また i_{13} は指数 x_1/x_3 の平均値、 i_{24} は x_2/x_4 の平均値、とおく、 Σ_1 と Σ_2 は夫々 x_1/x_3 x_2/x_4 の標準偏差である。 N はグループの全数である。

それらを作っているものの平均値からの偏差は非常に小さいのでその三乗は無視出来ると仮定する。そうすると

$$\begin{aligned} i_{13} &= \frac{1}{N} S\left(\frac{x_1}{x_3}\right) = \frac{1}{N} \frac{m_1}{m_3} \cdot S\left\{\left(1 + \frac{\epsilon_1}{m_1}\right)\left(1 + \frac{\epsilon_3}{m_3}\right)^{-1}\right\} \\ &= \frac{1}{N} \frac{m_1}{m_3} \left\{ N + \frac{S(\epsilon_1)}{m_1} - \frac{S(\epsilon_3)}{m_3} - \frac{S(\epsilon_1 \epsilon_3)}{m_1 m_3} + \frac{S(\epsilon_3^2)}{m_3^2} \right\} \end{aligned}$$

よか $S(\epsilon_1) = S(\epsilon_3) = 0$ また $S(\epsilon_1 \epsilon_3) = N \sigma_1 \sigma_3 r_{13}$

$$S(\epsilon_3^2) = N \sigma_3^2$$

であるから

$$i_{13} = \frac{m_1}{m_3} \left(1 + \frac{\sigma_3^2}{m_3^2} - \frac{\sigma_1}{m_1} \cdot \frac{\sigma_3}{m_3} - r_{13} \right)$$

また

$$i_{24} = \frac{m_2}{m_4} \left(1 + \frac{\sigma_4^2}{m_4^2} - \frac{\sigma_2}{m_2} \cdot \frac{\sigma_4}{m_4} - r_{24} \right)$$

II 標準偏差と二つの絶対測定値の間の相関係数とによつて指数の標準偏差をみつける事

$$\begin{aligned} N \Sigma_{13}^2 &= S\left(\frac{x_1}{x_3} - i_{13}\right)^2 \\ &= \frac{m_1^2}{m_3^2} S\left\{\left(1 + \frac{\epsilon_1}{m_1}\right)\left(1 + \frac{\epsilon_3}{m_3}\right)^{-1} - \left(1 + \frac{\sigma_3^2}{m_3^2} - \frac{\sigma_1}{m_1} \cdot \frac{\sigma_3}{m_3} - r_{13}\right)\right\}^2 \\ &= i_{13}^2 S\left\{\frac{\epsilon_1}{m_1} - \frac{\epsilon_3}{m_3} + \text{二乗の項}\right\}^2 \\ &= i_{13}^2 \left(N \frac{\sigma_1^2}{m_1^2} + N \frac{\sigma_3^2}{m_3^2} - 2N \frac{\sigma_1}{m_1} \cdot \frac{\sigma_3}{m_3} r_{13} \right) \end{aligned}$$

あるいは

$$\Sigma_{13} = i_{13} \sqrt{\left\{ \frac{\sigma_1^2}{m_1^2} + \frac{\sigma_3^2}{m_3^2} - 2 \frac{\sigma_1}{m_1} \cdot \frac{\sigma_3}{m_3} r_{13} \right\}}$$

III 四つの絶対測定値の相関係数とそれらの標準偏差によつて二つの指数の相関係数をみつける事。

x_1/x_3 と x_2/x_4 を二つの指数とする。そうすると ρ がこれらの二つの間の相関係数であれば三乗の項は無視出来るので

$$\begin{aligned} NP \Sigma_{13} \Sigma_{24} &= S\left(\frac{x_1}{x_3} - i_{13}\right) \left(\frac{x_2}{x_4} - i_{24}\right) \\ &= \frac{m_1 m_2}{m_3 m_4} S\left(1 + \frac{\epsilon_1}{m_1} - \frac{\epsilon_3}{m_3} - \frac{\epsilon_1 \epsilon_3}{m_1 m_3} + \frac{\epsilon_3^2}{m_3^2} - 1 - \frac{\sigma_3^2}{m_3^2} + \frac{\sigma_1}{m_1} \cdot \frac{\sigma_3}{m_3} r_{13}\right) \\ &\quad \times \left(1 + \frac{\epsilon_2}{m_2} - \frac{\epsilon_4}{m_4} - \frac{\epsilon_2 \epsilon_4}{m_2 m_4} + \frac{\epsilon_4^2}{m_4^2} - 1 - \frac{\sigma_4^2}{m_4^2} + \frac{\sigma_2}{m_2} \cdot \frac{\sigma_4}{m_4} r_{24}\right) \\ &= i_{13} i_{24} S\left(\frac{\epsilon_1}{m_1} - \frac{\epsilon_3}{m_3}\right) \left(\frac{\epsilon_2}{m_2} - \frac{\epsilon_4}{m_4}\right) \end{aligned}$$

従つて

$$\rho \Sigma_{13} \Sigma_{24} = i_{13} i_{24} \left(\frac{\sigma_1}{m_1} \cdot \frac{\sigma_2}{m_2} r_{12} - \frac{\sigma_1}{m_1} \cdot \frac{\sigma_4}{m_4} r_{14} - \frac{\sigma_3}{m_3} \cdot \frac{\sigma_2}{m_2} r_{23} + \frac{\sigma_3}{m_3} \cdot \frac{\sigma_4}{m_4} r_{34} \right)$$

故に

$$\rho = \frac{\frac{\sigma_1}{m_1} \cdot \frac{\sigma_2}{m_2} r_{12} - \frac{\sigma_1}{m_1} \cdot \frac{\sigma_4}{m_4} r_{14} - \frac{\sigma_2}{m_2} \cdot \frac{\sigma_3}{m_3} r_{23} + \frac{\sigma_3}{m_3} \cdot \frac{\sigma_4}{m_4} r_{34}}{\sqrt{\left\{ \frac{\sigma_1^2}{m_1^2} + \frac{\sigma_3^2}{m_3^2} - 2 \frac{\sigma_1}{m_1} \cdot \frac{\sigma_3}{m_3} r_{13} \right\}} \sqrt{\left\{ \frac{\sigma_2^2}{m_2^2} + \frac{\sigma_4^2}{m_4^2} - 2 \frac{\sigma_2}{m_2} \cdot \frac{\sigma_4}{m_4} r_{24} \right\}}}$$

定理Ⅰは、指数の平均値は対施する絶対測定値の平均値の比でない事を示し、定理Ⅱは指数の作る四つの対象が全く無相関のときに、 ρ は0になる筈である事を示している。しかし σ_3 と σ_4 の二つが等しければそのとき $r_{34}=1$ 且 $\frac{\sigma_3}{m_3} = \frac{\sigma_4}{m_4}$ であるから、 ρ の式をうる。

$$\rho = \frac{\frac{\sigma_1}{m_1} \cdot \frac{\sigma_2}{m_2} r_{12} - \frac{\sigma_1}{m_1} \cdot \frac{\sigma_3}{m_3} r_{13} - \frac{\sigma_2}{m_2} \cdot \frac{\sigma_3}{m_3} r_{23} + \frac{\sigma_3^2}{m_3^2}}{\sqrt{\left\{ \frac{\sigma_1^2}{m_1^2} + \frac{\sigma_3^2}{m_3^2} - 2 \frac{\sigma_1}{m_1} \cdot \frac{\sigma_3}{m_3} r_{13} \right\}} \sqrt{\left\{ \frac{\sigma_2^2}{m_2^2} + \frac{\sigma_3^2}{m_3^2} - 2 \frac{\sigma_2}{m_2} \cdot \frac{\sigma_3}{m_3} r_{23} \right\}}}$$

これは前に引用した支払制限付養老保険の場合に適用出来るかも知れない。

対象 X_1, X_2, X_3 が互に無相関で且 X_1/X_3 と X_2/X_3 の指数を作るときは興味ある特別な場合が起る。そこでは

$$\rho = \frac{\frac{\sigma_2^2}{m_2^2}}{\sqrt{\left\{ \frac{\sigma_1^2}{m_1^2} + \frac{\sigma_3^2}{m_3^2} \right\}} \sqrt{\left\{ \frac{\sigma_2^2}{m_2^2} + \frac{\sigma_3^2}{m_3^2} \right\}}}$$

S 11

前述の事から学ぶべき擬相関についての実際の教訓は(1)同質的なデータを扱う事であり、また、これらの経験表が同質的であると言う事が確信出来るまで混合した経験表の場合についての相関係数の値を余り信じてはならない事である。(2)つぎのような相関表を作る事をさけるか、あるいは説明に注意深くする事である。即ち相関表での相関連した函数は分母が同じであるか、あるいは分母自身相関連する事も出来るような指数として表わされている。擬相関は相互対が年に関係し、しかも時間に関しては無作意に

取られないとき起る事を付け加えよう。しかし又つの n 次階差の間の相関が又つの $(n+1)$ 次階差の間の相関に等しいなら我々は時間に無関係な相関に到達しそして各変量が時間に関して従属であれば $a + bt + ct^2 + \dots$ の形をとる。

「10ヶ年を前とした11年」 (8ヶ年の賃時がどの組にも共通の場合)

年 の 出 た 数	最初の投げて出た数										計 合	行の 平均値
	0	1	2	3	4	5	6	7	8	9	10	
0	1	2	1								4	10
1	2	12	18	8							40	18
2	1	18	61	72	28						180	26
3		8	72	176	168	56					480	34
4			28	168	322	252	70				840	42
5				56	252	392	252	56			1008	50
6					70	252	322	168	28		840	58
7						56	168	176	72	8	480	66
8							28	72	61	18	180	74
9								8	18	12	40	82
10									1	2	4	90
合計	4	40	180	480	840	1008	840	480	180	40	4096	
列の平均値	10	18	26	34	42	50	58	66	74	82	90	

(5ヶの貨幣がどの組にも共通の場合)

第II回の投げ で表の出た数	第I回の投げで表の出た数										合計	行の平均
	0	1	2	3	4	5	6	7	8	9	10	
0	1	5	10	10	5	1					32	2.5
1	5	30	75	100	75	30	5				320	3.0
2	10	75	235	400	400	235	75	10			1440	3.5
3	10	100	400	860	1100	860	400	100	10		3840	4.0
4	5	75	400	1100	1780	1780	1100	400	75	5	6720	4.5
5	1	30	235	860	1780	2252	1780	860	235	30	8064	5.0
6		5	75	400	1100	1780	1780	1100	400	75	6720	5.5
7			10	100	400	860	1100	860	400	100	3840	6.0
8				10	75	235	400	400	235	75	1440	6.5
9					5	30	75	100	75	30	320	7.0
10						1	5	10	10	5	32	7.5
合計	32	320	1440	3840	6720	8064	6720	3840	1440	320	32768	
列の平均	2.5	3.0	3.5	4.0	4.5	5.0	5.5	6.0	6.5	7.0	7.5	

(2ヶの貨幣がどの組にも共通の場合)

第II回の投げ で表の出た数	第I回の投げで表の出た数										合計	行の平均
	0	1	2	3	4	5	6	7	8	9	10	
0	1	8	28	56	70	56	28	8	1		256	4.0
1	8	66	240	504	672	588	336	120	24	2	2560	4.2
2	28	240	913	2024	2884	2744	1750	728	184	24	11520	4.4
3	56	504	2024	4768	7280	7504	5264	2464	728	120	30720	4.6
4	70	672	2884	7280	11956	13328	10192	5264	1750	336	53760	4.8
5	56	588	2744	7504	13328	16072	13328	7504	2744	588	64512	5.0
6	28	336	1750	5264	10192	13328	11956	7280	2884	672	53760	5.2
7	8	120	728	2464	5264	7504	7280	4768	2024	504	30720	5.4
8	1	24	184	728	1750	2744	2884	2024	913	240	11520	5.6
9		2	24	120	336	588	672	504	240	66	2560	5.8
10			1	8	28	56	70	56	28	8	256	6.0
合計	256	2560	11520	30720	53760	64512	53760	30720	11520	2560	262144	
列の平均	4.0	4.2	4.4	4.6	4.8	5.0	5.2	5.4	5.6	5.8	6.0	

10ヶのサイコロを組にして投げた場合（各組に共通な8ヶのサイコロ）

第II回の投げで表の出る数	第I回の投げで表の出る数										合計	行の平均
	0	1	2	3	4	5	6	7	8	9	10	
0	4096	4096	1024								9216	0.6
1	4096	20480	17408	4096							46080	1.46
2	1024	17408	45312	32768	7168						103680	2.26
3		4096	32768	58368	35840	7168					138240	3.06
4			7168	35840	48384	25088	4480				120960	3.86
5				7168	25088	26880	11648	1792			72576	4.6
6					4480	11648	10080	3584	448		30240	5.46
7						1792	3584	2496	704	64	8640	6.26
8							448	704	384	80	1620	7.06
9								64	80	32	180	7.86
10									4	4	9	8.6
合計	9216	46080	103680	138240	120960	92576	30240	8640	1620	180	531441	
列の平均	0.6	1.46	2.26	3.06	3.86	4.6	5.46	6.26	7.06	7.86	8.6	

質へ1投げの2つの表を合計して作った表（17の表は180と同じでも1つの表は見出しを）

第II回の投げで表の出る数	第I回の投げで表の出る数													合計	行の平均
	0	1	2	3	4	5	6	7	8	9	10	11	12		
0	1	5	10	10	5	1								32	2.5
1	5	30	75	100	75	30	5							320	3.0
2	10	75	235	400	400	235	75	10						1440	3.5
3	10	100	400	861	1105	870	440	105	11					3872	4.012
4	5	75	400	1105	1810	1855	1200	475	105	10				7040	4.568
5	1	30	235	870	1855	2487	2180	1260	470	105	11			9504	5.226
6		5	75	440	1200	2180	2640	2300	1260	475	105	10		10560	6.045
7			10	105	475	1260	2200	2640	2180	1200	410	75	5	10560	6.955
8				11	105	470	1260	2180	2487	1855	870	235	30	9504	7.774
9					10	105	475	1200	1855	1810	1105	400	75	7040	8.432
10						11	105	410	870	1105	861	400	100	3872	8.988
11							10	75	235	400	400	235	75	1440	9.5
12								5	30	75	100	75	30	320	10.0
13									1	5	10	10	5	32	10.5
合計	32	320	1440	3872	7040	9504	10560	10560	9504	7040	3872	1440	320	65536	
列の平均	2.5	3.0	3.5	4.012	4.568	5.226	6.045	6.955	7.774	8.432	8.988	9.5	10.0	10.5	

第 IX 章 定量的に測れない特性の相関

§ 1

本節での理論を討論する前に我々は種痘統計と *Sheffield*⁽¹⁾ 地方に 1887 年〜8^{*} 年に発生した天然痘に関するものから取られた問題を取扱ったこの種の問題を示している表を与えよう。

さきめのある種痘の度数	天然痘にかかったときの抵抗する強さ			
	傷 痕	回復したもの	死 亡 数	計
	あるもの	3,951	200	4,151
	ないもの	278	274	552
	計	4,229	474	4,703

我々が関心を持つ特性は“天然痘にかかった時の抵抗の強さ”と“さきめのある種痘の度数”である。そして統計量はこれより詳しい方法で並べる事は出来ない。特性は定量的に測る事は出来ないが、そのような測定値がないからと言ってそこに相関がない事は意味しない。我々はそのような場合にどんな係数が得られるかを考えなければならぬ。

§ 2

一般的考察から相関係数にある値を指定出来るような二三の例を書き当てる事が出来るかどうかをみる事からこの問題を考察しよう。貨幣投げを行えば“表”か“裏”が出なければならない。

* *Biometrika* I 375 頁以後。この論文は W. R. Macdonell によつてかけられたものでありその附録には種痘の側に関する明確な効力を示す方法に関する問題を取扱っている、順向の種類も研究されている。実際的問題点はしばしば省略されている。

(1) イングランド中部の工業地帯

対のやり方を次々に行つて第四章のような表を作ればそこには相関はないしまた前節で作つたような表では各分割の中の数は等しくなるであらう。

しかしもしテーブルの上に残した貨幣を射を作りオ II 回の投げの数を数えれば我々は絶対的相関を持つであらう。そして表の左上と右下の分割の数は等しくなり他の二つの分割は空白である。もしこれら二つの表を混みにし各々の計が 4 である事を仮定すれば下に示すような 0.5 の相関係数をもっている表をうる。

第 II 回の投げ	第 I 回の投げ ^a		計
	表	裏	
表	3	1	4
裏	1	3	4
計	4	4	8

これらの簡単な場合では 4 つの分割は 4 分離点での度数を表わしている。Y の正確な値は次式で与えられる。

$$\frac{ad - bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}$$

ここに a, b, c, d は § 5 の表に示す意味のものである。

§ 3

この式はここでは 4 つの分割しかないけれども、測られた事柄がたとえ詳しい方法でそれを表わせないにしても連続尺度の測定値を意味するから凡ての場合に用いてもよいものであると解してはいけない。従つて前の種痘統計でさきめのある種痘の程度は従来のときに 40 才であった人に対しては子供の時の種痘と最近おこなわれた種痘までの各種痘の系列の間を変化する。更に病気に罹つたときの回復力は“回復”と“死亡”の二つの分割に示されているものよりももっと広い尺度にあるように考えられている。

他の例をとると親子の目色の研究をすれば黒色から薄青色或は黒色から白子の場合の色素の不足に至るまでの色の段階を作る事が出来るが統計量は唯単に“茶色”であるか“茶色でない”と言う形で用いられる。換言すれば四重分類の相関表その統計量は154頁に示したような連続的度数分布に關係するのである。しかしこの方法のために集めねばならない事実、あるいは陳述されねばならない事実は全資料に対して唯四つの分割があるにすぎない。当初の問題は5.2の終りに述べた簡単な公式がこのような事情の下で満足出来る答を与えてくれるかどうかを考察する事であつた。そしてもしそれが失敗だつたならどんなものを採用したらよいかを考える事であつた。

5.4

第兩章では貨幣投げに關した表を与え、これらの表の一つの資料を四つの分割にグループ化問題の公式がどんな答を与えるかを見る事が出来る。

もし0.5の相関をもっている表をとり“5ヶの表”と“6ヶの表”の間で切れれば下の様な表をうる。

二回目の投げで 表の出た数	最初の投げで表の出る数		計
	0~5	6~10	
0~5	15.330	5.086	20.416
6~10	5.086	7.266	12.352
計	20.416	12.352	32.768

公式は

$$\frac{15.330 \times 7.266 - 5.086 \times 5.086}{20.416 \times 12.352} = 0.34$$

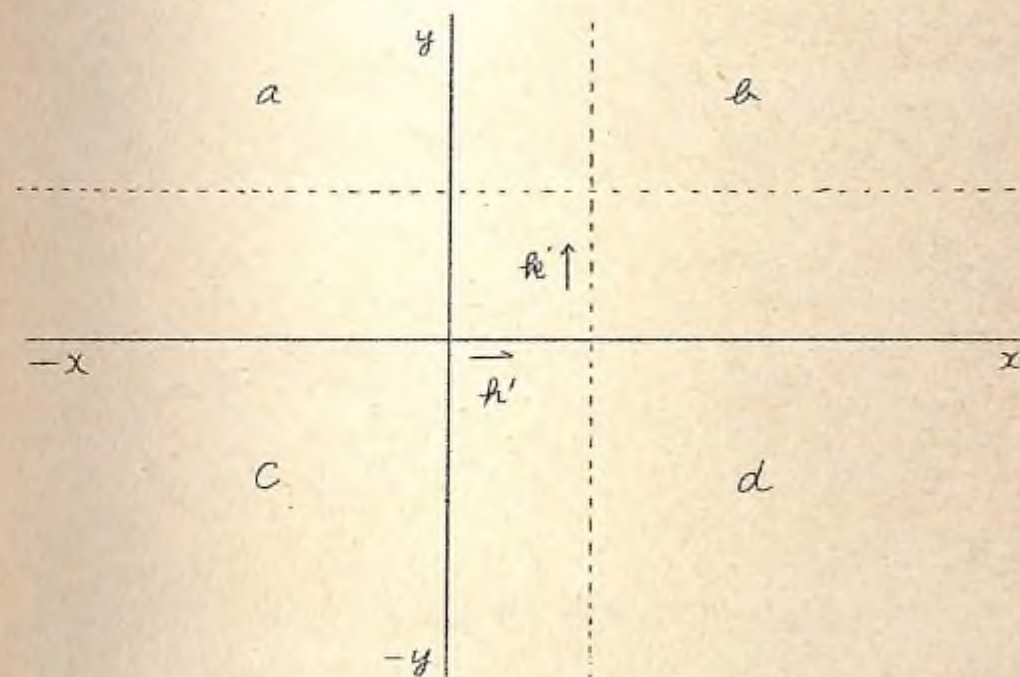
を与える。これは0.5の真値からかなり変つてゐる。

5.5

他の解答を探さねばならない事はこの事から明らかである。また度数曲面を次のような式で表わせる事は前章で學んでゐる。

$$Z = \frac{N}{2\pi\sqrt{(1-r^2)\sigma_1\sigma_2}} e^{-\frac{1}{2(1-r^2)}\left(\frac{x^2}{\sigma_1^2} + \frac{y^2}{\sigma_2^2} - \frac{2xyr}{\sigma_1\sigma_2}\right)}$$

そこで下の図に示されているように原点から a' 、 b' の距離にあり x 、 y 軸に直角な二つの平面で上の度数曲面を4つの部分に分割すればどんな結論が導かれるかを考えよう。



度 数 表		
a	b	a + b
c	d	c + d
a + c	b + d	N

そうすると

$$d = \frac{N}{2\pi N(1-r^2)\sigma_1\sigma_2} \int_{h'}^{\infty} \int_{h''}^{\infty} e^{-\frac{1}{2(1-r^2)}(\frac{x^2}{\sigma_1^2} + \frac{y^2}{\sigma_2^2} - \frac{2rxy}{\sigma_1\sigma_2})} dx dy$$

ここで $\frac{x^2}{\sigma_1^2}$ に対し X^2 , $\frac{y^2}{\sigma_2^2}$ に対し Y^2 と変数変換を行い $h = \frac{h'}{\sigma_1}$

$h'' = \frac{h'''}{\sigma_2}$ とかけば上式は

$$= \frac{N}{2\pi N(1-r^2)} \int_h^{\infty} \int_{h''}^{\infty} e^{-\frac{1}{2(1-r^2)}(Y^2 + X^2 - 2rXY)} dX dY$$

その上

$$b + d = \frac{N}{\sqrt{2\pi}\sigma_1} \int_{h'}^{\infty} e^{-\frac{1}{2}\frac{x^2}{\sigma_1^2}} dx$$

$$= \frac{N}{\sqrt{2\pi}} \int_h^{\infty} e^{-\frac{1}{2}X^2} dX$$

また

$$c + d = \frac{N}{\sqrt{2\pi}} \int_{h''}^{\infty} e^{-\frac{1}{2}Y^2} dY$$

全度数 $N = a + b + c + d$ に注意すれば

$$N - 2(b + d) = N - N\sqrt{\frac{2}{\pi}} \int_h^{\infty} e^{-\frac{1}{2}X^2} dX$$

$$\therefore \frac{(a+c)-(b+d)}{N} = \sqrt{\frac{2}{\pi}} \int_0^h e^{-\frac{1}{2}X^2} dX$$

同様にして

$$\frac{(a+b)-(c+d)}{N} = \sqrt{\frac{2}{\pi}} \int_0^{h''} e^{-\frac{1}{2}Y^2} dY$$

a, b, c, d は既知であるから h と h'' は Sheppard の表から求める事が出来る。そして問題は次の方程式から r の値を求める事に帰着する。

$$\frac{N}{2\pi\sqrt{2\pi}} \int_h^{\infty} \int_{h''}^{\infty} e^{-\frac{1}{2(1-r^2)}(X^2 + Y^2 - 2rXY)} dX dY = d$$

そこで d, N, h, h'' は既知である。

この解は次の方程式を導く (附録 IV をみよ)

$$\begin{aligned} \frac{ad-bc}{N^2HK} = & r + \frac{r^2}{2} h h'' + \frac{r^3}{6} (h^2-1)(h''^2-1) + \frac{r^4}{24} h(h^2-3)h''(h''^2-3) \\ & + \frac{r^5}{120} (h^4-6h^2+3)(h''^4-6h''^2+3) + \frac{r^6}{720} h(h^4-10h^2+15)h''(h''^4 \\ & -10h''^2+15) + \frac{r^7}{5040} (h^6-15h^4+45h^2-15)(h''^6-15h''^4+45h''^2-15) \\ & + \dots \end{aligned}$$

$$\therefore H = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}h^2} \quad \text{また} \quad K = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}h''^2}$$

数値解は近似法で得られねばならない。Newton * の方法は近似根を求めるのに便利である。

§ 6

§ 1 の例題の数値的解法は次のようである。

$$\begin{aligned} \sqrt{\frac{2}{\pi}} \int_0^h e^{-\frac{1}{2}X^2} dX &= \frac{(a+c)-(b+d)}{N} = \frac{3755}{4703} \\ &= 0.7984265 \end{aligned}$$

故に Sheppard の表 (Tables for Statisticians 参照) の補挿によつて

$$h = 1.27716$$

本方程式の根に対するニュートンの近似法:

$f(x)=0$ を満たす x を求めるような方程式である。 h は $x=h+h'$ を満たすような x に近い一つの値とする。 $\therefore$ h は小さいものである。そうすると $f(x)=f(h+h')=f(h)+h'f'(h)+\dots$ (Taylor の定理により h' の高次の項を含む項) となりまた $f(x)=0$ だから $h' = -\frac{f(h)}{f'(h)}$ あるいは $x = h - \frac{f(h)}{f'(h)}$ をうる。この方法のおもな欠点は根に近い値が沢山あると言う事であるがこの事は相関の応用にまで考える事は必要ない。(年金から利率を近似する事は Todhunter の Interest and annuities Certain 177 頁 公式 2 を参照せよ)

をうる。このような目的に対してこれらの表を用いるときは数値
0.7984265は *Sheppard* の記号 α に対応している事に注意
せよ。そうすると

$$\frac{1}{2}(1 + 0.7984265) = 0.8992132$$

は *Sheppard* の表 II までに見て行かぬばならない。
もし *Sheppard* 表を使えば 0.7984265 を加えねばならない。

同じように

$$\int_0^{\sqrt{2}} e^{-\frac{1}{2}y^2} dy = 0.7652561$$

$$\therefore k = 1.18833$$

次に $\frac{ad-bc}{N^2HK}$ を求める *Sheppard* の表から先づ

$$H = 0.1764870 \quad \therefore \log H = T.2467127$$

$$K = 0.1969111 \quad \therefore \log K = T.2942702$$

をうる。故に $\log \frac{ad-bc}{N^2HK} = 0.1258266$

$$\frac{ad-bc}{N^2HK} = 1.336062$$

Macdonell は最後の二つの数字として 62 の代りに 56 を与
えている。この違いは多分補挿によるものであろう。

r の式に立ちもどつて考えれば k は r^2, r^4, r^6 等の係数に
なっている積であることに注意すればその値は充分に算出され係
数が見つけられているからその事については注意が必要である。
また k と k のあるいは 7 次の中から書き下す事によって計算を
始める事が勧められている。

Macdonell は次の級数を与えている。

$$0.0970838r^7 + 0.0081708r^6 + 0.119614r^5 + 0.137450r^4 \\ + 0.043352r^3 + 0.758844r^2 + r = 1.336056$$

r を求めるために我々は第 1 近似として r の真値に近い値を見付
けねばならない。

$$0.758844r^2 + r - 1.336056 = 0$$

をとれば

$$r = \frac{-1 + \sqrt{1 + 4 \times 1.336 \times 0.7588}}{1.5177} = 0.8$$

をうる。

さてこの値は真値より大きいがそれは r を見付けるための方程
式の左辺の二つの項しか使われなかつたためである。また仮想根と
して 0.77 をとつてもよからう。Newton の法則を応用すれば次
をうる。

$$\begin{aligned} & -1.336056 + (0.77) + 0.7588(0.77)^2 + 0.0434(0.77)^3 \\ & + 0.1375(0.77)^4 + 0.1196(0.77)^5 + 0.0082(0.77)^6 + 0.0971(0.77)^7 \\ r = 0.77 & \frac{1 + 2(0.77)(0.7588) + 3(0.77)^2(0.0434) + 4(0.77)^3(0.1375) \\ & + 5(0.77)^4(0.1196) + 6(0.77)^5(0.0082) + 7(0.77)^6(0.0971)}{1 + 2(0.77)(0.7588) + 3(0.77)^2(0.0434) + 4(0.77)^3(0.1375)} \\ & = 0.77 - \frac{0.0022}{2.961} \\ & = 0.7692 \end{aligned}$$

このような計算には自然数の市乗の表等が便利である。

57

数値計算に必要な種々の函数表は *Tables for Statisticians* に
求められる。テトラコリック函数 (Tetrachoric function) *
もここに採用されている。これらの表が準備出来るから我々は次
のような方程式を使う事が出来る。

$$d/N = T_0T_0' + T_1T_1'r + T_2T_2'r^2 + \dots \quad (1)$$

ここに T の値は T_{19} まで表にされていてそれ以上の値は表の序文
に説明してある階差公式で求める事が出来る。

係数の計算は詳しく上に述べられているが *Tables for Statisticians Part II Tables VII, IX* を用いればこれらの計算の大部
分を避ける事が出来る。これらの表が採用出来れば、(1) 式とに示
したような d/N の計算、(2) 全数に関して象限にある数 d の比
すなわち d/N の計算、また (3) r を求めるために表を補挿する事

の凡てがなされる筈である。

* *Tetrachoric function* は *Hermite* 多項式の似たものであり、
有用な完全な表が用意されている

8番目の *Tetrachoric function* は

$$\tau_8(h) = \frac{(-1)^{8-1}}{\sqrt{8!}} \frac{d^{8-1}}{dh^{8-1}} \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}h^2} \right)$$

であり、また (8-1) 番目の *Hermite* 多項式は次の通りである。

$$H_{8-1}(h) = \frac{(-1)^{8-1}}{\sqrt{(8-1)!}} \frac{d^{8-1}}{dh^{8-1}} \left(\frac{1}{\sqrt{(2\pi)}} e^{-\frac{1}{2}h^2} \right)$$

(1)

188頁から

$$\left. \begin{aligned} \frac{1}{\sqrt{(2\pi)}} \int_{f_0}^{\infty} e^{-\frac{1}{2}y^2} dy &= \frac{c+d}{N} \\ \frac{1}{\sqrt{(2\pi)}} \int_{f_0}^{\infty} e^{-\frac{1}{2}x^2} dx &= \frac{b+d}{N} \end{aligned} \right\} \text{ から } f_0, f_0 \text{ が求められる}$$

$$\text{また } \frac{d}{N} = \frac{1}{2\pi\sqrt{(1-r^2)}} \int_{f_0}^{\infty} \int_{f_0}^{\infty} e^{-\frac{1}{2(1-r^2)}(x^2+y^2-2rxy)} dx dy$$

である。

次に x, y の同時分布の特性函数による表示

$$\begin{aligned} f(x, y) &= \frac{1}{2\pi\sqrt{(1-r^2)}} e^{-\frac{1}{2(1-r^2)}(x^2+y^2-2rxy)} \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(t^2+2rtu+u^2)} e^{-itx-iuy} dt du \end{aligned}$$

を使って

$$\begin{aligned} \frac{d}{N} &= \int_{f_0}^{\infty} \int_{f_0}^{\infty} f(x, y) dx dy \text{ を } r \text{ の巾級数に展開すると} \\ &= \sum_{j=0}^{\infty} \tau_j(h) \tau_j(f_0) r^j \end{aligned}$$

が証明される。この値は表から求められ r の高次の項を省略した方程式から
求められる [M. G. Kendall (前掲書) P354 統計数値表 I, P52 参照]

58

さて我々は貨幣投げの問題に立ちもどしてみよう。若し今論議
した方法で 54 に与えられた表の相関係数を計算すれば 0.51 と
0.52 の間にある値をうるだろう。この値は、特に貨幣投げが絶
対的な連続尺度でなかつた事を思い出せば非常によい結果である。
予想される普遍的な結論は我々が試みたこの種の場合に一つの仮
定が適当な結果を導くと言う事である。度数曲面が平均値からか
なり離れた所で切られるときにはこの方法はそんなに具合よくは
解けないしまたそんな場合の数値結果は正確さを仮定してはいけ
ない。

59

今までは利用出来るデータは唯4つの分割に分けられる事を
仮定して来た。そして特性値が定量的には測定出来ないが数種に
分割される時の相関の論究は後(第12章)まで延期しよう。さ
て次に一つは変量であり他は定量的に測定出来ないような場合
例えば194頁にある子供(少年)の体重に関する (*enlarged*^{増大線}
glands) の効果についての表のような場合を扱う事が出来る。^{*}
統計量は“良い腺”と“悪い腺”に分けられているが、腺の状態
は連続的変量である。悪い腺をもっている少年はそうでないも
のよりも体重は軽い。

読者が養老保険に対して作ったような完成された表からか、あ
るいは夫と妻の年令の関係を与える相関表から容積度数を考える
なら完全な分布を得た事を知るであろう。さてもし容積度数を一
つの変量のある値での垂直面によつてこの全度数から切り離すと
此の量の重心を通る垂線は回帰線を切る。= 重分類表での垂直面
は行の分割にある。すなわち我々の例では悪い腺から始まり良い
腺で終る。F と G が区割点の座標ならこの容積の重心を通る垂線
もこの方法は K. Pearson によつて *Biometrika* 頁 96 頁以后に与え
られている本例はそこから引用したものである。

は回帰線を切っている。そうすると σ_1 , σ_2 が二変量の標準偏差となり r が相関係数であるから 次をうる

$$\bar{P} = r \frac{\sigma_1}{\sigma_2} \bar{Q} \quad \text{あるいは} \quad r = \frac{\bar{P}}{\sigma_1} / \frac{\bar{Q}}{\sigma_2}$$

体 重	良い腺をもつ少年	悪い腺をもつ少年	計
14	2	—	2
16	3	5	8
18	15	26	41
20	20	40	60
22	28	47	75
24	34	30	64
26	30	31	61
28	29	20	49
30	30	30	60
32	21	14	35
34	18	11	29
36	18	5	23
38	6	7	13
40	5	2	7
42	7	3	10
44	1	—	1
46	—	—	—
48	3	—	3
50	1	—	1
52	1	—	1
—	—	—	—
62	2	—	2
計	274	271	545

今戸は二つの変量の或る一つとの凡ての組に対して定量的に測定したものの平均値である。我々の例では悪い腺の少年の平均体重である。 σ_1 は全少年の標準偏差である。同じように $\bar{Q}$ と σ_2 は計算出来ない。何故かと言えばそれらは定量的な測定が意味を持たない腺に関係しているからである。正規確率分布に従う測定可能な変量(腺)を仮定し、確率積分の表を使ってこの変量の母集団はこの変量が現われる区間での平均値からの距離に対して $\frac{y}{\sigma_2}$ の

比を与えると仮定すれば そのとき

$$\begin{aligned} \frac{\bar{Q}}{\sigma_2} &= \frac{N}{\sqrt{2\pi}\sigma_2} \int_{-\infty}^{\infty} y e^{-\frac{1}{2}\frac{y^2}{\sigma_2^2}} dy / \frac{N}{\sqrt{2\pi}\sigma_2} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\frac{y^2}{\sigma_2^2}} dy \\ &= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y}{\sigma_2}\right)^2} / \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}y^2} dy \end{aligned}$$

確率積分に関して *Sheppard* の表で用いられている記号では分子が z であり分母は $\frac{1}{2}(1-\alpha)$ である。

数値例を解く事は方法をはっきりさせるに役立つだろう。それは次の通りである。

全少年の平均体重は 27.7522

標準偏差は 6.7502

悪い腺をもつ少年の平均体重は 27.3737

$$\frac{1}{2}(1-\alpha) = 271/545 = 0.4972$$

$$\frac{1}{2}(1+\alpha) = 0.5028 \quad \text{そしてこの値が } \textit{Sheppard} \text{ の表での } z=0.3989 \text{ に対応する。}$$

腺と体重の間の相関係数は

$$\frac{(27.7522 - 27.3737)}{6.7502} / \frac{0.3989}{0.4972}$$

$$= 0.070$$

$\frac{1}{2}(1-\alpha)$ を使うのは合計度数の小さい方の列をとる事が仮定されていることは注意されたい。従つて我々の例では良い腺をもつ少年より悪い腺をもつ少年の方がより少ないのである。

§ 10

この例は実際の相関係数を計算する前に統計量の全般をみる事と、そこに何らかの相関があるかどうかについて予備的観念を考

えるのが得策であろうと言う実践的見地を指摘している。154頁のような表で凡ての行の平均値が同じような値でありまた凡ての列の平均値がどれも同じようであればそこに相関はない。同じように184頁のような表において各棒内の数値が合計に比例していればもはやそこに相関はない。59の例で合計の列に内側の又列を比較すれば、もしその間に何らかの相関があつたとしてもそれは小さくなければならぬ事を示している。何故なら合計の列の分布は中の列の各々によつて補整される可能性を与えるからである。

第X章 標準誤差

§1

統計的工作において我々は沢山の測定値から一つの平均値を計算し、そこで我々が関係したものの持つ平均値をはつきりと確定する事であると考える事も出来る。数値計算は非常に詳細な点までなされており、また測定が正確になされていても統計量から求められる平均値は特性値の真平均値と異つてゐる。何故なら測られるものは数に限界があるからである。換言すれば標本によつてそれが扱かれた無限母集団を正確に表わす事は出来ないからである。もし5ヶの貨幣を投げて表の出た回数を記録すると、次のような表を得るだろう。ここでは140回の反復実験の結果が与えられている。*

各回の試行で表の出た数	左のランに記された数だけ表の出た実験回数
5	4
4	24
3	49
2	40
1	20
0	3
	140

さてこれを単なる統計的問題として取扱うとき、我々は表のみについて何も先験的に知つていないから「表」の出た平均回数として2.6を計算するかも知れない。この事はこの平均値以外のものは起らないと言うことを証明はしない。もう一度実験すると異なる値を与えるかも知れない。それ故我々の計算からどの平均値が真値であると言う事は出来ない。しかし他の方法でこの問題に近づく事が出来る。すなわち、真平均値からの偏差がどんな分

*このような実験は $(\frac{1}{2} + \frac{1}{2})^5$ のような無限母集団から140の標本を無作為に抽出したものと考える事も出来る。

布をするかを定める事が出来るし、ある実験あるいは実験の系列から計算した平均値は真値とどれ位異なるかを評価出来る。

實際上真平均値は100回の試行で1回だけある小さな量(ε)以上は計算値とちがわないと言う事が出来れば我々は満足する事が出来る。実際に使われている尺度に入る前に準備的方針に關する二三の点を考へよう。

凡ての統計的実験は次の事を確信させる。すなわち1,000回も繰返した実験は50回繰返された同様の実験よりもあてになるに違いない。

だから真平均値と計算値の間の差異とは、ある幾つかの点で実験回数に依存する筈であると予想する。更に、広い範囲に広がった分布は一点に集中した分布よりも真平均値からせれる機会が多い平均値を与える。だからとは分布の広がり、すなわち分布の標準偏差に依存すると予想する事が出来る。

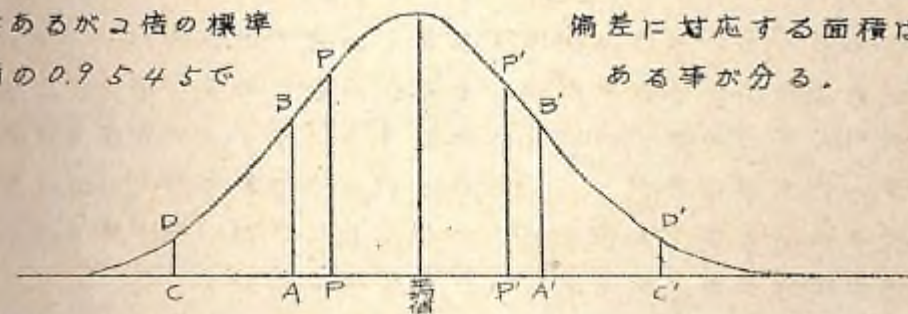
§2

これらの注意は凡ての統計的尺度に適用されるものである。測定値は不正確であり、真値への近似に過ぎないが、もし無限の事象を取扱う事が出来れば、そこから得られた結果は真値とある量以上違わないだろうと言う事が非常な確らしさを与える。今までの議論は平均値について話されて来たが他の尺度も同じような一般的考察に従うし、それ故我々はどんな値が平均値、標準偏差、相関係数あるいは他の統計的尺度に対する小さな誤差を割り当てられるかを考へねばならない。この誤差は標準偏差に依存する筈だと予想して来たしこの事は時にはつまらぬ混乱を導くこともある。何故なら標準偏差に關する誤差の分布に注意を払わねばならないからである。さて夫と妻の結婚時の年齢と言うような100対の相関係数を計算したと仮定しよう。それから他の100対の相関係数を計算すると言うようにこれらの結果の大きな数をうるまで続ける、係数はずつと前の章で論議したある度数分布に似た分布

をする。我々に關係のあるのはその分布の平均値からの標準偏差である。同様に§1の貨幣投げの実験を再三再四繰返し各回の実験から平均値を計算する。平均値は5と0の間の任意の値をうるだろう。これらの両極端は凡ての試行で5ヶとも表が出るか凡ての試行で一ヶも表が出ないかの最も起りそうもない場合が起つたにすぎない。最も起りそうな平均値は2.5である。そしてもし十分に実験を繰返せば平均値が分布する状態についての考え方を形作る事が出来よう。すなわち我々はそれ自身の平均値、標準偏差をもっている平均値の度数分布に準ずるだろう。また平均値が計算されていることから§1の表にあるような度数分布と混合してはならない。統計的測度(平均値、相関係数等のような)の度数分布に關するときに“標準偏差”と言う替りに標準誤差と言へば考え方の混乱をさける助けになるだろう。せうすると標準誤差は我々が考察している特別の度数分布の標準偏差である。

§3

標準誤差の導入について最も簡単な鈴型度数曲線即ち誤差の正規曲線を考へる事が出来る。もし統計的測度の分布がこの型を取ればどんな結論が引き出せるかをみる事が出来る。実際に統計的測度の分布は標本その実験数が多いときに至しそうな型でなければならぬ事が示されている。従つてたとえ§1の分布が対称的の代りに歪んでいても平均値の分布は前に得られた歪んだ分布よりも正規分布型に近い型となる筈である。正規曲線の表を参考にすれば標準偏差に対応する面積は全面積のほぼ3分の2であるが2倍の標準偏差に対応する面積は全面積の0.9545である事が分る。



換言すれば若し統計的尺度の分布がこの正規分布すれば、2倍の標準誤差より大きい誤差を生ずる事は200回の実験で9回以上起らないと言う事が出来る。それ故 我々の取扱っている実験にこの特別の場は起りそうもないのである。上の図はこの議論を理解するのに役立つでしょう。平均値から標準偏差の距離にあるAとA' B'の向にある面積は全面積のほぼ2/3である。線分CDとC'D'は平均値から2倍の標準偏差だけ離れていてそれは全曲線の殆ど全部を含んでいる。従つてこれらの線分外の部分は相対的に小さな面積でなければならない。

§4

以前は確率誤差として知られている他の函数を使う事が習慣であつた。その確率誤差は標準誤差の0.67449倍である。確率誤差は絶対的誤差を表わしている正規曲線を二つの等しい部分に分けるXすなわち(P)の値を与えている。故に確率誤差は

$$\int_0^P \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx = 0.25$$

によつて与えられる。ここで曲線の全面積(正の偏差+負の偏差)は1である。標準偏差の方からPを発見するために *Tables for Statisticians Part I, Table III*、或は附録IXの表にある $\frac{1}{2}(1+\alpha) = 0.75$ に対応するXの値を求めねばならない。ここで α は

$$\int_{-x}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx$$

である。この表は逆検索する事が出来て近似的に $P=0.67449$ である事が分る。平均値、あるいはむしろ平均値を通る垂線は全分布を二つの等しい部分に分割する。従つて確率誤差は全分布を4つの部分に分割し、Galtonが四分位数と呼んだものを与える。その位置は図で文字P、P'で示されている。確率誤差の3倍は2倍の標準偏差とほぼ同じような面積を含む。

我々は次のような一般的方法を設ける事が出来る。

- (1) 平均値や他の統計量の計算値と真値は標準誤差の2倍以上の差がある事は起りそうにない。
- (2) もしある目的のための実験が2倍の標準誤差以上に期待値から離れている結果を導いたら、それを任意標本として取扱えないと思わなければならない。

§5

我々の問題は限られたデーターから計算した統計的尺度がその期待値のまわりにどのように変化するかを考えることである。二つの処理法が考えられる。§2で述べたように多数回の実験を行う事—あるいは多数の標本を集める事—が出来る。そして各々に対して同様の統計的尺度を計算する。処置法は一般にかなりの仕事であるので我々は別の処理法をとろう。確率論に基礎をおく代数的解析は標本抽出過程が無限に繰返之されれば標準誤差は極限として求められ、そのために凡ての標本がその期待される割合の中に含まれているような標準誤差の決定を可能にする。更に遠くへ行つて標本数が増加するにつれてある統計的尺度の度数分布が狭いような実際の曲線をしばしば決定することが出来る。

次に簡単な説明をしよう。度数(n)の標準誤差、すなわち生起確率がPで不生起確率がqのm回の独立試行で起る現象をもつ度数nの標準誤差を求めよう。m, m-1, m-2, ..., 3, 2, 1, 0 に等しいnの確率は二項展開の各項で与えられる。

このmによつて表わされる点の周りのモーメントを使えば一次モーメントは次の通りである。

$$mP^{m-1}q + m(m-1)P^{m-2}q^2 + \dots + mP^m = mP(P+q)^{m-1} = mP$$

同じ点の周りの二次モーメントは

$$mP^{m-1}q + 2m(m-1)P^{m-2}q^2 + \frac{3}{2}m(m-1)(m-2)P^{m-3}q^3 + \dots + m^2P^m$$

$$\begin{aligned}
&= m p^{m-1} q + m(m-1) p^{m-2} q^2 + \frac{m(m-1)(m-2)}{2!} p^{m-3} q^3 + \dots + m q^m \\
&+ m(m-1) p^{m-2} q^2 + m(m-1)(m-2) p^{m-3} q^3 + \dots + m(m-1) q^2 \\
&= m q + m(m-1) q^2
\end{aligned}$$

従つて平均値周りの二次モーメントは

$$m q + m(m-1) q^2 - m^2 q^2 = m p q$$

$$\therefore \text{標準誤差} = \sqrt{\mu_2} = \sqrt{(m p q)}$$

もし m 回の独立試行を繰返せば生起回数の観測度数 n は $\sqrt{(m p q)}$ の標準誤差をもつて期待値 $m p$ の周りに変化する筈であると言う事が分る。

§ 6

次にこの結果を二三の例に適用してみよう。

- (a) 男子が生れる数と女子が生れる数は 1050:1000 である事が認められている。換言すれば子供が男である確率は $\frac{1050}{2050}$ である。もしある地方で 100,000 人の子供の中 51,350 人が男である事がわかつていれば普通の確率論から変動に關するどの理論をこの統計量に基礎づけても安全であろうか？男子出生の期待値は 51,220 人であり 標準誤差は

$$\sqrt{\left(100,000 \cdot \frac{1050}{2050} \cdot \frac{1000}{2050}\right)} = \pm 158.07$$

である。

実際の場合と期待値との差は 130 人でこれはこれは標準誤差よりも小さいのでこの結果からせれると言う事ははつきり結論づける事は出来ない。

- (b) 場合の数が 10,000,000 人で実際の男子出生数が 5,135,000 人であればそのとき標準誤差は 1580.7 で期待値と実際の値の差は

13,000 である。したがつて比 1050:1000 は特別の場合に適合しないと言う結論に対して充分な証明になる筈である。

- (c) 年間死亡確率が 0.007 であれば 200 件の確率誤差は $0.67449 \sqrt{(200 \times 0.007 \times 0.993)} = 80$ である。だからある 1 年間の危険について 200 件のうちに 1 年間に期待される死亡数として 1.4 の代りに 2.2 とすれば非常時のための生命保険料の割増に対してほぼ近似的なものが得られるだろう。この目的に対しては確率誤差が標準誤差よりもすぐれているとおもう。すなわち、場合の数は略 200 件であり、年間死亡確率として 0.007 であるような平均年々であると言う仮定の上で保険期間に対する非常時割増しいくつかのアイデアを得るために死亡率として 0.007 の代りに 0.011 を取扱う事は合理的であろう。その上あたかも各年々級がその級独自の死亡率に従っているかの如く各級を取扱う事が正しいと仮定されて来たし、そして事業の余剰金に独立に取扱わなければならない。しかし、これは少し異論のある点である。
- (d) m が定数であれば $\sqrt{m p q}$ は $p = q = \frac{1}{2}$ のときに最大値をもっている事は注意されたい。そしてこの事は保険会社は次の事を一般に見出す事を示している。すなわちもし保険会社が同じ大きさの二つの階級を持つていて、その一方が他方よりも高い死亡率に従つていれば前者は期待請求数から実際に大きな偏差をもつ筈である。何故かと言うと 1 年間だけの死亡確率は死亡率表の最後で値 1 になるからである。

§ 7

さて S 番目のグループに入っている割合が P_s であるように度数分布を k 個のグループに分ける事を考えよう。明らかに $P_1 + P_2 + \dots + P_k = 1$ である。もしこの分布から無作意に一つの場合をとるとそれが S 番目のグループから抜かれる機会が P_s であり、他のグループからとられる機会は $1 - P_s = q_s$ である。今 n 個の場合が無作意にとられ、それらのうち $n P_s$ が S 番目の

グループ^oに落ちているとしよう。そうすると n_s の期待値が mP_s であるからこの度数は二項定理 $(P_s + Q_s)^m$ の各項で与えられる確率をもつて $m, m-1, \dots, 2, 1, 0$ の値をとる。 n_s の標準誤差は

$$\sigma_{n_s} = \sqrt{mP_s(1-P_s)} \quad \dots\dots\dots (1)$$

である。

実際に、与えられた標本から度数分布の正確な型を知らないから $P_s = \frac{n_s}{m}$ とおく事によつて標準誤差に近似してもよい。これは s 番目のグループにある標本の観察割合であるから近似式として次をうる

$$\sigma_{n_s} = \sqrt{\left\{ n_s \left(1 - \frac{n_s}{m} \right) \right\}} \quad \dots\dots\dots (2)$$

§ 8

凡ての度数 $n_1, n_2, \dots, n_k$ の和は m であるからその結果次の事が言える。すなわち特別の標本で n_s が mP_s より非常に大きければ他の度数は平均的に小さくしなければならない。そしてこの事は各グループ間の誤差は相関がある事を示している。次に研究すべき点は s 番目のグループ^oと t 番目のグループ^oの度数の相関の間の相関の量である。

n_s についてその期待値からの偏差 δn_s は $n_s - mP_s$ である。 n_s と n_t の偏差間の関係を考へているのだから残つた $(k-2)$ 個のグループを n_R とする一つのグループにまとめた組を作るのが便利である。そうすると

$$n_s + n_t + n_R = m$$

$$P_s + P_t + P_R = 1$$

$$\delta n_s + \delta n_t + \delta n_R = 0$$

そして

$$(\delta n_s + \delta n_t)^2 = (-\delta n_R)^2$$

あるいは

$$\delta n_s \delta n_t = \frac{1}{2} (\delta n_R^2 - \delta n_s^2 - \delta n_t^2)$$

次にもし任意標本の非常に大きな数 N がとられてそれについて上の最後の方程式の両辺の式を加えて全数 N で割ると考えれば

$$\frac{1}{N} S(\delta n_s \delta n_t) = \frac{1}{2} \left\{ \frac{1}{N} S(\delta n_R^2) - \frac{1}{N} S(\delta n_s^2) - \frac{1}{N} S(\delta n_t^2) \right\}$$

右辺の表わす式は極限において上の方程式 (1) に与えられているグループ度数の平方標準誤差を表わしている。故に極限において

$$\frac{1}{N} S(\delta n_s \delta n_t) = \frac{m}{2} \{ P_R(1-P_R) - P_s(1-P_s) - P_t(1-P_t) \}$$

$$= \frac{1}{2} m \{ (1-P_s-P_t)(P_s+P_t) - P_s(1-P_s) - P_t(1-P_t) \}$$

$$= -mP_sP_t \quad \dots\dots\dots (3)$$

n_s と n_t の間の相関は

$$r_{n_s n_t} = \frac{\frac{1}{N} S(\delta n_s \delta n_t) \text{ の極限}}{\sigma_{n_s} \sigma_{n_t}}$$

$$= - \frac{P_s P_t}{\sqrt{\{ P_s(1-P_s) P_t(1-P_t) \}}}$$

$$= - \sqrt{\left\{ \frac{P_s P_t}{(1-P_s)(1-P_t)} \right\}} \quad \dots\dots\dots (4)$$

更に P_s と P_t に対して標本の比例度数 $\frac{n_s}{m}, \frac{n_t}{m}$ をおく事によりこの式に近似する事が出来る。

§ 9

m 個の標本からその平均値の標準誤差を見出す事

前の度数分布を k 個のグループに分ける事を仮定する。ここぞ x は s 番目のグループ^oに属した変量 x の値である。本章の始めの節で既に説明した理由のために (i) 度数分布によつて表わされる集団平均値 すなわち

$$\bar{x} = \sum (P_s x_s)$$

と(ii)この母集団から任意に抽かれた m 個の特別の標本から計算した平均値 すなわち

$$\bar{x} = \sum \left(\frac{n_s}{m} x_s \right)$$

の間には区別しなければならない。ここで最初の $\sum$ は各グループの凡てについての和を示す。

$\bar{x}$ の標準誤差すなわち $\sigma_{\bar{x}}$ は標本平均値が母集団平均値と異なるための広がり の尺度を示すであろう。 $\sigma_{\bar{x}}$ の値は前節(1)と(3)の結果を用いて求められる。

同じ記号を使って

$$\begin{aligned} s\bar{x} &= \bar{x} - \bar{x} \\ &= \frac{1}{m} \sum (n_s - m P_s) x_s \\ &= \frac{1}{m} \sum (\delta n_s x_s) \end{aligned}$$

δn_s の期待値は0であるから $\delta \bar{x}$ の期待値は0である。あるいは標本平均の繰返しサンプリングから求められる平均値は母集団平均値に同じものである。上式の両辺を二乗して

$$(\delta \bar{x})^2 = \frac{1}{m^2} \left\{ \sum (\delta n_s^2 x_s^2) + 2 \sum' (\delta n_s \delta n_t x_s x_t) \right\}$$

をうる。ここで $\sum'$ は S 以外の S と t の凡ての対に対する和を数えわたす。

もし始めに取られる標本を大きさ N と仮定し $(\delta \bar{x})^2$ の対応値を加え、その結果を N で割れば次式をうる。

$$\frac{1}{N} S(\delta \bar{x})^2 = \frac{1}{m^2} \left\{ \sum \left(x_s^2 \frac{1}{N} S(\delta n_s^2) \right) + 2 \sum' \left(x_s x_t \frac{1}{N} S(\delta n_s \delta n_t) \right) \right\}$$

ここで S は N 標本についての和を表わしている。この式の左辺は標本平均の平方標準誤差 $\sigma_{\bar{x}}^2$ である。右辺の $\frac{1}{N} S(\delta n_s^2)$ は方

式(1)の $\sigma_{n_s}^2$ である。また $\frac{1}{N} S(\delta n_s \delta n_t)$ は方程式(3)で与えられている。故に

$$\begin{aligned} \sigma_{\bar{x}}^2 &= m^2 \left\{ \sum \left[x_s^2 m P_s (1 - P_s) \right] - 2 \sum' (x_s x_t m P_s P_t) \right\} \\ &= \frac{1}{m} \sum (P_s x_s^2) - \frac{1}{m} \left\{ \sum (P_s x_s) \right\}^2 \\ &= \frac{1}{m} (\mu_2' - \bar{x}^2) \end{aligned}$$

ここで μ_2' は母集団分布の x に対するの原点周りの二次モーメントである。しかる $\mu_2' - \bar{x}^2 = \sigma_x^2$ であるから

$$\sigma_{\bar{x}} = \sigma_x / \sqrt{m} \quad \text{----- (5)}$$

従つて平均値の標準誤差は母集団標準偏差に対する標本サイズ m の平方根の比である事がわかる。

§10

この最後の結果は統計的仕事によく使われている。大数が観察されし平均値は他の同じような特別の実験の比較のために使われている。平均値間の実際の差は無作為抽出の代に何が正当な理由があるのだろうか？ 実際的な応用は数年間にわたる色々な会社の平均利益の比較でもよからう。多年間の利益の標準誤差は平均値周りの二次モーメントの平方根を年数の平方根で割つて得られる。その商は(5)の $\sigma_{\bar{x}}$ を与えるであろう。事業のある部分での平均利益が低いのは偶然に支配されるか、あるいはそこに除去する必要があるいくつかの要因のためだと言えるのは標準誤差を使つてゐるからに過ぎない(標準誤差は標準偏差から推定出来る)

§11

本章§5で注意した事であるが我々は統計的尺度の度数分布が近づいて行く実際の曲線を決定する事が出来る。先に4章で述べ

β_2 は平均値と共にそれが Pearson の曲線系であれば度数曲線を指定するのに用いられる事を導いた。それは次のようである。即ちもし、統計的尺度の度数分布に対して β_1 と β_2 を見付ける事が出来れば曲線の型を決定するのにずっと前の長つたらしい方法を使うだろう。もし $\beta_1(x)$ と $\beta_2(x)$ を x の母集団分布に対するモーメント比とし $\beta_1(\bar{x})$ と $\beta_2(\bar{x})$ は大きさ m の繰返し標本の標本平均値 $\bar{x}$ の分布に対するモーメント比とすれば次の事が示される (R. H. Anderson, J. Inst. Actu. XLII. 429 参照) すなわち

$$\left. \begin{aligned} \beta_1(\bar{x}) &= \beta_1(x)/m \\ \beta_2(\bar{x}) &= 3 + \{\beta_2(x) - 3\}/m \end{aligned} \right\} \dots \dots (6)$$

もし $\bar{x}$ の分布が $\beta_1(x) = 0$, $\beta_2(x) = 3$ の正規曲線によつて表わされていれば

$$\beta_1(\bar{x}) = 0, \quad \beta_2(\bar{x}) = 3$$

であつて $\bar{x}$ の分布もまた正規であると言う事が分る。

たとへ $\bar{x}$ の分布が正規でなくても方程式 (5) から $\beta_1(\bar{x})$ が 0 に近似し m が小さくなければ $\beta_2(\bar{x})$ が近似的に 3 になることが分る。

§.12

標準偏差の標準誤差は大標本に対しては $\frac{\sigma_x}{2} \sqrt{\frac{\beta_2(x)-3}{m}}$ として取られる。そして母集団分布が近似的に誤差の正規曲線 ($\beta_2(x) = 3$ のとき) であれば標準誤差は $\frac{\sigma_x}{\sqrt{m}}$ になる。その他、標準誤差は 2 つの百分率あるいは割合の間の差に関係させて有用である。従つてもし m_1 回の試行を行えば n_1 回の事象が起りまたこれとは独立に m_2 回の試行で n_2 回が起つていればどういふ事情の下で $P_1 = P_2$ と結論出来るだろうか？ ここに P_1 の標本推定は $\frac{n_1}{m_1}$ であり P_2 の標本推定値は $\frac{n_2}{m_2}$ である。この解法は二つの死亡率や契約脱退率、罹病率を比較するときに役立つだろう。

もし $P_1 = P_2 = P$ ならば差 $n_1/m_1 - n_2/m_2$ の標準誤差は $\sqrt{\{P(1-P)(\frac{1}{m_1} + \frac{1}{m_2})\}}$ である。我々の実験からの P_1 と P_2 は近

似値であるがこれに基づく割合 P については全く未知であるがそこに共通値があると言う仮定から

$$(n_1 + n_2) / (m_1 + m_2)$$

をその推定値にとる事が出来る。

これは次の標準誤差を導く。

例として (1) 22 人の脱退者をもつ 1000 件 (2) 19 人の脱退者をもつ 600 件をとらう。前者は 0.0220 の脱退率を与え后者は 0.0317 の脱退率を与えている。この差 $0.0317 - 0.0220 = 0.0097$ は有意であろうか？ 二つの経験表を組合せると脱退率として $41/1600$ あるいは 0.0256 を与える。最後の式で与えられる標準誤差は 0.0082 であるこの差は有意ではない。しかしながらもし件数をどれも 3 倍以上にとれば標準誤差は 0.005 になるだろう。そしてこの差が有意であることはもう説明の必要がないだろう。

§.13

同じような方法でモーメントや度数の標準誤差をみつける事が出来る。しかしこれは本章で取扱わない方がよいような理論的部分を導く。しかし統計的仕事でその函数が重要なために、相関係数の標準誤差に注意する必要がある。平均値の場合のように母相関係数に ρ の記号を用い、観測値の m 対の標本から計算した相関係数に対して r とは異なる r を用いければ混乱をさける事が出来る。標本相関係数 r は ρ の周りに変化する。そして r の標準誤差は大標本 m に対しては近似的に

$$\sigma_r = (1 - \rho^2) / \sqrt{m-1} \dots \dots \dots 7$$

m が充分大きいときは、ここぞ $\sqrt{m-1}$ の代りに $\sqrt{m}$ を用いる事が出来て同じ結果になる

(1) R. A. Fisher は幾何学的考察から直接に積分して r の標本分布の式として次の式を得た $f(r) = \frac{(1-\rho^2)^{\frac{m-2}{2}}}{\pi(m-3)!} (1-r^2)^{\frac{m-4}{2}} \frac{d^{m-2}}{d(r\rho)^{m-2}} \left\{ \frac{\cos^{-1}(-r\rho)}{\sqrt{(1-r^2\rho^2)}} \right\}$

である。もし ρ の値が未知ならその近似値として γ を用いるべきである。

5/4

上の結果は始め K. Pearson と L. N. G. Filon に よつて m が大きいとき近似的に $\sqrt{m}$ をもつ形で与えられている (Philos Trans. A. CXCI. 231~41)

後に R. A. Fisher は標本が 1/3 以上の正規相関曲面に従う母集団から抽かれたとき r の正確な分布を得ている (Biometrika X 507~521) (4) このような詳しさを γ の標本分布の形と同じように公式 (7) による正確な近似値は Tables for Statisticians Part II. Table XXXII あるいは Biometrika XI 328 頁以後に与えられている表によつて研究出来る。 $\rho=0$ なら公式 (7) は m のどんな小さな値にも σ_r の良い値を与える事がこれらの表から分る。しかし ρ が大きくなるにつれてこの近似値はあまり満足なものとはいえない。何故なら公式は正確な値を与えないためでありそしてたとえ σ_r が正確に求められていても γ の分布は ± 1 の偏差が等しくないと言う事と標準誤差の 2 倍が殆ど全面積をカバーすると言う普通の法則が適用できないためである。これ以上明確な説明は困難であるが $m > 400$ なら公式 (7) が用いられると言う事は役に立つ。* しかし $m=100$ まで ρ が 0.5 より小さくなければまた $m=50$ まで ρ が 0.3 より小さくなければ σ_r の解釈に注意が必要である

R. A. Fisher は有用な変換

$$\frac{1}{2} \{ \log_e (1+r) - \log_e (1-r) \}$$

を提案している (Metron I. 1921) これは ρ の値の如何に拘り (前頁下へ)

* $m=400$, $\rho=0.9$ に対して $\sigma_r=0.00957$ をうる。公式 (7) からの σ_r は 0.00951 である γ の分布は 平均 $\gamma=0.8998$, モード $=0.9011$, $\beta_1=0.07402$, $\beta_2=3.1342$ によって表わされる。これは正規曲線によって概略を表わされるにすぎない。

ら標準誤差 $\frac{1}{\sqrt{m-3}}$ をもつて正規分布するような変換である。")

5/5

公式 (7) の応用として匪章の例をとろう。

そこでは養老保険の満期年令と未満期々間の間の相関係数が 0.254 であつた事を知っている。しかしこの相関係数が両者間の相関々係を正確に表わしていると主張する事は正しくない。真の相関係数はこれより大きいかも知れない。あるいは小さいかもしれない。そして γ で例示したような考えを持ち上げる。しかしそこには相関係数に関して別の問題があつた —

我々は標準誤差を調べるまでは γ にどんな真の関係があると言う事は出来ない。

我々の例では $m=2870$, $\gamma=0.254$ であつたから $\sigma_r = \pm 0.016$ である。故にこの場合は標準誤差が非常に小さいのでこの結果が信頼しうるのである。しかし、もし 0.05 の標準誤差をもつ $\gamma=0.073$ を求められたら相関は単に偶然から起つたものでないと断言することは不可能であらう。

5/6

この事は公式 (7) で m が 30 位の小さいときでも安全に使えると

この変換は Z 変換とよばれている。

$$Z = \frac{1}{2} \log_e \frac{1+r}{1-r} \quad \gamma = \frac{1}{2} \log_e \frac{1+P}{1-P} \quad \text{とおけば } m \text{ が充分大き}$$

いときには Z の分布は近似的に

$$E(Z) = \gamma + \frac{P}{2(m-1)} \left\{ 1 + \frac{1+P^2}{8(m-1)} \right\} = \gamma$$

$$V(Z) = \frac{1}{m-1} \left\{ 1 + \frac{4-P^2}{2(m-1)} \right\} = \frac{1}{m-3} \quad (\rho=0 \text{ として})$$

なる正規分布に従つて分布する。これは相関係数の近似的検正法特に標本相関係数の差の検定に便利である。詳しくは M. G. Kendall 前掲文献 (I) 11 章参照の事。普通標本抽出に対してはスネデカー 統計的方法 (岩波書店) (E) オフ章にある図によつて充分正確に一方を他方に変換出来る。

る標準誤差の重要な応用を教えている。もしそこにどんな相関もなければ $\rho=0$ であり(?) 或は次の称になる。

$$\sigma_r = \frac{1}{\sqrt{n-1}} \quad \dots \dots \dots (8)$$

従つて四章の例にもどつて そこに相関がないと仮定すれば $\frac{1}{\sqrt{2869}}$ 或は 0.0187 の標準誤差をもつて $r - \rho = 0.254$ である。この差 $r - \rho$ は標準誤差の 12 倍以上である。故に 2870 の標本が抽かれたと考へている母集団の相関は 0 であるという事は殆ど不可能である。

§ 17

上の公式(7)は簡単に述べた方法で計算された相関係数にだけ適当なものである。四重分割表を使った場合の標準誤差はグループ分けが極端であるためにその期待値より大きい。また、厳密に計算した公式によれば完全な値になる筈である。次の公式は r の標準誤差を与えるために参考になる。

$$\frac{1}{2\sqrt{N}} \sqrt{\left\{ \frac{(a+d)(c+b)}{4N^2} + \psi_2^2 \frac{(a+c)(d+b)}{N^2} + \psi_1^2 \frac{(a+b)(d+c)}{N^2} \right.}$$

$$\left. + 2\psi_1\psi_2 \frac{ad-bc}{N^2} - \psi_2 \frac{ab-cd}{N^2} - \psi_1 \frac{ac-bd}{N^2} \right\}}$$

$$\text{そこで } \chi = \frac{1}{2\sqrt{N(1-r^2)}} e^{-\frac{(R^2+R^2-2Rr)/2}{(1-r^2)}}$$

$$\psi_1 = \frac{1}{\sqrt{2\pi}} \int_0^{\frac{R-rR}{\sqrt{1-r^2}}} e^{-\frac{1}{2}x^2} dx$$

$$\psi_2 = \frac{1}{\sqrt{2\pi}} \int_0^{\frac{R-rR}{\sqrt{1-r^2}}} e^{-\frac{1}{2}x^2} dx$$

そして四重分割表は $a+c > b+d$ 、 $a+b > c+d$ のように配列されている事が仮定されている。ここに a, b, c, d は 187頁の図で示した意味を持っている。18章の例ごとの標準誤差を求める数値例は次の通りである。

$$\psi_1 = \frac{1}{\sqrt{2\pi}} \int_0^{\frac{R-rR}{\sqrt{1-r^2}}} e^{-\frac{1}{2}x^2} dx = \frac{1}{\sqrt{2\pi}} \int_0^{0.56821} e^{-\frac{1}{2}x^2} dx = 0.21585$$

$$\psi_2 = \frac{1}{\sqrt{2\pi}} \int_0^{\frac{R-rR}{\sqrt{1-r^2}}} e^{-\frac{1}{2}x^2} dx = \frac{1}{\sqrt{2\pi}} \int_0^{0.32230} e^{-\frac{1}{2}x^2} dx = 0.12639$$

これらの値は *Tables for Statisticians Part I Table II* を用いて求められる。

$$\chi = \frac{1}{2\sqrt{N(1-r^2)}} e^{-\frac{(R^2+R^2-2Rr)/2}{(1-r^2)}} = \frac{1}{2\sqrt{N} \times 0.67900} e^{-0.66744}$$

$$= 0.10462$$

$$\therefore \log \frac{1}{\chi} = 0.98039 \quad \log \psi_1 = \bar{1}.33254, \quad \log \psi_2 = \bar{1}.10171$$

$\therefore r$ の標準誤差は

$$\frac{1}{0.10462 \sqrt{4703}} \left\{ 0.002283 + 0.00145 + 0.00479 + 0.00252 - 0.00408 \right\}$$

$$= \pm 0.018 \quad \quad \quad - 0.01615$$

§ 18

この方法で計算した標準誤差は公式(7)で計算した結果より多くの場合3倍程大きい値を示す。現に我々の例でもそうである。正確な公式は面倒であるが *Tables for Statisticians Part I Tables XXIII XXIV* を概算に用いて数値計算を少なく出来る。この近似法は相関表の区割が極端に違ふ時以外は安心して使う事が出来る。

§ 19

我々が予想したように標準誤差のどの式も分母に場合の数の平方根を含んでいる事は注意すべきである。本章の始めで論じた標準誤差は場合の数が増加するにつれて減少する。そしてまた各々の場合で討論された事であるが標準誤差は場合の数の平方根に逆比例する事ができる。學生は標準誤差を計算するのに習熟すべきであつて *Tables for Statisticians* についている普通の「標準誤差」の表を使

えば多くの計算がたしかる事を発見するだろう。

標準誤差を計算する目的は計算した平均値や他の尺度から余り多くの判断をしないようにするためである。しかし極端に逆の方向に陥ってはいけないうし、また理論的な正当性以上に標準誤差に拘りて信頼をしてはならない。

だから我々の理論は特性値が誤差の正規曲線に似た型で分布している事がある点で仮定されているしまた既に示された証明の大部分は場合の数が30以上のときあるいはある種の特性値にとつてはそれ以下の場合でもこの仮定が適当なものである事を示している。この仮定には十と一の誤差が多分等しい筈だと言う事が暗に含まれている。しかし丁型分布の標本平均値は真平均値の上と下に二倍の標準誤差以内に等しいように分布していると主張する事は正しくないであろう。公式(6)はこの上下の限界を示すための助けとなる筈である。

§20.

次にサンプリングでの二三の実際的な点に関して簡単に考えよう。サンプリングの本質は標本を調べる事によつて全体を評価する事であり①また誤差は標本を作る場合の偏りのために生ずるか②標本が少ない数から作られているために全体とかなりちがっている特定の標本のために起るのである。

保険数理や社会学上の応用での偏りを防ぐ事は普通には困難なことではない。例えば生命保険のついた人の死亡率の推定値を求めたいならAで始まる名前の人に関する情報を集めるであろう。これは二三の場合を与えるだろうが、そこにはこのような人がアルプベットの他の文字から始まる名前の人と違ふのだと疑う事の根拠もない。特別の文字を選ぶ事は *Alsace-Lorraine* 地がのように混合社会その人種の問題が入るなら疑念を深くかもしれ

(1) フランス北東部の地方。その *alsace* と *Lorraine* の両州を含む。

1871~1919及び1940~44年間はドイツ領

ない。例えば *Alsace-Lorraine* でWで始まる名前の人を取り上げればフランス系統の人を除きドイツ系統の人を含むだろう。もう一つの方法は100人のうちから一人を抜くと言う場合である。たとえば生命保険の死亡率は保険会社の登録簿から各100番目の場合を調べる事によつて研究される。

この種のサンプリングは社会学研究で有用である。学徒の家庭条件を調べたいと思つても、衛生の項目、住居あるいは居住の習慣を全家庭から得る事は望めない。しかし2000や3000の場合の莫大な調査を得る事は出来る。

勝手気ままに任意標本を得る事は困難ではないしまた僅かであるがそこに必要なものは常識的なものである。

我々が小さな標本を用いているに過ぎないと言う事実には他の誤差の危険がある。ここでの問題は標準誤差に関するものである。標本が無作為に選ばれた事と標本それ自身は小さくなくてそれが抽かれた母集団に比して小さいと仮定出来れば、本章の始めで述べた方法に従う事が出来る。

§21

いくつかの実験には特殊事情が起る。その一つの型について特に注目しよう。それは植物の異品種の収獲量比較試験にしばしば必要である。そのような場合の面倒なことは小さな圃場でさえも個々別々の置き場所が大きく違つた結果を導くと言う事のみならず小さな接近したプロットが互に類似すると言う事である。だから公平に比較するためには相隣接プロットの対を作らねばならない。比較は各対の間で行なわれこれらの対の間の差に関係があるのだから我々は

$$s^2 = \frac{S'(y-x)^2}{m}$$

を計算しなければならない。ここに m は対の数であり x, y はそれぞれの平均値からはかられた対の対応するものである。

隣接プロットから作られた対が単位になつてこの種の場合

は、 σ_1 なる標準偏差をもつ母集団記録から m_1 観察値を抜き、標準偏差が σ_2 である場合の他の記録から m_2 ケのオス標本を抜く場合とは区別することが大切である。この場合二つの平均値間の差の標準偏差は

$$\sigma^2 = \frac{\sigma_1^2}{m_1} + \frac{\sigma_2^2}{m_2}$$

で与えられる。

これは変量間に相関のない事が仮定されているが“対”にした場合の問題では相関が期待できるから“対”をとっている。代数的に相関は $S(y-x)^2 = S(y^2 - 2xy + x^2)$ の xy の項によって示される。

標本から計算された σ は母集団の標準偏差として扱えると言う仮定を証明するために標本は充分大きい事が仮定されている事は本章の他の処で与えられた事から分るだろう。

不適當な公式を用いる事は間違つた結論を導くかも知れない。平均値間の実際の差は 30 ぞ $\sigma^2 = S(y-x)^2/m$ によつての標準誤差は 6 であつても $\sigma^2 = \frac{\sigma_1^2}{m_1} + \frac{\sigma_2^2}{m_2}$ では 12 である。前者で判定すれば差は殆ど有意であるが、後者で判定すれば疑わしいのである。

今述べた種類の問題は条件を変えた方法での結果の比較が必要であるときはいつでも起り得る問題である。また収量試験のために初めになされた理論は他の処で貴重な事が分るでしょう。

第 XI 章 適合度の検定

§1

度数曲線の色々な型の例で縦線の値や面積が計算されているとき、得られた結果が適當であつたかどうかを確かめるために補整法を決定する系統的企画は何も作られていない。

一般に係数数理士は實際上の表の補整値に責任を課して行ふのが普通であつたし、概略のチェックを行えば色々のグループの合計を比較出来たし、また補整値と未補整値間の差における符号の変化もみつけられた。適合度の問題はもつと正確な取扱いが必要である。すなわち検定のために各グループに対する標準誤差の計算によつて助成したときでも、ある差が大きいと言えるにすぎない。また標準誤差が二三の場合にある限度を超えていればそれは補整の残りでどんな方法で等しく釣合せられるかどうかと言う事は不可能である。全補整値で判定されるような不適合のある尺度を与えるような検定が必要である。

§2

今 m ケのグループに分けられた N ケの観察値があり、各グループの中に $m'_1, m'_2, \dots, m'_n$ ケが入つていれば、我々は系列 $m'_1, m'_2, \dots, m'_n$ が正しい補整であるときに決定を可能にする一つの判定基準を求めねばならない。観察値 (m') が任意標本抽出で期待される偏差以上理論値 (m) から違わない補整が正しいと考えるのは明らかである。我々が知る必要があるのは m が理論値を表すならば m' の特別な系列の起る確率ではなくて m' の系列あるいはこれに似たもの、あるいは少くとも同じような系列の起る確率である。この問題の困難を認識するために貨幣投げの最も簡単な場合を考えよう。その実験で貨幣は 6 回投げられて“表”が 4 回“裏”が 2 回現れたと仮定する。我々が作る“補整”は表が 3 回、裏が 3 回である。検定のために観察結果よりも起りそうになり結果の得られる確率を求める必要がある。この確率は次の結果

のどれか一つが起る確率と同じものがある。

表	裏
6	0
5	1
4	2
3	3
2	4
1	5
0	6

とにかく偏差を算くかんたんな確率が分つてゐる時でさえも最も簡単な場合を除いてはこのような確率を直接計算する事は不可能である。しかしかんたんな確率を知らないときや複雑な場合には可能な場合の先験的推論は実際に起つた場合の推論よりもいくらか大きいとか小さいとかと言う事からは説明出来ないためにもつと困難が導入される。例えば20ケのサイコロが投げられた時に“6”の目が10ケあるいはそれ以上出る確率は6の目が2つあるいはそれ以下得る確率よりも大きいかどうかは大変な数値計算をしないう事は困難であろう。しかしこれは数の系列に関する偏差を考えねばならない場合での一般的定理に比して非常に簡単な場合である。

53

この目的に内するどんな測定値にも平均値からの偏差が誤差の正規曲線の形をとる事が仮定されている。またある値(たとえば σ という値)よりも大きい偏差を得る機会を推定する必要があるれば平均値の両側に σ まで正規曲線の凡ての値を加える必要がある。すなわち我々は次式

$$\int_{-\infty}^{-t} e^{-\frac{1}{2}x^2} dx + \int_t^{\infty} e^{-\frac{1}{2}x^2} dx = 2 \int_t^{\infty} e^{-\frac{1}{2}x^2} dx$$

を計算しこれを全曲線の面積すなわち全偏差で割らねばならない。今二つの測定値(例えばちがつた年々での危険率)があると考へ

るは偏差はいわば2方向にある。そしてこれを表すためには二つの変数をもつ式をとることが必要である。その式は正規曲線に類似の相関曲面である

$$Z = Z_0 e^{-\frac{1}{2} \left\{ \frac{x^2}{\sigma_1^2} - \frac{2rxy}{\sigma_1\sigma_2} + \frac{y^2}{\sigma_2^2} \right\} / (1-r^2)}$$

をとればよい。この式は既に取扱つたものである。積分は両変数に向して σ から σ' まで実行され全体に比較されねばならない。もし n ケの測定物があれば n 変数の函数を扱う必要がある。これは読者に数学的見地から問題のざつとしたアイデアを与えるであろう。そして n 重複分の二つの商は確率を与えると期待してよい事を示している。次の段階はこれらの n 重積分を普通の積分型にまどすことである。その結果は

$$P = \frac{\int_0^{\infty} e^{-\frac{1}{2}x^2} x^{n-1} dx}{\int_0^{\infty} e^{-\frac{1}{2}x^2} x^{n-1} dx} \quad *$$

になる事が示されている*

* これは K. Pearson の論文 "On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from Random Sampling" Phil. Mag. July 1900. にある。その短い証明は H. E. Soper によつて Frequency Arrays に与えられている

3から30までの $n' = n+1$ の凡ての値に対する P の表は1から30までの x^2 の値に対応している。いくつかの加算定数と一緒にこれ以上の値を計算するための補助表は Tables for statisticians Part I にあるその短縮した表が附録 IX にある。

又はこの式が展開された事から変数に依存する複合函数を表わしていて特別の分布の確率で示された位置を表わしその補正に対する検定に必要である。

§ 4

確率Pの尺度を得る前に補正された統計量からεに対する値を求めねばならない。既に扱った関係論文にその値が次式のようなもので与えられる事が示されている。

$$\chi^2 = \sum \left\{ \frac{(m_r - m'_r)^2}{m_r} \right\}$$

負の差を正の差と同じように扱うために二乗を用いている。しかし比はグループの大きさを計算に持ち込む必要がある。何故なら20のグループで15の誤差は非常に大きいけれども1000のグループでは無視しうるからである。

§ 5

適合度検定の実際面とその応用を次に扱う。それは次の8つの事項で述べられる

(1) 補正数値や未補正数値の表している事実がグループ単位でのみ得られるならば検定のための確率の値はグループの数が増加するにつれて一般に小さくなる。この実際的な補正が5年毎のグループで検定される事がしばしば起きるので注意しなければならない。しかし個々の年令に対する補正数値は条件に使われているけれども厳密に言えばそれらは補正値よりもよい値ではありえない。

(2) 検定は一つの分布を考える。そしてもし数値が縦線の系列であれば適當ではないだろう。しかし縦線の大きな数が系列として与えられていれば検定の応用に有望な美しい適合度のアイデアを支えるであろう。

(3) 経験表の最終数値は非常に小さく、また正確に適合する事

は失っていない。

このような場合には少なくとも度数単位に等しいような最終面積をカバーする最後の理論グループをとらねばならない。

(4) もし観察値がε倍されていれば偏差もまたε倍されており従って χ^2 の値もε倍されている。そして検定は適合が悪かった事を示す筈である。このことは始め不思議に思うかも知れないが僅かな観察でその合理的な事が示される。場合の数が大きいのは小さい場合よりもスムーズな系列を与えるから。全度数は異なるが同じ理論分布に従う二つの例での結果が比例的に同じであれば大きな度数での結果は小さな度数での結果よりも起りにくいことが分る。貨幣投げで表が2回表が2回出るのが理論的結果である。表が3回表が1回よりも悪いがあるいは同じように悪い結果の起る確率は0.625である。しかし3×2=6回が表で1×2=2回が裏であるよりも悪いがあるいは同じように悪い結果の起る確率は0.259である。

ある103480例にもとづいた分布があつてこの分布を示すためには合計が1000になるように数値を変えれば、そのときあたかも全度数が1000であつたとして検定された補正は、それが実際よりもよゝあてはまるとの印象を与える。

(5) 検定の応用で曲線が統計量に非常に厳密な適合を表しているとも取扱う数が多いときにはその確率はしばしば小さい事を私はいっている。この解釈は実際に取扱う統計量は大抵いつでも無関係の事柄のある量を含んでおりまた不平等性がデータの概略のために小さな経験をかきしてしまふためであると言えよう。観察された場合の数の増加は凸凹を除去はするが不平等性は依然として残る。曲線適合の立場からこれの意味する処は経験表は一つ以上の度数曲線によつて作られるのが本当である事を意味している。しかしある曲線は計算された曲線への近似がすぐれている。他に示される説明はこの問題の解決が偏差の正確な分布を与えない数学的公式の検定に立っている点でありそして大きな観察を行へばこ

の仮定の大体的性質は明かにされる。

(6) 良い適合から悪い適合の間で P の実際の値はどんなものであろうか？ そのような値を定める事は不可能である。我々は全表に対する距離の尺度を持っているにすぎないしもし補正に対する見込みが $20:1$ あるいは $30:1$ であれば結果は不満である。もしそれらが $10:1$ であれば補正は合理的なものとはならないが結果を捨てなければならぬときには正確な値を与えることは出来ない。しかしながら完全に確定した標準を固定出来るどんな検定を考える事も不可能であるからそこには特別の方法に反対しなければならぬと言う理由は何もない。何故ならそうする事ができないからである。

(7) 附加定数の導入は曲線の適合を必然的に改良するにちがいないと言う事はしばしば考えられるいくつかの場合にそのようにすることもあるが 10 ケの定数をもつ曲線を使う事も確に可能であるが 3 つ大しか定数を持たない他の曲線を使うよりも悪い結果を与える事もみられる。そのうえ下方補正 (under graduation) の可能性がある。従って P に対して高い値 (例えば 0.95 と言うような) をとる事を期待してはならない。貨幣投げに例をとれば一ケの貨幣を投げる実験が理論分布に近い分布を与える事は起りそうにないのである。だからその結果を得る確率もしくは誤る確率を推定しているのであればその確率に対して非常に高い値かもしくは非常に小さい値をうるのはきわめてまれである。

我々はしばしばそうするだろうがそれを期待してはならない。またある補正が P の非常に高い値あるいは非常に低い値を与える時の説明を探すのが賢明である。

(8) 簡単なためたか、もしくは保険数理士に *Makeham* の仮説⁽¹⁾ の採用をすすめる理由のために他のものよりも悪い一致を与える曲線を使う方がしばしば得策である

(1) 死亡の危険はまたは死力は一生を通じて定数、算術級数的増加、および幾何級数的増加の3要素よりなる事即ち $\mu_x = A + Hx + Bc^x$ を発見している。これを *Makeham* の法則と言う。

56 "On the Comparative Reserves of Life Assurance Companies etc" (J. Just. Actu. XXX VII 458~459) で *Ginge Kinge* は O^M 表に対して H^M 表モデルオフィスを用いる事は差支えないしまた、もし H^M が理論分布として取られるならば O^M 分布の確率が何であるかをみるために上に与えられた公式を応用する事は興味あるものであると注意している。下頁の表では 10 ケのグループがあり $\chi^2 = 1.79$ である $\chi^2 = 1$ と 2 のときに夫々 $P = 0.999438$ と 0.991468 が *Tables for Statisticians* に与えられている。しかしながらこれは 100 枚の新しい保険証券に対する検定のためには充分でない。 950 なら 0.05 に確率を定めるだろう。そしてこの事は O^M に示されていると同じ位大きい H^M 表からの偏差系を導く無作意標本である事は 20 回に 1 回だけだと言う事を意味している。この結果は調査した実際の数と考えるので百分率を取扱う危れのある学生は注意すべきである。 *Kinge* の他の表は 1 連した年令に従う保険証券) 彼の研究の中で最も重要な部分であるが 10000 例に対して $P = 0.831051$ のような厳密な一致を示している

年令群の 数	年令群に配列した証券発行 数		$O^M - H^M$		$(O^M - H^M)^2$ H^M
	H^M	O^M	+	-	
20	6.97	7.30	0.33	----	0.02
25	17.75	20.45	2.70	----	0.41
30	21.04	23.11	2.07	----	0.20
35	18.41	18.40	-----	0.01	0.00
40	13.82	13.05	-----	0.77	0.04
45	9.45	8.44	-----	1.01	0.11
50	6.23	5.07	-----	1.16	0.22
55	3.51	2.58	-----	0.93	0.25
60	1.97	1.20	-----	0.77	0.30
65	0.85	0.40	-----	0.45	0.24
	100.00	100.00	5.10	5.10	$\chi^2 = 1.79$

§ 7

本章の§ 2にもどろう。そこではこの問題の始まりは $n+1$ 個のグループに分けられた N 個の観測値があると言う事であった。我々は分配するための N 個のグループを選ぶ事ができる。何故ならこれら n 個のグループを固定すれば最後の一つは必然的に定められる。すなわち選択の自由はこの範囲に限定されるのである。またどんな問題でもグループの数が用いられる方法では選択の自由はできるだけ注意しなければならない。この事は既に与えられた諸公式を誘導した証明の中に暗に示されている。次にこの論述に続いて読者に同じ n 個の値を使って或る資料のⅠ型による補整とⅡ型による補整と比較してどちらがきれいに一致するかを尋ねることが出来る。そのとき前者を使おうとすれば4つの定数が必要であり、后者によって補整するためには3つの定数が必要である。読者は前者の場合には后者の場合よりも多く、選択の自由を制限つけるのが本当ではないかと尋ねるかも知れない。何故なら極端な場合をとれば4つのグループだけの分布をⅠ型を使って正確に再現出来るし、またⅡ型を用いてもそれを変えることができさらに選択の自由もあるかである。

§ 8

この質問を考える前に本章の前の諸節で“補整”という用語を紛らして来た二つの別個の問題を説明しよう。これらの問題は次のⅠ、Ⅱでまとめられる。

【Ⅰ】 現実の分布あるいはそれに等しいかもしくは少なくとも似たような分布をうる確率を調べるための理論分布が与えられる。

ここに n 個の貨幣が投げられたときの“表”の出る理論度数が現実分布と比較出来る例がある。

表の数 (1)	理論値 (2)	現実値 (3)	(2) - (3) (4)	(4) の二乗 (5)	(5)/(2) (6)
0	1	0	1	1	1.00
1	6	6	0	0	.00
2	15	12	3	9	.60
3	20	23	-3	9	.45
4	15	18	-3	9	.60
5	6	3	3	9	1.50
6	1	2	-1	1	1.00
計	64	64			$\chi^2 = 5.15$

$n' = 7$ $\chi^2 = 5.15$ のとき $p = 0.52$ である。

【Ⅱ】 偏差が理論分布から求められたものよりも大きいのか、あるいは同じである確率を調べるために現実分布の補整が与えられる。

この事は補整に用いた公式に含まれている定数の数に依存している。もしそこに r 個の定数があれば実際は $n' = n - 1$ に対して最後の例でやったようにグループの数から1を引く替りに定数の数 r を引くべきである。平均値は曲線の位置を固定するために必要でありそして一つの定数のように計算されねばならない。その結果もし正規曲線を使つたら3を引かねばならない。(すなわち合計に対して1つと平均に対して1つ、標準偏差に対して1つ) またⅡ型を使えば4を他の主要型*に対しては5を引かねばならない。一般的におえば計算に用いたモーメントの数が引かれていれば同じ結果が得られるのである。これは前節§ 7の最後の方で置った問題の理論的解答を与えている。

*Tables for Statisticians, Part I には $n' = n - 1$ であるから既に1が引かれている。従つて正規曲線が補整に用いられたときには $n' = n - 2$ が用いられる数である。

我々は定数を定めるためには3つのモーメントが必要であるにすぎないが4次モーメントを用いてⅢ型とⅣ型のどちらかを選ぶというようにこの法則を応用する際にモーメントの数を判断する事がいつでも可能だとは限らない。

更にⅠ型で曲線の出発点が固定されていれば3つのモーメントが必要だけである。一方範囲が固定されていれば2つのモーメントが用いられるにすぎない。(實際上定数の数は同じように減少している)4つの未修正モーメントを使ったⅠ型曲線によって強いて概略の補正をしようとするれば、Ⅴ章 §10で述べたように曲線のスタートは変り最後の補正は3つのモーメントを用いるだけである。定数の全部が仮定されていることと4つのモーメントが用いられている事は説明できる。

§9

§8の問題Ⅰのために与えられた例は下方補正について§5で注意した点を説明するのに使う事が出来る。特別の場合には理論分布に全く等しい実際分布を得ることも出来る。そのとき x^2 は0で $P=1$ である。

同じように殆ど不可能にみえる程理論から非常にかげ離れた分布に到達する事も出来る。これらのあるものは例外のようにもみえる。そしてもし充分長い実験を繰返せば P の凡ての値を与える分布を得られるだろう。同じような事は補正に関しても言える。すなわち曲線の正しい型を知つていても非常に小さいかまたは非常に1に近い P の値を求める事は見込みがないし且そのどちらも不可能である。

§10

同じ分布からの数種類の補正を比較する事だけを欲するときには x^2 を計算するに止められる。従つて色々の修正を使つてゐるⅠ型で補正を作るかあるいは同じ数の定数が使われているA型かB

型にそれらと比較すれば、 x^2 の最小値が最も厳密な補正を示している。定数の数が違ふときでも x^2 の値は補正が実際への接近する度合を示している。そしてこの事は保険数理のある種の仕事にとつて確率の研究よりも大切である。ある場合に自由度の数の説明が困難である事に注意すれば問題Ⅱに代つて問題Ⅰの解答を使う時に何が意味されているかを考えよう。(P. x^2)

検定の凡ての古い応用はこの場合になされて来たのである。このような事情の下で要するに我々は次のように言つてゐるのである。即ち、補正は必ずしも現実分布から得られたものではなくて一般論的推論あるいは他の明確な経験によつて得られた理論分布である。そして我々はその理論分布から離れる確率が現実分布から離れる確率より大きいがあるいは同じ位大きい確率を測つてゐるのである。

これらの節で説明された点は次の事を想起させるために述べられたのである。すなわち我々は基本理論によつて証明もされないし実際の仕事に必要でもないような補正の一般的な良い検定の改良を意味するものだとして解釈してはならない。

§11

ここで死亡率表の補正の検定を作ることを附記しよう。データ—は各年齢(あるいはグループの年齢)での“危険にさらされた”数(E_x)と“死亡数”(θ_x)を表わしている。死亡率の補正がなされて“期待死亡数”(θ'_x)は適当に補正された死亡率 q_x に E_x の値を掛けて計算される。

次に

$$\theta'_x = E_x - \theta_x, \quad \theta'_{x+1} = E_{x+1} - \theta_{x+1} \text{ 等々}$$

によつて補正系列

$$\theta_x, \quad E_x - \theta_x, \quad \theta_{x+1}, \quad E_{x+1} - \theta_{x+1} \text{ 等々}$$

をうる。

E_x は各計で固定されている。だから40才の場合には自由度

は80ではなくて40である。しかしでは凡ての80ヶの値から計算されねばならない。けれども E_x が θ_x に較べて大きいときにはほとんど凡ての年令において $E-\theta$ は0を与える。8の合計が8'の合計を正確に表す必要がない事に注意すればこの議論を理解する事は読者にとってたやすい事と思う。調整に用いた定数の数だけこの自由度40から引かねばならない。

第 XII 章

相関比—コンテングエンス

1. 回帰が直線であるとき、すなわち行(又は列)の平均値が直線に近似しているときは相関係数を合理的に使える事を知っている。しかしそうでない事状のもとで相関係数を使うと欠点を導きやすい。前章の方法はこの欠点を解放的でない事を述べた。先づ相関比 η として知られている函数を取扱う。これはある場合には有効な尺度である。

η_{yx} の値は次式で与えられる。

$$\eta_{yx}^2 = \frac{S\{x(\bar{y}_x - \bar{y})^2\}}{N\sigma_y^2} \quad (1)$$

ここで $\bar{y}_x$ は x の個有な配列に対応している y の平均値である。また x は x 配列での場合の数である。 N は全度数で σ_y^2 は y の標準偏差である。 $\bar{y}$ は凡ての y についての平均値であり、和は凡ての配列の上を動くものである。同じような方法で y 配列からも計算出来てその値は次式から得られる。

$$\eta_{xy}^2 = \frac{S\{y(\bar{x}_y - \bar{x})^2\}}{N\sigma_x^2} \quad (2)$$

両方向への回帰が直線で $\eta_{yx} = \eta_{xy} = 1$ の限られた場合を除き二つの η の値は同じではない。相関比 η_{yx} は y の標準偏差に対する平均値の標準偏差の比として表わされている事が分る。ここで右側には配列の中の数が重みづけられている。

154の例をとると次のようにして η を求める事が出来る。

各列での平均 木満期期間 $\bar{y}_x$	全平均(20.3/2) を差引いたもの $\bar{y}_x - \bar{y}$	$(\bar{y}_x - \bar{y})^2$	n_x	$n_x(\bar{y}_x - \bar{y})^2$
10.333	-9.977	99.6	6	598
13.250	7.062	49.9	4	200
13.176	7.136	50.9	17	865
16.113	4.199	17.6	62	1,091
17.230	3.052	9.50	584	5,548
20.141	171	0.029	643	19
21.877	+1.565	2.45	1,098	2,690
21.665	1.353	1.83	385	710
21.500	1.198	1.41	60	85
27.625	7.313	53.5	8	428
....			2,870	12,234

$$\therefore \eta_{yx}^2 = \frac{12234}{2870 \times (7.6067)^2} = 0.07367$$

$$\eta_{yx} = 0.2708$$

数値 7.6067 は 161 頁の $\sigma_y (=1.52135)$ にグループの単位 5 を乗じたものである。満期年令も使って同じような計算をすれば次表をうる。

$\bar{x}_y - \bar{x}$	$(\bar{x}_y - \bar{x})^2$	$n_y(\bar{x}_y - \bar{x})^2$
-3.48	12.11	678
2.20	4.84	532
1.38	1.90	821
0.64	0.41	273
+0.35	0.12	81
0.65	0.42	226
2.71	7.34	1,813
3.81	14.52	1,118
5.27	27.77	222
7.77	60.37	60
....		6,124

$$\eta_{xy}^2 = \frac{6124}{2870 \times (5.6818)^2} = 0.06610$$

$$\eta_{xy} = 0.2571$$

四章の §11 ~ §13 で述べた数値計算の過程はここで必要な数値の大部分を供給して呉れる。

§2. さて公式(1)にもとり分母を変型しよう。標準偏差の二乗は個々の観測値 O_{xy} と平均値の間の差の平方であることを思い出すと(Ⅲ章 §14 参照)次式を得る。

$$\begin{aligned} N\sigma_y^2 &= \sum_y \sum_x (O_{xy} - \bar{y}_x + \bar{y}_x - \bar{y})^2 \\ &= \sum_y \sum_x (O_{xy} - \bar{y}_x)^2 + \sum_x (n_x(\bar{y}_x - \bar{y})^2) \\ &\quad + 2 \sum_x \{(\bar{y}_x - \bar{y}) \sum_y (O_{xy} - \bar{y}_x)\} \end{aligned}$$

最後の項は 0 になるから

$$= \sum_y (O_{xy} - \bar{y}_x)^2 + \sum_x \{n_x (\bar{y}_x - \bar{y})^2\}$$

従つて

$$\eta^2_{yx} = \frac{\sum_x \{n_x (\bar{y}_x - \bar{y})^2\}}{\sum_y (O_{xy} - \bar{y}_x)^2 + \sum_x \{n_x (\bar{y}_x - \bar{y})^2\}} \quad (5)$$

$\sum_x \{n_x (\bar{y}_x - \bar{y})^2\}$ は列間の変動の量を計り $\sum_y (O_{xy} - \bar{y}_x)^2$ は列の内部の変動の量を計るものである。分母のどの項も負でないから結局

$$1 \geq \eta^2_{yx} \geq 0$$

である。

η^2 が大きくなるためには $\sum_x \{n_x (\bar{y}_x - \bar{y})^2\}$ が $\sum_y (O_{xy} - \bar{y}_x)^2$ に比して大きくならなければならない。換言すれば η^2 が大きくなればなる程配列内の変動に比して平均値の変動の変動が大きくなること (3) 式から分る。また η^2 が小さくなるにつれて配列の平均値間の差はあまり重要でなくなる。

§3. 相関係数は次の三つの目的、すなわち

- (a) x と y の間の関係を計る — これは既に例で示されている。
- (b) 列の平均 $\bar{y}_x$ の中に標本抽出から予期される以外の真の差があるかどうかを検定する。
- (c) 回帰直線を直線とみなす理由があるかどうかを検定する、のために用いられる。

(b) と (c) を扱う場合に各 x 配列に対する y の分布は正規分布から余り離れていない。またある x に対する y 配列の標準偏差は各々近似的に等しいことが仮定されていなければならない。これらの条件の下で上の (b) で述べた検定によつてもし母集団で k かの列の平均値が凡て等しく従つて η^2_{yx} の母集団値が 0 であれば x と y の値の N 対の標本で

(i) η^2 の期待値は

$$\bar{\eta}^2 = \frac{(k-1)}{N-1} \quad (4)$$

(ii) 標準誤差は

$$\sigma_{\eta^2} = \frac{1}{N-1} \sqrt{\{2(k-1)(N-k)/(N+1)\}}$$

であることが示される。

それ故にもし観察された η^2 が $\bar{\eta}^2 + 2\sigma_{\eta^2}$ より大きくなければ、平均値の差が有意である。あるいは母集団における配列の平均値は異なるという事を確信をもつて自覚することは出来ない。 η^2 の分布は配列の数が小さいならば非常に歪むので 2 倍の標準誤差という偏差は又章 §19 で述べたように考えねばならない。

同じ条件の下で検定 (c) に対して母集団で配列の平均値 ($\bar{y}_x$) が直線上にある、すなわち回帰が直線であり $\eta^2 - r^2 = 0$ であれば、そのとき x と y の値の N 対の標本で比 $\frac{\eta^2 - r^2}{1 - r^2}$ は

$$(i) \quad (k-2)/(N-2) \quad (6)$$

の期待値と

$$(ii) \quad \frac{1}{N-2} \sqrt{\{2(k-2)(N-k)/N\}} \quad (7)$$

の標準誤差をもっている。

そして我々は (b) で行つた同じような検定を応用して標本における回帰の直線性からの離れを判定出来ることが示される。

§14. 同じ数値例 (§1 の最初の表をみよ)

では $k=10$, $N=2870$ 公式 (4) と (5) からの値は

$$\bar{\eta}^2 = 0.0031$$

$$\sigma_{\eta^2} = 0.0015$$

である。我々は実際に $\eta^2 = 0.0737$ を算しているから配列平均値に真の差がある。

第二の表で公式(6)と(7)の検定を用いれば $r^2 = 0.06474$ に非常に近い $\eta^2 = 0.06610$ となる。従つて比 $(\eta^2 - r^2)/(1 - r^2)$ は 0.0014 である。この期待値は 0.0028 と標準誤差は 0.0014 である。従つてこの表は直観性を示している。最初の表は殆どその様になる事はないであらう。

§5. 相関を説明する他の方法を与えるコンティンゼンシー (Contingency) の理論に入ろう。これは、回帰が直観でないとき、あるいは、データが、IX章で述べた多重分割表の区割よりも多い区割を持ち定量的でない形で与えられているときに用いることができる。コンティンゼンシーの基礎理論はある与えられた表と相関はないが同じ周辺合計を持つ対応する表との間に作られる比較である。従つて最初の段階は相関のない表を作る方法を知らなくてはならない。そして考慮すべき事は最終合計欄にある分布に比例して他の列の合計を分けねばならないことが分るのである。従つて第一ランは

未満期々向	2	7
相関のない度数	$6 \times \frac{56}{2870}$	$6 \times \frac{172}{2870}$

である。他の部分も同じようにして作られる。各ランは合計に比例して構成されているから各ランの平均値は全体の平均値と同じでなければならぬ。そしてそのような表に相関が存在しないことは定義から直に示される。

§6. 次の表は普通の型では無相関を示す数値を与えている。これは現に小さな活字にみられる。今もしこれら二つの表の集合がある特別な場合に正確に一致すれば、この表に相関がないこと

は明らかである。もしそこに少しでも差があれば表には僅かの相関がある。またもし大きく違えば考慮してもよい程の相関がある。その結果二つの事柄の間の相関を見付けるとの方法も実際の相関表の数値と、そこにどんな相関もない時に現われるよ

最老経験の 未満期々向 の中心値	満期時の中点年令										計
	30	35	40	45	50	55	60	65	70	75	
2	0.1 2	0.1	0.3	1.2	11.4	12.5	21.4	7.6	1.2	0.2	56
7	0.4 1	0.2	1.0	3.7	35.0	38.6	65.8	23.2	3.6	0.5	172
12	0.9 2	0.6	2.6	9.3	87.8	96.8	165.4	58.4	9.0	1.2	432
17	1.4 3	0.7	3.7	14.4	135.3	147.0	250.4	82.7	13.9	1.7	665
22	1.4 1	0.9	4.0	14.6	137.2	151.0	257.8	91.1	14.1	1.7	674
27	1.1	0.8	3.2	11.6	107.5	120.6	205.9	72.7	11.2	1.4	538
32	0.5	0.4	1.5	5.3	50.3	55.3	74.4	33.4	5.2	0.7	247
37	0.2	0.1	0.5	1.7	15.7	18.2	27.4	10.4	1.6	0.2	77
42	0	0	0	0.2	1.6	1.8	3.1	1.1	0.2	0	8
47	0	0	0	0	0.2	0.2	0.4	0.2	0	0	1
計	6	4	17	62	584	643	1078	388	60	8	2870

うな数値の間の差を測ることである。XI章では二つの数値の集合の間の適合度(あるいは一致の量)を測る方法を論議して来た。それは表の各対の数値間の差を二乗しその表に相関がないときは度数で割つて χ^2 を計算してもよい事を提案したものであった。分母として相関のない表からの数値を選ぶ理由はそれが常に一つの値を持つているためである。しかし相関表は0の値

数を与えるかもしれない。

§7. x^2 の値が同連の尺度を与えることは明らかであるから x^2 と相関係数 r の間の同係をみることは興味がある。また次の証明は相関係数正規相関曲面で近似的に表わされることが出来ればグループ分けの数が大きく

$$r = \sqrt{\left\{ \frac{\phi^2}{1+\phi^2} \right\}} \quad (8)$$

であることを示している。ここで $\phi^2 = x^2/N$ 第IV章と同じ記号を用いると無相関度教曲面は次式で与えられる。

$$Z_0 = \frac{N}{2\pi\sigma_1\sigma_2} e^{-\frac{1}{2}\left(\frac{x^2}{\sigma_1^2} + \frac{y^2}{\sigma_2^2}\right)}$$

しかし相関を持つ度教曲面は

$$Z = \frac{N}{2\pi\sqrt{(1-r^2)}\sigma_1\sigma_2} e^{-\frac{1}{2}\frac{1}{1-r^2}\left(\frac{x^2}{\sigma_1^2} - \frac{2rxy}{\sigma_1\sigma_2} + \frac{y^2}{\sigma_2^2}\right)}$$

で与えられる。

そのとき

$$\phi^2 = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{(Z-Z_0)^2}{NZ_0} dx dy$$

$$= \frac{1}{N} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \left(\frac{x^2}{\sigma_1^2} - 2Z + Z_0 \right) dx dy$$

$$= \frac{1}{2\pi} \left[\frac{1}{1-r^2} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}\left\{ x^2 \frac{1+r^2}{1-r^2} - \frac{4rxy}{1-r^2} \right.} \right.$$

$$\left. + y^2 \frac{1+r^2}{1-r^2} \right\} dx dy$$

$$- \frac{2}{1-r^2} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}\left\{ x^2 \frac{1}{1-r^2} - \frac{2rxy}{1-r^2} + y^2 \frac{1}{1-r^2} \right\}} dx dy$$

$$+ \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}\{x^2 + y^2\}} dx dy \Bigg]$$

$$\text{ここで } x' = \frac{x}{\sigma_1} \quad y' = \frac{y}{\sigma_2}$$

$$= \frac{1}{1-r^2} \frac{1}{\sqrt{\left\{ \left(\frac{1+r^2}{1-r^2} \right)^2 - \frac{4r^2}{(1-r^2)^2} \right\}}} - \frac{2}{1-r^2} \frac{1}{\sqrt{\left\{ \frac{1}{(1-r^2)^2} - \frac{r^2}{(1-r^2)^2} \right\}}} + 1$$

は附録IVの(VI)式にて

$$= \frac{1}{1-r^2} - 2 + 1 = \frac{r^2}{1-r^2}$$

$$\therefore r = \pm \sqrt{\frac{\phi^2}{1+\phi^2}}$$

§8. 今得られた結果はもう少し厳密に考えてみよう。

- (1) r は -1 と $+1$ の間になければならないことを示している。
- (2) ϕ^2 の値は列(あるいは行)の順序で影響されない。その順序を変えることも許す。勿論、どの列(あるいは行)も同時に動かせる条件である。
- (3) この証明は非常に小さいグループの数が使われたら r は必ずしも正確に得られないことを示している。何故なら積分計算を用いている事によりグループの数の無限性が仮定されているからである。
- (4) どのような系列を取扱う事も仮定されている。しかしこれは相関と無相関の数値の間の適合度の尺度であるからグループの大きな数は無作為標本を用いるために甚だしく集めた偶然偏差を与える。 ϕ^2 の値から求められる r の値は σ_y モーメ

ントから求められた値からかなり異つてゐる。余り細かいグループ分けはさうでないグループ分けの場合よりも結果の正確さに乏しい。

§ 9. これらの結論は実際の計算によつても証明される。色々なグループ分けを使って大きな表から二つの方法で γ の値を見つける問題をやろうとする人は誰でも細かすぎも、また粗雑すぎもしないグループ分けのときに最もよい一致が得られる事を知るのであろう。しかしこの一般的觀察はある困難性を示す。何故ならは学生は適当なクラスの教に教値を分類するためのグループの仕方も知りたくなろう。若し年令に従つて分けられた中興を取扱うならば5年や3年の細かいグループの代りに10年のグループを取る事が出来る。あるいは最後に小さなグループを一緒にまとめてもよい。資料が6つのクラス（あるいはそれ以下）に分けられているときに各クラス等しい度数は各クラス等範囲の結果よりもよい結果を与える。しかし6つ以上のクラスに分けられているときには同じ間隔で作らねばならない。この法則は広く応用出来る。テーターの性質から度数の等しいグループを作る事はできないが近似的に等しい度数を含んでいなければならぬ。教値計算の仕方と進め方を示すために次の表の表から準備されている。

満期々間 の中心値	満期時の中心年令				合計
	50 以下	55	60	60 以上	
2.7 12	154.6 (202)	147.9 (141)	252.6 (181)	104.9 (91)	660
17	155.9 (198)	149.0 (155)	254.4 (237)	105.7 (95)	665
22	158.1 (137)	151.0 (167)	257.8 (271)	107.1 (99)	674
27	126.2 (99)	120.6 (123)	205.9 (231)	85.3 (85)	538
32 以上	78.2 (12)	74.5 (57)	127.3 (128)	53.0 (86)	333
合計	673	643	1,098	456	2,870

§ 10. 上表の各合計は互に等しくはない。60才満期の1098件はこれらの均等性を防げてはいるがもとの表の各列の合計よりはほぼ等しさが増している。さて χ^2 を計算しその値が178.8*である事が分つたから

$$\phi^2 = \chi^2/N = 0.0693$$

そしてコンティンジェンジー係数は0.254である。これは § 1*で γ に対して与えられた教値とは異つてゐるしまた VII 章の方法で求められた γ ととも異なつてゐる。すなわちもとの表は仮定された数学的型に充分厳密に従つていない。しかし個々の場合のどんな特殊性も別として一般に困難がある。と言うのはグループの数が有限であるので γ の係数を導くことは決して出来ない。だから ϕ^2 は γ が1であれば無限大になることが必要であるのだが決して無限大にはならない。同じように γ の値が他の値ととりこれが資料のグループ分けのある範囲に依存するときにはコンティンジェンジーの方法での γ の値は誤つて述べる傾向がある。全く簡単な性質ではあるが修正されねばならない。

§ 11. § 7 で ϕ^2 と γ の間の関係を作り上げたときに度数が正規相関曲面の形をとることを仮定した。これは行、または列の合計が誤差の正規曲線であると仮定していることを意味する。今 § 9 の表より細かくないグループ分けをしたと仮定しよう。そのとき、列の合計は 673, 643, 1098, 456 でありこれは合計 2870 である。あるいは全度数を1にしてこれらを表えれば

* この計算をたやすく理解するために最初の列から χ^2 の計算に参与するものは 55.1, 3.1, 2.8, 56.0, であることに注意しておこう。

* VII 章からの γ に一致することは起る。この場合に中の広いグループからの誤差と仮定された型のための偏差による誤差は均等を誤つて起る。その一致は孤立した場合から法則化する危険の例証である。

0.2345, 0.2240, 0.3826, 0.1589. である。正規曲線から度数の各グループ端での縦線を算出することが出来る。次の表を作ることも出来る。

グループ	(1)に対する	始まりからの合計面積	各グループの始めの方での縦線の大	その差	(5)ランの	(6/70)
度	数	(2)に足えることによつて得られる	き	の	二	次
(1)	(2)	(3)	(4)	(5)	(6)	(7)
列から						
675	0.2345	0.2345	0.00000	-0.022624	0.07421	0.003
643	0.2240	0.4585	0.30894	-0.08934	0.08807	0.036
1098	0.3826	0.8411	0.39678	+0.15456	0.02387	0.668
456	0.1589	1.0000	0.24222	+0.24222	0.05867	0.267
...	...	...	0.00000	...	...	...
2.870	1.0000	...	...	...	平均値 =	0.067 0.722
行から						
660	0.2300	0.2300	0.00000	-0.30365	0.07220	0.401
665	0.2317	0.4617	0.30365	-0.07345	0.08873	0.038
674	0.2348	0.6965	0.39710	+0.04759	0.00226	0.010
538	0.1875	0.8840	0.34951	+0.15421	0.02378	0.127
333	0.1160	1.0000	0.19530	+0.19530	0.03814	0.324
...	...	...	0.00000	...	...	...
2.870	1.0000	...	...	...	平均値 =	0.705 0.751

最後の数値は

$$\frac{(Z_0 - Z_1)^2}{n_1} + \frac{(Z_1 - Z_2)^2}{n_2} + \frac{(Z_2 - Z_3)^2}{n_3} + \dots$$

の平方根である。ここで $Z_0, Z_1, Z_2, \dots$ は次々のグループの始めでの縦線である。また、 $n_1, n_2, n_3, \dots$ 等々は例度数、すなわち、前表の (2) ランの数値である。修正した値は

* *Tables for Statisticians Part I Table II.*

$$\frac{0.254}{0.932 \times 0.951} = 0.286$$

となる。

§10. 前節の表で始めの3ランは容易に構成される。その第3ランは正規曲線の面積の教が曲線の任意の点から末端までの面積、すなわち Z から ∞ までの積分値を与えるという理由のために必要である。次のランは縦線を与えている。この縦線は *Tables for Statisticians Part I* に求められそこでは縦線と面積は並んだランにある。あるいは直接には *Part II* の表からも求められる。

さて理論的な面にもどり $m_1, m_2, \dots, m_{s+1}$ という各々の面積の平均は何であろうか考えよう。それは次の通りである。

$$\begin{aligned} & \frac{\int_{Z_{s+1}}^{Z_s} x e^{-\frac{1}{2}x^2} dx}{\int_{Z_{s+1}}^{Z_s} e^{-\frac{1}{2}x^2} dx} \\ &= \left(e^{-\frac{1}{2}Z_s^2} - e^{-\frac{1}{2}Z_{s+1}^2} \right) / \int_{Z_{s+1}}^{Z_s} e^{-\frac{1}{2}x^2} dx \\ &= \frac{(Z_s - Z_{s+1})}{m_{s+1}} \end{aligned}$$

故にこの式は全分布の平均値から面積 m_{s+1} の平均値までの距離を与えている。しかし m_{s+1} は度数である。故に

$$m_{s+1} \times \left(\frac{Z_s - Z_{s+1}}{m_{s+1}} \right)^2$$

を S の Z について加え上げるともこれは分布の二次モーメントを与え、二次モーメントの平方根をこれは標準偏差である。

我々は標準偏差は1である筈だと仮定していた。すなわちわれわれは実際に用いられるようにそれを再計算し結果を修正した。この修正は実際に §10 で述べた欠点の大部分を除去している。

§13. この種の場合に我々が得る筈の相関の値を知らなければ修正の必要無やその出来ばえの判断にいさごかの困難がある。修正をしないコンティンゲンシーの方法で求めた χ^2 の値が修正することによって χ^2 の真値にどれ位近く求められるかを見るために貨幣投げの表をまとめてみるのがよからう。この目的のため表が次のようにして作られている。そこでは5つの貨幣投げに使われた“対”に共通である。

2回の投げ その数の数	最初の投げその数の数			計
	0-4	5-6	7-10	
0-3	2123(3706)	2541(1606)	968(120)	5,632
4	2534(3360)	3031(2580)	1155(480)	6,720
5	3038(2906)	3640(4032)	1386(1126)	8,064
6	2534(1580)	3031(3560)	1155(1580)	6,720
7-10	2123(600)	2541(2706)	968(2326)	5,632
計	12,352	14,784	5,632	32,768

零-コンティンゲンシー数値は括弧の中にある。

$$\chi^2 = 6971 / 32768 = 0.2127$$

$$\chi^2 (\text{未修正}) = 0.418$$

前の様に修正すれば χ^2 に対して 0.753 が求められ列に対して 0.872 が求められる。故に χ^2 の修正値は $0.418 / (0.753 \times 0.872) = 0.492$

同じ貨幣投げの表を次の形に投げたとすれば別の試行が作られる。

	0-4	5	6-10	計
0-4	7,266	2,906	2,180	12,352
5	2,906	2,282	2,906	8,064
6-10	2,180	2,906	7,266	12,352
計	12,352	8,064	12,352	32,768

$$\chi^2 = 0.171$$

$$\chi^2 (\text{未修正}) = 0.382$$

行に対する修正因子は 0.872 で列に対する因子も同じである。故に修正された χ^2 の値は 0.503 である。これらは共に 0.5 であるべきである。しかし 0.492 と 0.503 は広いグループ分けを使ったための良い近似値であることは明らかである。またこれらの例は修正の重要性と正確さが得られる率を示している。

§14. しかし既に論及されたことに対するこの種の修正の問題はもっと他にもある。コンティンゲンシーの方法は列と行の順序を変えられることが含まれていると注意した (§8)。しかしこの仮定をしなければ修正によって何が起るだろうか？ この点はいくらも重要である。分割が定量的でない広いグループにおいて我々は次のこと、すなわち定量的に尺度を表わすことが出来ればそれは仮定した正規曲線のような分布を与える筈であるとは確信出来ない。これを試みるため前節の表の括弧の中の数値をとりそれらを任意に並べかえてみよう。我々は次表をうる。

χ^2 の特性	オ - の 特 性			計
	a_0	a_1	a_2	
b_0	4,032	1,126	2,906	8,064
b_1	3,560	1,580	1,580	6,720
b_2	2,980	480	3,360	6,720
b_3	2,706	2,326	600	5,632
b_4	1,606	120	3,906	5,632
計	14,784	5,632	12,352	32,768

明らかに χ^2 と未修正 χ^2 は異なる。だから我々は合計だけを考慮しの手順によればよい。これらの計算をして修正因子として 0.944 と 0.856 を得る。故に χ^2 の値は 0.517 である。これは未修正値より大部よい。—— 事実少しよい値である。説明はこうである。すなわち我々が念数でそれを割ったとはいえグループ分けが異常でなければ極端な範囲に渡って変動しない修正値を得るということである。

§15. 既にみたように二重記入の表は相関がないときにでも相関の尺度に対しては小さい値を示すしまた表面上の相関がその係数の標準誤差も正確に測ることよりも大切であるかどうかを決定することは一般的に重要である。 χ^2 の標準誤差を考慮する場合にやらねばならないことは無相関を仮定して作られた表と実際の表とを比較することであり、その相関が有意であるかどうかをみることである。実際の仕事においては係数を算出する前にこの検定を行う方が有利であろう。例えば 238 頁の表に例をとれば $\chi^2 = 178.8$ 、そして我々はそこに相関がない。また特にこの表はサンプリングでまれに起ったものであるという仮定の上で推定された値より大きいか、あるいは同じ位に大きいばらつきの確率を求めねばならない。

我々には 20 の枠がある。しかしその各行、各列の合計も固定しているのをこれを n' に用いるために余り大きすぎる。自由な枠の正確な数は $(k-1)(r-1)$ である。但し k は行の数であり、 r は列の数である。この場合では $(k-1)(r-1) = (4-1)(5-1) = 3 \times 4 = 12$ *Tables for Statisticians* の Part I Table 20 では n' は自由な枠の数より 1 つ大きい数、 $n' = n + 1$ が用

* この修正の基礎理論は正規度数曲面が上の表を表わすように分割が出来るといふ点にある。しかし特別な並び換えの場合にはこのような分割は出来ないと私は思っている。しかしこの修正は先かよくなされる。

いられている。それ故に $n' = 13$ 、 $\chi^2 = 178.8$ 。に対する表の値を代入しなければならぬ。前の計算から χ^2 の値が分つてはるから逐々相関からこのように離れる機会はいくとも小数以下 0.1 折までのことであることがみられても不思議でない。

四重分割表では自由枠は一つだけである。この節で述べた方法を詳しく表わす他の式は次のものである。

$$(i) \text{ 平均 } \chi^2 = (n-1)(k-1) \dots\dots\dots (9)$$

$$(ii) \sigma_{\chi^2} = \sqrt{\frac{2(k-1)(r-1)}{N}} \dots\dots\dots (10)$$

但しそこに相関はないとする。

公式 (9) と (10) は既に公式 (4) と (5) で χ^2 として与えたと同じような結果を表わす。

§16. コンテンジエンシーの計算で、そこに相関がないときに表にある実際の数値と対応する数値との間の差の二乗を計算する。しかし符号は無視して差を加えることによつて進めてもよいことが今まで仮定されて来た。それから場合の数の合計で割つてこれらの平均を得る。 *Tables for Statisticians* の Part I にある四に χ^2 の値が与えられている。この方法を進めるための数学的な方法は上に与えられた ミーンスクエア、コンティンジェンシー (Mean Square Contingency) に対するよりも困難である。実際の仕事の場合に右の方より危険しう。

§17. この他にも相関推定に役立つものがある。それは順位相関として知られており、Spearman* によつて提示されたものである。この方法で三つの身長とスパン (Span) の間の相関を推

* C. Spearman. Amer. J. Psychol. XV 72; K. Person, Drapers' Company Memoir. No. 4.

定しよう。但しどの子供についても正確に計ることはしない。
先づ身長順に子供を並べそれに小さいものから始めて 1, 2, 3, ... と最長迄番号をつける。次に子供をスパンの大きさ順に並べ変えて再び番号付けをする。すなわち最長のスパンの子供は 1 で最長スパンの子供は n であるようにする。そうすると身長順の順位が 1 である子供はスパンの順位では n であるかもしれない等々。このようなデータが n ケある。

次に二つの特性に対する順位の間隔の差の二乗和 $S(d^2)$ を計算する。(身長とスパンの例での計算要素の一つは $(3-1)^2$ あるいは 4 である) そして次のように書く事が出来る。

$$R = 1 - \frac{6S(d^2)}{n(n^2-1)} \quad (1)$$

$$Y = 2 \sin \frac{\pi}{6} R \quad (2)$$

ここで R は順位相関係数であり、 Y は前章で論じたと同じような変量間の係数に対応する。この関係は正規相関曲面の仮定に依存している。

この方法で求められる Y の標準誤差は X 章のモーメント法で求めたものよりも大抵 5% 位大きい。

註(1) 二つの属性 A, B の順位組合せは $n!$ 通りある。これらが同じ確率をもつとき、 R の標本分布は近似的に

$$f(R) dR = \frac{1}{B(\frac{1}{2}, \frac{n-2}{2})} (1-R^2)^{\frac{n-4}{2}} dR \quad (-1 \leq R \leq 1)$$

となる。即ち 3 次のモーメントまではこれと一致し 4 次のモーメントにおいて凡そ 36/35 n^2 の差があるにすぎない。従つて

$t = R \sqrt{n-2} / \sqrt{1-R^2}$ とおけばこれは近似的に自由度 $n-2$ の t -分布に従う。詳しくは M.G. Kendall 前掲書 (I) 16 章参

照。前順位に同じものがあるときは例えば次のような方法が採用される。

測定値 75, 80, 85, 85, 85, 90, 90, 95

1 2 4 4 4 6.5 6.5 8

註(2) Spearman の相関係数と類似の目的に用いられる Kendall の相関係数 $T = (P-Q)/T$ がある。但し $T = N(N-1)/2$ 。ここに P は例えば X_j 2, 3 ($2 < 3$)、 Y_j 5, 15 ($5 < 15$) のように大小の順序の一致する組に対して 1 と数えたときの数で Q は大小の順序が相反した組に対して -1 と数えた数である。 $P+Q = T$ に注意して

$$T = \frac{2P}{T} - 1$$

と表わされる。

註(3) n 通りの方法で n ケの物の順序付けがされている際、それらの順位付けがどの程度類似しているかを知るために一致係数 (W) がある。

$$W = 12S / n^2(n^2-n)$$

は 1 を越えない量でこれが 1 に近くなれば n ケの順位付けの類似性を表わすと考えられる。有意性の検定のためには $\frac{(n-1)W}{1-W}$ が $n_1 = (n-1) - \frac{2}{n}$ 、 $n_2 = (n-1) \{ (n-1) - \frac{2}{n} \}$ の F -分布することを用いる。ここに S は n 通りの順位付けの和の平方和である。

	$X_1, X_2, \dots, X_n$
Y_1	
Y_2	
$\vdots$	
Y_m	
T	$T_1, T_2, \dots, T_n$

第 VIII 章

偏 相 関

§1. 今までには関連した事柄を対として扱えることを仮定して来た。しかし多くの調査研究、特に社会統計ではこの問題は変量の数が多いだけに複雑である。例えば幼児の死亡調査を行い高い死亡率の第一の原因を探究しようとするとき幼児の死亡と母親の習性との間に何か関係があるかどうかをみるために、特定の地方の子供の家庭環境を調べるかも知れない。しかし母親の健康は重要であるかもしれない。また仮に (1) 幼児の死亡と母の習性、(2) 幼児の死亡と母の健康についての相関係数を知っていたとしても我々は今までに母の健康と習性の間に可能な関係を除くどんな方法も知らなかった。換言すればもし幼児の死亡率の原因が母の習性に関係していれば、それは健康と習性が関係していたとの理由を考えるだけでよいだろうか？

§2. 二つの変量だけの相関を扱うときにやったように、 $x, y, z, x_2, y_2, z_2, \dots$ 等は関連した偏差であると仮定して進めよう。先づ次のようにおく。

$$z = a + bx + cy \quad (1)$$

前のように各変量がその平均値から測られていれば a は省略できる。モーメント法を用いて次式をうる。

$$(bx_1 + cy_1) + (bx_2 + cy_2) + \dots = z_1 + z_2 + \dots$$

$$(bx_1 + cy_1)x_1 + (bx_2 + cy_2)x_2 + \dots = z_1x_1 + z_2x_2 + \dots$$

あるいは

$$bS(x^2) + cS(xy) = S(xz) \quad (2)$$

同様に

$$bS(xy) + cS(y^2) = S(yz) \quad (3)$$

158頁157頁を用いた 記号を少し変えれば次のようにかける。

$$S(xy) = N\sigma_x\sigma_y r_{xy}$$

$$S(x^2) = N\sigma_x^2$$

$$S(y^2) = N\sigma_y^2$$

故に

$$S(xz) = N\sigma_x\sigma_z r_{xz}$$

$$S(yz) = N\sigma_y\sigma_z r_{yz}$$

これらを(2)、(3)式におきかえると次式をうる。

$$b\sigma_x^2 + c\sigma_x\sigma_y r_{xy} = \sigma_x\sigma_z r_{xz}$$

あるいは

$$b\sigma_x + c\sigma_y r_{xy} = \sigma_z r_{xz}$$

また

$$b\sigma_x r_{xy} + c\sigma_y = \sigma_z r_{yz}$$

故に

$$\left. \begin{aligned} b &= \frac{\sigma_z}{\sigma_x} \cdot \frac{r_{xz} - r_{yz}r_{xy}}{1 - r_{xy}^2} \\ c &= \frac{\sigma_z}{\sigma_y} \cdot \frac{r_{yz} - r_{xz}r_{xy}}{1 - r_{xy}^2} \end{aligned} \right\}$$

これを(1)に入れ $Q=0$ に注意すれば

$$\bar{z} = \frac{\sigma_z}{\sigma_x} \cdot \frac{r_{xz} - r_{yz}r_{xy}}{1 - r_{xy}^2} \bar{x} + \frac{\sigma_z}{\sigma_y} \cdot \frac{r_{yz} - r_{xz}r_{xy}}{1 - r_{xy}^2} \bar{y}$$

2. 変量も扱ったときは次のようであつた。

$$\left. \begin{aligned} \bar{y} &= r \frac{\sigma_2}{\sigma_1} \bar{x} \\ \bar{z} &= r \frac{\sigma_1}{\sigma_2} \bar{y} \end{aligned} \right\}$$

だから相関の尺度 r はこれらの係数

$$r \frac{\sigma_2}{\sigma_1} \quad \text{と} \quad r \frac{\sigma_1}{\sigma_2}$$

の幾何平均である。同様に三変量についても z と y によつて r を、あるいは x と z によつて r をそれぞれ書き下せる。また r は係数の適当な対の幾何平均を用いて

$$rR_{xz} = \frac{r_{xz} - r_{xy}r_{yz}}{\sqrt{(1-r_{xy}^2)(1-r_{yz}^2)}}$$

を y の単一型に関連させた x と z の向の純粋(偏)相関係数として得る。この分母の平方根は必ず正の方をととする。

§3. 相関係数は1より大きくてはならない。

$$(r_{xz} - r_{xy}r_{yz})^2 \leq (1-r_{xy}^2)(1-r_{yz}^2)$$

あるいは r_{xz} は範囲

$$r_{xy}r_{yz} \pm \sqrt{(1-r_{xy}^2)(1-r_{yz}^2)}$$

の間になければならぬ。このことから三つの変量を扱ったときに起る幾つかの制限を書き下せる。

もし $r_{xy} = r_{yz} =$ ならばそのとき $r_{xz} =$

0	0	任意の値
+1	+1	1
-1	-1	1
+1	-1	-1
0	± 1	0
0	$\pm r$	$\pm \sqrt{(1-r^2)}$ の間
r	r	$1 \pm 2r^2 - 1$ の間
$-r$	$-r$	
r	$-r$	$1 - 2r^2 - 1$ の間

§4. 次の数値例を考えよう。*

母の体格 (2)	母の健康 (4)		計
	良	い	
良	956	197	1,153
貧弱	257	286	543
計	1,213	483	1,696

$$r_{xy} = 0.567 \pm 0.033$$

* "Relative Value of factors influencing infant welfare" Annals of Eugenics, 1, 175-9. この統計量は Bradford, 1911 から引用されている。読者はこの論文に同じような表と対比をみかけるでしょう。

子供の生死 (x)	母の体質 (y)		計
	良　い	悪　弱	
生　存	797	420	1,417
死　亡	156	123	279
計	1,153	543	1,696

$$r_{xz} = 0.213 \pm 0.046$$

子供の生死 (x)	母の健康 (y)		計
	良　い	良くない	
生　存	1,065	352	1,417
死　亡	148	131	279
計	1,213	483	1,696

$$r_{yz} = 0.329 \pm 0.045$$

さて“母の健康”を一定にしたときの“幼児の死亡”と“母の体質”の間の偏相関係数の計算をすれば次をうる。

$$yR_{xz} = \frac{0.213 - (0.329)(0.567)}{\sqrt{(0.691)} \sqrt{(0.678)}} = 0.034$$

換言すればこの値は始めには 0.213 のようにみえるけれども今は 0.034 にすぎないことが示された。また標準誤差は略 0.04 であるから結果が有意味であるとは言えない。

体質を一定にしたときの母の健康と幼児の死亡の間の偏相関係数を計算すれば 0.26 をうる。この値は注の数値 0.329 より小さいけれども有意味である。

図 5. 変数の数が多い場合への理論の拡張も可能である。しかしここでそれを行う必要はないようである。この例ではそのような場合にどうしたらよいかの使い方の指示を与えている。また給

別に別替する他の原因が考慮されるまで係数の数値を信ずることは危険であることを暗示している。

附 録 !

モーメントの修正

§ 1. 曲線が一方または両端で急激に上昇するときには次の方法が E. Pearson と K. Pearson によつて示されている (Biometrika XII. 231 頁以下参照)

n_1, n_2 等々はオノグループ、オノグループの比例度数である。そうすると次のようにおける。

$$a_1 = -\frac{1}{60} \{137n_1 - 163n_2 + 137n_3 - 63n_4 + 12n_5\}$$

$$a_2 = \frac{1}{12} \{45n_1 - 107n_2 + 105n_3 - 51n_4 + 10n_5\}$$

$$a_3 = -\frac{1}{4} \{17n_1 - 54n_2 + 64n_3 - 34n_4 + 7n_5\}$$

$$a_4 = \{3n_1 - 11n_2 + 15n_3 - 9n_4 + 2n_5\}$$

$$a_5 = -\{n_1 - 4n_2 + 6n_3 - 4n_4 + n_5\}$$

同じように b_1, b_2 等の値が分布のもう一つの端から求められる。

そうするとモーメントの値は次の通りである。

$$\mu'_1 = V'_1 + \left\{ \frac{1}{12} (a_1 - \frac{1}{60} a_3 + \frac{1}{2520} a_5) + \frac{1}{12} (b_1 - \frac{1}{60} b_3 + \frac{1}{2520} b_5) \right\}$$

$$\mu'_2 = V'_2 - \frac{1}{12} + \left\{ -\frac{1}{120} (a_2 - \frac{5}{126} a_4) + \frac{1}{6} A (a_1 - \frac{1}{60} a_3 + \frac{1}{2520} a_5) \right\}$$

$$\mu'_3 = V'_3 - \frac{1}{4} V'_1 + \left\{ -\frac{1}{40} (a_3 - \frac{5}{63} a_5 + \frac{1}{240} a_5) - \frac{1}{40} A (a_2 - \frac{5}{126} a_4) + \frac{1}{4} A^2 (a_1 - \frac{1}{60} a_3 + \frac{1}{2520} a_5) \right\}$$

$$\mu'_4 = V'_4 - \frac{1}{2} V'_2 + \frac{7}{240} + \left\{ \frac{1}{126} (a_2 - \frac{7}{90} a_4) - \frac{1}{10} A (a_1 - \frac{5}{63} a_3 + \frac{1}{240} a_5) - \frac{1}{20} A^2 (a_2 - \frac{5}{126} a_4) + \frac{1}{3} A^3 (a_1 - \frac{1}{60} a_3 + \frac{1}{2520} a_5) \right\}$$

ここで A は分布の出発点からモーメントが計算された原点までの距離であり B は同じ原点から分布の終点までの距離である。

A と B に關する式も同様にして得られる。もしモーメントがオノグループの始まる点で計算されていれはオノグループは $\frac{1}{2}$ 余が掛けられオノグループは $\frac{3}{2}$ 余が掛けられる等々である。この式は A を含む各項が 0 になるという理由で A の項に關する限り簡単になる。

§ 2. これらの修正を得る方法は Euler-Maclaurin 展開に始まり、そして曲線はその始めと終りで次の形

$$1 + \frac{a_1}{1!} (x-A) + \frac{a_2}{2!} (x-A)^2 + \frac{a_3}{3!} (x-A)^3 + \dots$$

をとると仮定されている。この式から a の値を導く。Euler-Maclaurin 展開に必要な各端点での微係数が展開されて上に与えられた結果に連する。

出発点での度数は近似的に

$$\frac{N}{60} \{137n_1 - 163n_2 + 137n_3 - 63n_4 + 12n_5\}$$

である。この式によつて度数曲線が範囲の終りで 0 になる場合の度数でも発見することが出来る。

§ 3. ニ三の数値例を与えよう。高位の接觸がある場合の面積は縦線を加えることによつて求められる。小數以下 5 桁まで計算された正規曲線の 12 の縦線を加えることによつて試せは 1.25000 の代りに 1.24498 が与えられる。高位の接觸を持つオノ型曲線の 9 つの縦線を加えると 24475 の代りに 24473 を与える。

§ 1 の方法の例はそこで引用された論文から取りられている。 $x=0$ から $x=10$ までの $\sqrt{2 \times 100,000}$ のモーメントが計算され、その正確な結果が分つてゐる。データーとして取られている比例度数は次の通りである。

$$\begin{aligned} a_1 &= 0.03623 & a_6 &= 0.111205 \\ a_2 &= 0.057820 & a_7 &= 0.120704 \\ a_3 &= 0.074874 & a_8 &= 0.127850 \\ a_4 &= 0.088665 & a_9 &= 0.133273 \\ a_5 &= 0.100571 & a_{10} &= 0.146135 \\ & & & 1.000000 \end{aligned}$$

これらの数値から次が得られる。

$$\begin{aligned} a_1 &= -0.0131.0643 & b_1 &= 0.14777857 \\ a_2 &= -0.0444.8167 & b_2 &= -0.00747253 \\ a_3 &= 0.0238.4150 & b_3 &= -0.0003.7450 \\ a_4 &= -0.0148.8400 & b_4 &= -0.0000.2600 \\ a_5 &= 0.0045.0200 & b_5 &= -0.00003800 \end{aligned}$$

$$a_1 - \frac{1}{60} a_3 + \frac{1}{2520} a_5 = -0.0135.3533$$

$$a_2 - \frac{5}{126} a_4 = -0.04389104$$

等々。

$A=0$, $B=10$ とおき出発点周りでモーメントを計算すれば a に対する修正

$$\frac{1}{12} (a_1 - \frac{1}{60} a_3 + \frac{1}{2520} a_5) = -0.0011.2794$$

とある。他の修正も順に

$$\frac{1}{120} (a_2 - \frac{5}{126} a_4) = -0.0003.6576$$

$$\frac{1}{40} (a_1 - \frac{5}{63} a_3 + \frac{1}{240} a_5) = -0.0003.7846$$

$$\frac{1}{126} (a_2 - \frac{7}{30} a_4) = -0.0003.4267 \text{ である。}$$

8項に対しては

$$\frac{1}{12} (b_1 - \frac{1}{60} b_3 + \frac{1}{2520} b_5) = 0.0121.0043$$

$$\frac{1}{8} B (b_1 - \frac{1}{60} b_3 + \frac{1}{2520} b_5) = 0.2500.0860$$

この他の項も順に次の通りである。

$$\begin{aligned} 3.7501.2825 & & 50.0017.1000 \\ -0.0000.6244 & & -0.0001.8731 \\ -0.0374.6365 & & 0.0037.5074 \\ 0.1500.2772 & & -0.0000.5745 \end{aligned}$$

修正されたモーメントに対して最後に次表をうる。

	モーメント法則	Sheppard の修正による	完全修正による	真 値	
V_1'	5.7880	5.7880	5.7774	6.0000	M_1'
V_2'	42.6700	42.6067	42.8570	42.9571	M_2'
V_3'	331.0854	327.5834	333.3347	333.3333	M_3'
V_4'	2.673.7735	2.677.4576	2727.2757	2727.2727	M_4'

34. 上述の方法は計算が大変であるが良い結果を与えている。尤もグループが完全基準単位に内包せずまた曲線が急に上昇する場合にこの近似法は充分満足なものではない。着者等は丁型分布に対する方法も与えている。

しかし私は指数型 (X型) に試みる事によつて簡単な近似法を得たのでこの方法は使つていない。

指数函数 $y = y_0 e^{-\frac{x}{\sigma}}$ で表わされる統計量の X の等しい範囲 A の各グループに分けられているとき次々のグループは

$$\int_0^A y_0 e^{-\frac{x}{\sigma}} dx; \int_A^{2A} y_0 e^{-\frac{x}{\sigma}} dx; \text{ 等々。あるいは } y_0 \sigma (1 - e^{-\frac{A}{\sigma}}), y_0 \sigma (1 - e^{-\frac{A}{\sigma}}) e^{-\frac{A}{\sigma}}, y_0 \sigma (1 - e^{-\frac{A}{\sigma}}) e^{-\frac{2A}{\sigma}}, \dots$$

等である。これらの各項は初項 $y_0 \cdot (1 - e^{-\frac{h^2}{2\sigma^2}})$ 、公比 $e^{-\frac{h^2}{2\sigma^2}}$ の幾何級数^{*}と考えることも出来る。もし面積を無限に展開した幾何級数として扱い、この仮定の上で、モーメントを計算しまた幾何級数を補正した項のように結果を読みとれば我々は正確な補正面積に当連し次にやや面倒な曲線に対する方程式を書き降せることとなる。

しかしながらこれ以外のことも含まれている。仮定として幾何級数を書き $A = (1 - a)^{-1}$ とおくとときその平均値まわりのモーメントは次の通りになる。

$$\begin{aligned} \text{2次モーメント} & \quad A^2 - A \\ \text{3次モーメント} & \quad 2A^3 - 3A^2 + A \\ \text{4次モーメント} & \quad 9A^4 - 18A^3 + 10A^2 - A \end{aligned}$$

また β_1 と β_2 を計算すれば $\mu + \frac{h^2}{2\sigma^2}\mu$ と $\sigma^2 + \frac{h^2}{2\sigma^2}\mu$ が夫々求められる。

指数函数を用いれば平均値周りのモーメント等は $\mu_2 = \sigma^2$, $\mu_3 = 2\sigma^3$, $\mu_4 = 9\sigma^4$, $\beta_1 = 4$, $\beta_2 = 9$ である。故にモーメントを計算するときにモーメントはほんとうは曲線からの面積であり、また Pearson 系の曲線で判定規準から曲線型の選択をするがために統計量が幾何級数をなしていると仮定するとわれわれは永久的の誤差に当連するであろう。判定基準から曲線型を決めるために統計量から求められた β_1 と β_2 は $\frac{h^2}{2\sigma^2}$ だけ引かねばならない。

この法則はすべての丁型曲線に近似法として使うことが出来て充分満足な結果を与えることがわかるであろう。

今までは曲線の出発点と面積のすべての区間が等しいことを知っていると仮定していた。この仮定が応用出来なければ指数曲線の場合には最初の(不完全な)項を除外して曲線の適合が

* 幾何級数は離散的系列を表わすのに至る処で用いられている。また指数曲線は連続的系列を表わすのに用いられている。

出来る。そしてその項は残りの補正から適當な区間の補正に内した項と考える事ができる。これは勝手なとりきめであるが、実際には有利である。

他の丁型曲線でも同じような事情で最初の段階は指数曲線を仮定し初めの不完全グループの近似的区間を逐次しそれから面積は中点に集中していることを仮定するのである。これは一般により結果を与える。すなわち指数型の仮定は急激に下降する曲線の少ないことを仮定している。故にそこには誤差の均衡がある。

統計的見方にもどると 114 頁の例は $\mu_2 = 2.045$, $\beta_1 = 4.627$, $\beta_2 = 9.502$ を与えている。これらの数値は未修正のモーメントから求めたものである。 β_1 , β_2 に対する上の値から 0.49 を引くと 4.137⁽¹⁾ と 9.012⁽¹⁾ を得る。指数曲線を用いねばならぬときの理論値は 4 と 9 である。他の丁型の場合に一つの近似法としてこの法則を適用すればねじれた丁型曲線を示した 118 頁の例では $\mu_2 = 4.266$, $\beta_1 = 0.761$, $\beta_2 = 2.646$ で、これらに対する修正値として $\beta_1 = 0.527$, $\beta_2 = 2.412$ と導くことが出来る。故に $5\beta_2 - 6\beta_1 - 9$ は、 -0.368 の代りに -0.098 となる。理論的判定規準は $5\beta_2 - 6\beta_1 - 9 = 0$ を期待させる。

これらの例は勿論完全な証明ではないがこの思いつきが正確な値を導くかもしれないことと簡単さという長所を持っていることを示すものである。 $\frac{h^2}{2\sigma^2}\mu$ による β の修正に因してこの法則は 115 頁に与えた近似法と一緒にしてもよい。そこで注意されたこと、すなわち平均値は $\mu_3 > 0$ のとき略 $\frac{h^2}{12\sigma^2}$ 大さくなり真平均値、すなわち $\frac{h^2}{12\sigma^2}$ だけ修正された平均値周りの 2 次モーメントは $\frac{h^2}{12}$ だけ小さい。 σ^2 の値は正確に分らな

詳しくは

(1) $4.627 - 0.49 = 4.137$, $9.502 - 0.49 = 9.012$ である。0.49 は $\frac{h^2}{12} = \frac{1}{2.045} = 0.435777$ を丸めた数値である。

い、しかし計算から求めた2次モーメントの平方根を用いることが出来る。もし元々単位として取られており、またモーメントが元、(すなわち計算の単位)によって求められていれば修正値は $\frac{1}{2\sqrt{\mu_2}}$ と $\frac{1}{12}$ である。

§5. §1の方法はグループに対処する中央縦線を見つけ、これらの縦線をⅢ章 §13で展開した方法で扱う事であつた。先づ中央縦線 m_1, m_2 等は次の方程によって求められる。

$$m_3 = \frac{1}{1920} \{ 2134m_3 - 116(m_2 + m_4) + 9(m_1 + m_5) \}$$

$$m_2 = \frac{1}{1920} \{ -71m_1 + 2044m_2 - 26m_3 - 36m_4 + 9m_5 \}$$

$$m_1 = \frac{1}{1920} \{ 1689m_1 + 654m_2 - 746m_3 + 364m_4 - 71m_5 \}$$

合計度数は正確に再現されないが求められたモーメントは良い近似である。

§6. 曲線が一方の端で急に上昇するか両端で上るときにモーメント計算に困難な点は既に指適されている。何故なら曲線の真の出発点あるいは終点が分からないからである。換言すれば最初の面積あるいは最後の面積の区間(あるいは両方とも)は他の面積の区間よりも小さい。実際に未修正モーメントの場合でもしばしばよい結果が得られるが最初の試みは曲線の範囲を色々に変えることによる修正が必要である。(132頁-133頁参照) これがなされたとき最初、あるいは最後あるいは両方の面積に対するモーメント修正は面積がより小さな区間の midpoint に集中すると仮定して再計算しなければならない。

E. S. Martin は *Biometrika* XXIV 12 の論文にもっと系統計にこの問題を扱っている。また彼が作った表から曲線の

正整数が推定される。

附 録 II

B-函数 と Γ -函数

$$B(m, n) = \int_0^1 x^{m-1} (1-x)^{n-1} dx$$

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

I 部分積分によると

$$\int x^{p-1} e^{-x} dx = -e^{-x} x^{p-1} + (p-1) \int x^{p-2} e^{-x} dx$$

で $p-1$ が正のとき $e^{-x} x^{p-1}$ は $x=0$ の時に0となり $x=\infty$ のときは $\frac{x^{p-1}}{e^x}$ と書かれて不定形の評価に対する法則が適用できて結果として0を与える。

$$\int_0^\infty x^{p-1} e^{-x} dx = (p-1) \int_0^\infty x^{p-2} e^{-x} dx$$

あるいは

$$\Gamma(p) = (p-1) \Gamma(p-1)$$

p が整数であれば $\Gamma(p) = (p-1)!$

$$I. \quad B(m, n) = \frac{\Gamma(m) \Gamma(n)}{\Gamma(m+n)} \quad \text{の証明}$$

$\Gamma(m)$ の式で x の代りに zx とおくと、

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

そして

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

$$\therefore \Gamma(m) \int_0^{\infty} e^{-z} z^{n-1} dz = \int_0^{\infty} \left[\int_0^{\infty} e^{-z(1+x)} z^{m+n-1} dx \right] x^{m-1} dz$$

しかし $z(1+x) = y$ とおけば

$$\begin{aligned} \int_0^{\infty} e^{-z(1+x)} z^{m+n-1} dz &= \frac{1}{(1+x)^{m+n}} \int_0^{\infty} e^{-y} y^{m+n-1} dy \\ &= \frac{\Gamma(m+n)}{(1+x)^{m+n}} \end{aligned}$$

$$\therefore \Gamma(m) \Gamma(n) = \Gamma(m+n) \int_0^{\infty} \frac{z^{m-1}}{(1+z)^{m+n}} dz$$

この積分で $(1+z) = 1/(1-z)$ と変数変換すれば

$$\int_0^1 \frac{z^{m-1}}{(1-z)^{m-1}} (1-z)^{m+n} \frac{1}{(1-z)^2} dz$$

をうる。これは直ちに $B(m, n)$ に変えられるから

$$B(m, n) = \frac{\Gamma(m) \Gamma(n)}{\Gamma(m+n)}$$

III. $\Gamma(\frac{1}{2}) = \sqrt{\pi}$ の証明

我々は既に 92 頁でみたように $2 \int_0^{\infty} e^{-x^2} dx = \sqrt{\pi}$ を知っている

ここで $x^2 = z$ と変換して次をうる、

$$2 \int_0^{\infty} e^{-x^2} dx = \int_0^{\infty} e^{-z} z^{-\frac{1}{2}} dz = \Gamma(\frac{1}{2}) = \sqrt{\pi}$$

統計的な仕事では $\Gamma(x)$ 、あるいは $\log \Gamma(x)$ の表が必要になる (本書 296 頁 ~ 299 頁 また "Tables for Statisticians" あるいはもっと詳しくは J. Brownlee の "Tracts for Computers", No 1x Camb Univ Press を参照されたい。ここでは $\log \Gamma(x)$ が $x=1$ から 50.9 まで 0.01 の間隔で与えられている) "Tracts for Computers" No 11, 中には Γ -函数の他の数値も与えている。

x が大きいときは $\log \Gamma(x)$ に近似をすることができる。最も良く知られている近似方法は

$$\Gamma(x+1) = \sqrt{2\pi x} x^x e^{-x} \frac{1}{12x} \quad * (1)$$

あるいは

$$\log_{10} \Gamma(x+1) = \log_{10} \sqrt{2\pi} + (x+\frac{1}{2}) \log_{10} x - (x-\frac{1}{12x}) \log_{10} e$$

これは x が 8 以上のときに用いられる。次の計算は $\log \Gamma(x)$ の表の使い方と近似値がどれ位真値に近づくみを示している。

* (1) この良く知られた近似法の証明は Chrystal の Algebra. II. 368 頁。あるいは Boole の Finite Difference 11 章にある。我々では高木貞治著、解析概論等参照されたい。

x	log Γ(x)		
	真 値	近 似 値	Δ
1.372	7.943775		
2.372	0.086327	0.086743	0.000416
3.372	0.461444	0.461532	0.000088
4.372	0.787332	0.787369	0.000037
5.372	1.630012	1.630025	0.000013
6.372	2.360148	2.360156	0.000008
7.372	3.164424	3.164430	0.000006
8.372	4.032009	4.032010	0.000001
9.372	4.954838	4.954837	-0.000001
10.372	5.926670	5.926667	-0.000003

今までに述べた函数は完全Γ函数。完全β函数とされている函数であるが、ある目的のためには不完全函数が必要でもある。故にもし我々がⅢ型曲線の面積を表したいなら $\int_a^b e^{-x} x^p dx$ が必要である。ここで $b-a$ は区間の長さである。この区間の積分は $\int_0^b e^{-x} x^p dx$ から $\int_0^a e^{-x} x^p dx$ を引くことによって得られる。この不完全Γ函数 $I_x(p+1)$ は $\int_0^x e^{-t} t^p dt$ で定義される。 γ_0 は毎毎に完全Γ函数に相当するものを合んでおりその値はΓ函数を決定していることに注意すれば $\frac{I_x(p+1)}{\Gamma(p+1)}$ である所の $I(x, p)$ と呼ばれる函数を表形式にする方がよい事が分った。この函数は常に0と1の間に存在し、度数がⅢ型曲線で表わされるようなものであれば与えられた範囲内の変動の残差を表わす。しかしながら x の替りに独立変数 $u = \frac{x}{\Gamma(p+1)}$ を用いるために相偶する困難を表形式にする問題に起つて来る。 $I(u, p)$ の表は公報されている*。

* Tables of the Incomplete Γ function. Printed by Cambridge University Press and published by Biometrika.

不完全β函数は

$$B_x(p, q) = \int_0^x t^{p-1} (1-t)^{q-1} dt$$

と定義されている。これも次の形で表にこれ公報されている*。

$$I_x(p, q) = \frac{B_x(p, q)}{B(p, q)}$$

不完全Γ函数とβ函数の表はある目的のために重要であるけれども序注は普通の曲線適合にはⅤ章で述べた方法で十分向に合う事が分るでしょう。

附 録 III

正規曲面への方程式

$\eta_1, \eta_2, \dots, \eta_n$ は測定可能な特性の複合体の各々の平均値からの偏差であるとしよう。測られた函数の大きさ、あるいは残差はその構成に干与する多くの独立な原因の数で決められる。今これらの原因が m あり、それらの平均値からの偏差は $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_m$ であるとしよう。そうすると $\eta_1, \eta_2, \dots, \eta_n$ は $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_m$ の函数である。しかも $m > n$ ならば ε のあるものは η のあるものの中にのみ現われる。また ε の系列は与えられた η 複合体に対して全部は決定されない。今構成に干与する原因の強さの変動

* Tables of the Incomplete Beta-Function. Published by Biometrika.

はその絶対的値さに比して小さいことを仮定する。またこれらの変動は正規分布法則に従うことを仮定する。すなわち平均値からの偏差は正規誤差曲線で補正される事を仮定する。

平均値複合体は原因の平均強度で得られるから微小なる量に因しては高次のものを省略出来るという理論によつて次をうる。

$$\left. \begin{aligned} \eta_1 &= a_{11}E_1 + a_{12}E_2 + a_{13}E_3 + \cdots + a_{1m}E_m \\ \eta_2 &= a_{21}E_1 + a_{22}E_2 + a_{23}E_3 + \cdots + a_{2m}E_m \\ &\vdots \\ \eta_n &= a_{n1}E_1 + a_{n2}E_2 + a_{n3}E_3 + \cdots + a_{nm}E_m \end{aligned} \right\} \quad (1)$$

a は決められねばならない係数である。また a の任意の粗は0であるかもしれない。何とならば特定の原因はその結果に関係しないかもしれないからである。しかも各原因が E_1 と $E_1 + SE_1$, E_2 と $E_2 + SE_2$; ... 等々 E_m と $E_m + SE_m$ の間にある確率は

$$P' = Ke^{-\left(\frac{E_1^2}{2K_1^2} + \frac{E_2^2}{2K_2^2} + \cdots + \frac{E_m^2}{2K_m^2}\right)} \times SE_1 SE_2 \cdots SE_m \quad (2)$$

で与えられる。

ここで分布の標準偏差が $K_1, K_2, K_3, \dots, K_m$ であり、 K は定数である。^{*}

* 正規曲線のどんな場合を考えても分布が“正規”であれば E_1 と $E_1 + SE_1$ の間に結果の起る找合は $\frac{1}{\sigma} e^{-\frac{E_1^2}{2\sigma^2}} SE_1$ で与えられる。ここで σ は標準偏差である。その他の E_i についても全く同じように出来る。 E_i は互に独立であるからその找合の積が求める結果を与える。

さて(1)によつて変数 E の初めの n 個を変数 η で置換えるとしよう。そうすると η_1 と $\eta_1 + S\eta_1$; η_2 と $\eta_2 + S\eta_2$; ... η_n と $\eta_n + S\eta_n$ の間に構成する複合体をうる確率は、 E_{n+1} と $E_{n+1} + SE_{n+1}$; E_{n+2} と $E_{n+2} + SE_{n+2}$; ... E_m と $E_m + SE_m$ の間にある各原因の系列と一緒にして次の通りである。

$$P' = C'e^{-\frac{1}{2}\phi^2} S\eta_1 S\eta_2 \cdots S\eta_n SE_{n+1} SE_{n+2} \cdots SE_m$$

ここで C' は K と a に関係した定数であり、 ϕ^2 は次の部分から構成されている。

- (i) η_1 から η_n までの η の二次函数である。
- (ii) E_{n+1} から E_m までの E の二次函数である。
- (iii) 次の型の函数の系列である。

$$E_{n+1}(b_{1,n+1}\eta_1 + b_{2,n+1}\eta_2 + \cdots + b_{n,n+1}\eta_n)$$

$$E_{n+1}(b_{1,n+2}\eta_1 + b_{2,n+2}\eta_2 + \cdots + b_{n,n+2}\eta_n)$$

$$E_m(b_{1,m}\eta_1 + b_{2,m}\eta_2 + \cdots + b_{n,m}\eta_n)$$

ここで b のいくつかは0であつてもよい。さて次に P' が原因 $E_{n+1}, E_{n+2}, \dots, E_m$ について $-\infty$ から $+\infty$ まで積分されれば我々は η_1 と $\eta_1 + S\eta_1$; η_2 と $\eta_2 + S\eta_2$; ... η_n と $\eta_n + S\eta_n$ の間に構成する複合体をうる全確率が得られる。しかし E_1 すなわち E_{n+1} 以下に因して積分する毎に ϕ^2 を構成している部分の定数は変わるが、(ii) と (iii) からある E の原因を消す以外に ϕ^2 を依つて三つの骨子を変えることはしない。同時に η のどんな項も C' の中に入つて来ないように C' を変えることが出来る。故に結局 $n-1$ 項のみとした后 ϕ^2 は(i)形に変えられる。あるいは η_1 と $\eta_1 + S\eta_1$; η_2 と $\eta_2 + S\eta_2$; ... ; η_n と $\eta_n + S\eta_n$ の間で複合体の構成が起る確率は

$$P = Ce^{-\frac{1}{2}\chi^2} S\eta_1 S\eta_2 \cdots S\eta_n \quad (3)$$

で与えられると結論される。

ここで x^2 は η の二次函数である。これは複合体に対する変数の法則である。

(iii) 式で x^2 に二次函数を置き換えて考えると

$$P = Ce^{-\frac{1}{2}\{C_0\eta_1^2 + C_{22}\eta_2^2 + \dots + 2C_{02}\eta_1\eta_2 + 2C_{23}\eta_1\eta_3 + \dots\}}$$

$$= Ce^{-\frac{1}{2}\{S_1(C_{pp}\eta_p^2) + 2S_2(C_{pg}\eta_p\eta_g)\}}$$

ここに C, C_{pp}, C_{pg} は定数である。 S_1 は p の凡てにわたる和を表わし、 S_2 は p, g のあらゆる“対”に対する和を表わす。

最も簡単な二変数の場合をとれば記号を少し修正して次のようになる。

$$P = Ce^{-\frac{1}{2}\{C_1\eta_1^2 + C_2\eta_2^2 + 2C_{12}\eta_1\eta_2\}}$$

P と η_i の凡てに対して $-\infty$ から $+\infty$ まで積分すれば η_2 -変動* について正規曲線を得る筈である。故に

$$\frac{1}{2\sigma_1^2} = C_2 \left(1 - \frac{C_{12}^2}{C_1 C_2}\right)$$

同じように η_2 に対して積分すれば

$$\frac{1}{2\sigma_2^2} = C_1 \left(1 - \frac{C_{12}^2}{C_1 C_2}\right)$$

全定数を得るために η_1 と η_2 の凡ての値に対して積分すれば

$$N = \frac{2C\pi}{\sqrt{(C_1 C_2 - C_{12}^2)^2}}$$

ゆへ $r = -\frac{C_{12}}{\sqrt{C_1 C_2}}$ とおき、 η_1, η_2 の代りに x, y とかけは成式をうる。

$$Z = \frac{N}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} e^{-\frac{1}{2}\left\{\frac{x^2}{\sigma_1^2(1-r^2)} - \frac{2xyr}{\sigma_1\sigma_2(1-r^2)} + \frac{y^2}{\sigma_2^2(1-r^2)}\right\}} \quad (iv)$$

* この結果 $\frac{1}{2\sigma_1^2} = C_2 \left(1 - \frac{C_{12}^2}{C_1 C_2}\right)$ は $\left\{\text{完全平方} - \frac{1}{2}C_2 \left(1 - \frac{C_{12}^2}{C_1 C_2}\right)\eta_2^2\right\}$ のように P 式の指数を配列しなおすと直に得られる。

附 録 IV

正規曲線に関する式の積分法

§1 92頁では、

$$\int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi} \quad \dots\dots\dots (i)$$

$$\therefore \int_{-\infty}^{\infty} e^{-x^2/k^2} dx = k\sqrt{\pi} \quad \dots\dots\dots (ii)$$

を示した。

曲線が対称であるから

$$\int_{-\infty}^{\infty} x^{2n+1} e^{-x^2} dx = 0 \quad \dots\dots\dots (iii)$$

をうる。

$\int x^{2n} e^{-x^2} dx$ を部分積分によって積分すれば

$$\int x^{2n} e^{-x^2} dx = \frac{x^{2n+1}}{2n+1} e^{-x^2} + \int 2x \frac{x^{2n+1}}{2n+1} e^{-x^2} dx$$

であり積分の上下限 ∞ と $-\infty$ を入れると

$$\int_{-\infty}^{\infty} x^{2n} e^{-x^2} dx = \frac{2}{2n+1} \int_{-\infty}^{\infty} x^{2n+2} e^{-x^2} dx \quad \dots\dots\dots (iv)$$

この式は次々の偶数次モーメントの間の関係を示す。

§2

$$Z = Z_0 e^{-\frac{1}{2(1-r^2)} \left\{ \frac{x^2}{\sigma_1^2} - \frac{2xyr}{\sigma_1\sigma_2} + \frac{y^2}{\sigma_2^2} \right\}}$$

これは次のように変形される。

$$Z = Z_0 e^{-\frac{1}{2(1-r^2)} \left\{ \frac{x}{\sigma_1} - \frac{yr}{\sigma_2} \right\}^2} e^{-\frac{1}{2(1-r^2)} \frac{y^2(1-r^2)}{\sigma_2^2}}$$

$$= Z_0 e^{-\frac{1}{2(1-r^2)\sigma_1^2} \left\{ x - \frac{y\sigma_1 r}{\sigma_2} \right\}^2} e^{-\frac{y^2}{2\sigma_2^2}}$$

そうすると (ii) から

$$\int_{-\infty}^{\infty} z dx = z_0 r \sqrt{2\pi(1-r^2)} \sigma_1 e^{-\frac{y^2}{2\sigma_1^2}}$$

$$N = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} z dx dy = \int_{-\infty}^{\infty} z_0 r \sqrt{2\pi(1-r^2)} \sigma_1 e^{-\frac{y^2}{2\sigma_1^2}} dy \\ = 2\pi \sigma_1 \sigma_2 r \sqrt{1-r^2} z_0 \dots \dots (V)$$

§3 上に与えられたと同じ方法で $ac > b^2$ なる次の式が示される。

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(ax^2 - 2bxy + cy^2)} dx dy = \frac{2\pi}{\sqrt{ac-b^2}} \dots \dots (VI)$$

何故なら指数は次式のように書かれ

$$-\frac{a}{2} \left\{ x - \frac{b}{a} y \right\}^2 + \frac{y^2}{2a} (ac - b^2)$$

また x に関して積分すれば次式を得る。

$$\sqrt{\frac{2\pi}{a}} e^{-\frac{y^2}{2a} (ac - b^2)}$$

そして $ac - b^2 > 0$ なる上の式を y に関して積分して次式を得る。

$$\sqrt{\frac{2\pi}{a}} \sqrt{\frac{2\pi a}{ac-b^2}} = \frac{2\pi}{\sqrt{ac-b^2}}$$

§4 $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} z xy dx dy$ を見出そう。

(V) で得たと同じ方法で

$$\int_{-\infty}^{\infty} z xy dx = \int_{-\infty}^{\infty} xy z_0 e^{-\frac{y^2}{2\sigma_2^2}} e^{-\frac{1}{2(1-r^2)\sigma_1^2} \left\{ x - \frac{y r \sigma_1}{\sigma_2} \right\}^2} dx \\ = z_0 y e^{-\frac{y^2}{2\sigma_2^2}} \int_{-\infty}^{\infty} \left(x + \frac{y r \sigma_1}{\sigma_2} \right) e^{-\frac{x^2}{2(1-r^2)\sigma_1^2}} dx$$

$$\text{ここで } x = x - \frac{y r \sigma_1}{\sigma_2}$$

$$= z_0 y e^{-\frac{y^2}{2\sigma_2^2}} \int_{-\infty}^{\infty} \frac{y r \sigma_1}{\sigma_2} e^{-\frac{x^2}{2(1-r^2)\sigma_1^2}} dx$$

∴ (iii) より $\int_{-\infty}^{\infty} e^{-\frac{x^2}{\sigma^2}} x dx = 0$ であるから

$$= z_0 \frac{\sigma_1^2}{\sigma_2} r \sqrt{2\pi(1-r^2)} y^2 e^{-\frac{y^2}{2\sigma_2^2}}$$

(IV) で $r=0$ とおき (ii) を用いると

$$\int_{-\infty}^{\infty} y^2 e^{-\frac{y^2}{2\sigma_2^2}} dy = \sigma_2^3 \sqrt{2\pi}$$

$$\therefore \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} z xy dx dy = z_0 \sigma_1^2 \sigma_2^2 2\pi r \sqrt{1-r^2}$$

何故なら (V) によつて

$$z_0 = \frac{N}{2\pi \sigma_1 \sigma_2 \sqrt{1-r^2}}$$

であるから。

§5 次に IX 章で関係した問題を扱う事が出来る。それは次の式から r の値を求めることである。

$$\frac{N}{2\pi \sqrt{1-r^2}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{1}{2(1-r^2)} (x^2 + y^2 - 2rxy)} dx dy = d$$

ここに d, N, σ_1, σ_2 は既知である。

次の式を考えよう。

$$\frac{1}{\sqrt{1-r^2}} e^{-(x^2 + y^2 - 2rxy)/2(1-r^2)} = U \dots \dots (A)$$

これを Maclaurin の定理によつて r に関して展開すると

$$U = e^{-\frac{1}{2}(x^2 + y^2)} \left(U_0 + \frac{U_1 r}{1!} + \frac{U_2 r^2}{2!} + \dots + \frac{U_n r^n}{n!} + \dots \right) \dots \dots (B)$$

$$\text{そこで } U_n = e^{-\frac{1}{2}(x^2 + y^2)} \left(\frac{d^n U}{dr^n} \right)_{r=0} \dots \dots (C)$$

r に関して U の対数微分法をとれば次をうる。

$$\frac{1}{U} \frac{dU}{dr} = -\frac{x^2+y^2-2xyr}{2} \frac{d}{dr} (1-r^2)^{-1} - \frac{(1-r^2)^{-1}}{2} \frac{d}{dr} (x^2+y^2-2xyr) - \frac{1}{2} \frac{d}{dr} \log(1-r^2)$$

$$= -(x^2+y^2-2xyr)r(1-r^2)^{-2} + (1-r^2)^{-1} xy + r(1-r^2)^{-1}$$

$$\therefore (1-r^2)^2 \frac{dU}{dr} = U \{ xy + r(1-x^2-y^2) + r^2 xy - r^3 \}$$

Leibnitz の定理により、 n 回の微分をとり $r=0$ とおくと

$$u_{n+1} = n(2n-1-x^2-y^2)u_{n-1} - n(n-1)(n-2)^2 u_{n-3} + xy \{ u_n + n(n-1)u_{n-2} \}$$

故に

$$\left. \begin{aligned} u_0 &= 1 \\ u_1 &= xy \\ u_2 &= (x^2-1)(y^2-1) \\ u_3 &= x(x^2-3)(y^2-3)y \\ u_4 &= (x^4-6x^2+3)(y^4-6y^2+3) \end{aligned} \right\} \dots\dots\dots (8)$$

(8) で表わされている法則は次のようなことである。

$$\left. \begin{aligned} u_n &= v_n \times w_n \\ v_n &= x v_{n-1} - (n-1) v_{n-2} \\ w_n &= y w_{n-1} - (n-1) w_{n-2} \end{aligned} \right\} \dots\dots\dots (9)$$

故に次のように (9) を書き直せる。

$$\frac{1}{2\pi} U = \frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)} \left(1 + \frac{v_1 w_1}{1!} r + \frac{v_2 w_2}{2!} r^2 + \dots \right) \dots (10)$$

これを x に関して $-\infty$ から ∞ まで積分する。 w_n に x が含まれていない事に注意して、

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} U dx = \frac{1}{2\pi} e^{-\frac{1}{2}y^2} \left\{ \int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2} dx + \right.$$

$$\left. \frac{r w_1}{1!} \int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2} v_1 dx + \dots + \frac{r^n w_n}{n!} \int_{-\infty}^{\infty} v_n e^{-\frac{1}{2}x^2} dx + \dots \right\}$$

$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} \left\{ v_0 + \frac{r v_1 w_1}{1!} + \dots + \frac{r^n v_n w_n}{n!} + \dots \right\}$$

ここで v_n は $\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} v_n e^{-\frac{1}{2}x^2} dx$ を表している。

またこれを y に関して $-\infty$ から ∞ まで積分する。 v_n は y を含まない事に注意し $\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}y^2} w_n dy$ を w_n で表わせば $\frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} U dx dy$ は一般項が $\frac{r^n}{n!} v_n w_n$ の級数として表わされる。として我々は w_n と v_n の値を評価しなければならぬ。

(8) から v_n の一般項が

$$x^n - n^2 x^{n-2} + 3n^4 x^{n-4} - 5n^6 x^{n-6} + \dots\dots\dots (11)$$

である事が帰納法で示される。ここに

$$n! = n(n-1)(n-2)\dots(n-t+1)/t!$$

今次の事、

$$\frac{dv_n}{dx} = n v_{n-1} \dots\dots\dots (12)$$

に注意すれば (11) によつて

$$v_n = x v_{n-1} - \frac{dv_{n-1}}{dx}$$

これを $e^{-\frac{1}{2}x^2}$ をかけて積分すると

$$\int e^{-\frac{1}{2}x^2} v_n dx = \int x e^{-\frac{1}{2}x^2} v_{n-1} dx - \int e^{-\frac{1}{2}x^2} \frac{dv_{n-1}}{dx} dx$$

部分積分法によつて第二項の積分を行へば、結局

$$\int e^{-\frac{1}{2}x^2} v_n dx = -e^{-\frac{1}{2}x^2} v_{n-1}$$

をうる。

あるいは

$$v_n = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} v_n e^{-\frac{1}{2}x^2} dx = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} (v_{n-1})_{x=x}$$

$$\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} = H, \quad \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} = K \quad \text{とかけば (u)}$$

式から次をうる。

$$\begin{aligned} \frac{d}{N} &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} U dx dy \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x^2+y^2)} dx dy + \sum_1 \left(\frac{r^n}{n!} HK (v_{n-1})_{x=x_0} \right. \\ &\quad \left. \cdot (w_{n-1})_{y=y_0} \right) \\ &= \frac{(b+d)(c+d)}{N^2} + \sum_1 \left(\frac{r^n}{n!} HK (v_{n-1})_{x=x_0} (w_{n-1})_{y=y_0} \right) \end{aligned}$$

あるいは $N = a + b + c + d$ に注意すれば次のようにもかける。

$$\begin{aligned} \frac{ad-bc}{N^2 HK} &= \sum_1 \left(\frac{r^n}{n!} (v_{n-1})_{x=x_0} \times (w_{n-1})_{y=y_0} \right) \\ &= r + \frac{r^2}{2} x_0 y_0 + \frac{r^3}{6} (x_0^2 - 1)(y_0^2 - 1) + \frac{r^4}{24} x_0 (x_0^2 - 3) y_0 (y_0^2 - 3) \\ &\quad + \frac{r^5}{120} (x_0^4 - 6x_0^2 + 3)(y_0^4 - 6y_0^2 + 3) \\ &\quad + \frac{r^6}{720} x_0 (x_0^4 - 10x_0^2 + 15) y_0 (y_0^4 - 10y_0^2 + 15) \\ &\quad + \frac{r^7}{5040} (x_0^6 - 15x_0^4 + 45x_0^2 - 15)(y_0^6 - 15y_0^4 \\ &\quad + 45y_0^2 - 15) \\ &\quad + \dots \dots \dots \text{(Vii)} \end{aligned}$$

附 録 V

曲線適合の他の方法

第II章で述べたモーメント法は曲線適合の唯一の有効な方法ではない。他の一般的な方法にも期待できることであるがモーメント法は次の三つの点で困難を与える。(i) 式

$$\int_{-\infty}^{\infty} x^n f(x, a, b, c, \dots) dx = v_n$$

に由来する方程式は $y = f(x, a, b, c, \dots)$ の定数 $a, b, c, \dots$ を求めるためには解けず。(ii) 充分正確な修正は統計的モーメントに対する凡ての条件においてまだ有効ではない。(iii) ある種のモーメントはそれらの計算が信じられず程大きき標準誤差を有する。この最後の点は Pearson 系の曲線のうちでモーメント法は充分に適合されない曲線例えば $y = y_0 \cdot (1+x^2)^{-1}$ を使つて例証もできる。

I 最小二乗法

観測値を調整する方法は実際の結果と調整された値の間の差の平方和を最小にするのである。この方法の基礎理論は誤差が正規曲線に従って分布すると言う事である。

この方法の応用例に

$$a + bx + cx^2 + \dots$$

を適合させると観測値が等荷重のとき

$$S \{ y - (a + bx + cx^2 + \dots) \}^2$$

を最小にすればよい。もし観測値をばらす $a, b, c, \dots$ 等を見出すための方程式は

$$S(y) = a S(1) + b S(x) + c S(x^2) + \dots$$

$$S(xy) = a S(x) + b S(x^2) + c S(x^3) + \dots$$

$$S(x^2y) = a S(x^2) + b S(x^3) + c S(x^4) + \dots$$

であり以下必要とだけ方程式をうるまで続けて行けばよい。

それからこれらの方程式を $a, b, c, \dots$ 等について解かねばならない。この方法は明確な条件が保たれている場合（例へば三角形の三つの角が測られたら三つの角の和は二直角に等しい）という条件に従って調整されねばならない。あるいは観察値が等しくない重みを持つている場合にも拡張されている。この種の計算方法の追求したい読者はこの問題に關した教科書を参照される事を望む。^{*}

曲線適合のこの方法の弱味は要々解けないような方程式を導く事である。これは $f(x)$ が $a + bx + cx^2 + \dots$ の代りに III 型変数曲線の如きをとつているとき前と同様な式を書き改めて解いてみれば分かる事である。すなわち曲線適合の一般的方法としてはモーメント法はむしろ望ましいものであることが分る。^{*} 二つの方法が比較されるときに都合よく共に良い結果を与へる——恐らく同じ位に良い結果を与へる。最小二乗法は最も良い結果を与へねばならない事が屢々主張されている。“最上”という意味の吟味が差の平方和を最小にすることであれば、明らかに結果は最良である。しかしこれは循環論法での主眼である。困難は最上の定義にある。曲線適合に關する美しい要訣は両方の方法とも非常に良い結果を与へることである。しかしモーメント法はより一般的应用性をもつてゐる。

II 最尤法

これは R. A. Fisher ^{*} によつて始められた方法である。もし

^{*} 例へば D. Brunt's *Combination of Observations* (Cambridge, 1917).

^{*} 羊不変系の Thiele の方法は (*Theory of Observations*) 最小二乗法よりもむしろモーメント法に類している。

^{*} *Philos. Trans. A. CCXXXII* 309 et seq;
Proc Camb. Phil. Soc XXXII 700 et seq.

分布が $y = f(x, a, b, \dots)$ で表わされていれば、 dx の範囲に観察値が落ちる確率は $f(x, a, b, \dots) dx$ で表わされる。そうすると n 々の根本のうち n_1 々は dx_1 の範囲にあり n_2 々は dx_2 の範囲にある等の確率は

$$\frac{n!}{\pi(n_p!)} \prod \{ f(x_p, a, b, \dots) dx_p \}^{n_p}$$

である。ここで積は P の凡ての値にわたつてとられるものであり $a, b, c, \dots$ は未知定数である。

最尤法はこの確率を最大にする $a, b, \dots$ の値を選ぶことからなりたつてゐる。 $a, b, c, \dots$ は f にのみ含まれてゐるから結局

$$\prod \{ f(x_p, a, b, \dots) dx_p \}^{n_p}$$

あるいは

$$S \{ n_p \log f(x_p, a, b, \dots) \}$$

を最大にすればよい。

本書で主に扱つてゐる I 型曲線に I のような適合方法を用ゐることを試みたのは著者だけのようである。また単なる一つの例でこの方法を評価するに好都合なものではないだろう。上述の場合でもこの方法はモーメントを用ゐることで始められ、それから利用するための定数を探す事である。一般的な曲線適合にこの方法を用ゐる場合の困難は最小二乗法に關して注意したのと同じような事である。すなわち得られた方程式は直接には解く事ができないし、また定数は近似法で求めねばならない。

Fisher の論文では尤度の他の快い方を大部分に取扱つてゐる。例へば Fisher はモーメント法は定数に対して有効^{*}な値を与へ

^{*} 有効度 (efficiency) の判定は大標本の場合には統計量の分布が正規性に近づくとき最小の確率誤差をもつ統計値 (statistic) を選ぶべきである。Fisher は統計的標本から得られた尤度や定数に統計値と云ふ語を使つてゐる。その尤度は上の統計量の分布云々の統計量を意味してゐる。

るかどうかを検定するために、また曲線が正規誤差曲線から遠く離れているときにその方法に疑をかけるためにこの尤度を使っている。第V. Ⅶ章で与へた例題はこの扶借的判定法を考慮に入れたもよ曲線が含まれている。しかし補整は悪いようにはみえないであらう。その上、多くの同じような曲線はモーメント法で適合されてゐて、且、実際的によい結果が得られている。試案として一定次の説明ができる。すなわち、我々が扱っている定数はそれらの分布を示している大きな標本を得たとしても正規に近づくかないこと。あるいは曲線適合に対して得られた普通の標本よりも大きな標本が正規に近づく傾向をうる前に必要であること。あるいは定数は独立ではなくて相関があること*。あるいは最尤法による検定は余りに洗練されているために実際の曲線適合では役立たない事等であると言へよう。とにかくこの他のもつと都合のよい曲線適合過程が利用出来るまで、統計的事実に満足に曲線をはめる実際的な方法を用いても悪い事はない。

Ⅲ 最小 χ^2 による方法

この方法は Dr. Kirstine Smith † によつて提出されたもので最小 χ^2 を依ることによつて曲線をあてはめる方法である。これは $S(\frac{23}{m_5})$ を最小にするのと同じことになつて来る。ここでの S は統計的度数であり m は補整度数である。

ここでも前述の方法で起つたような困難が起る。と言うのは求められた方程式は直接には解けない。K. Smith は定数は先づモーメント法で計算しそれから最小 χ^2 で修正しなければならぬ事を示した。曲線適合のこのような二役方式に伴う附加的な計算は適合法の改良を考へているか。あるいは問題の詳細な取扱ひが要求されれば、その正当性が分るであらう。引用論文には幾つか
 † 多くの曲線適合をやつた人は誰でも一つの定数計算での失敗が必ずしも凡そ曲線に影響することは限らないことを知つてゐる。
 ‡ *Biometrika* 21 262 et seq.

の例も与へられている。

この論文には最小 χ^2 方法の新しい応用が提せてありそれは実際に優れたものである。"Mortality Variations in Sweden"の研究で H. Cramér † 及 H. Wald * は死亡率の Makham の法則 ‡ $\mu_x = \alpha + \beta C^x$ を仮定し、そして

$$\chi^2 = \sum \frac{(\theta x - E_x \mu_x)^2}{E_x \mu_x}$$

の代りに

$$\chi^2 = \sum \frac{(\theta x - 2E_x - \beta C^x E_x)^2}{\theta x}$$

を用いた。ここに E_x は年齢 x での死亡の危険にさらされる数を表わし θx は死亡数に対応するものである。修正したとしても方程式は扱いにくい。故に最初の段階は C に定まつた値を仮定することであり、最小二乗法とモーメント法は χ^2 で検定ししかも χ^2 検定もまた満足であるときにより結果である α と β の値を与えることは注意してよい。それから C の二三の値で χ^2 が計算されて最終的な C の値は χ^2 の値を減らす通る拋物線を基礎として最小 χ^2 になるように選ばれる。次表はこの計算例を与へている。

1911 ~ 15; 男性。モーメント法で求められた χ^2 の値と 2 次、3 次の拋物線での補整値

$10^3 \log C$	χ^2	2 次拋物線による補整値	3 次拋物線による補整値
43.25	40.02	39.968	40.020
43.50	32.09	32.143	32.091
43.75	27.41	27.461	27.409

(V. 77)

* *Scandinavian Actuarial Tidsskrift*, 1935, pp 161 et seq.

† 保険数理士でない読者は国際保険数理士の記号での μ_x によつて表わされる死亡率をモーメントと取り違えてはならない。

(ヤツズ)

44.00	25.92	25.921	25.921
44.25	27.38	27.525	27.577
44.50	32.32	32.271	32.323
44.75	40.11	40.111	40.109

この場合に 44.00 が $103 \log c$ に対して用いられている。ついでに $\log c$ の値で比較的小さい変化でも x^2 では大きく変化する事が表から分るでせう。

c の値を決定するとき、 β と μ の新しい値はこの順序に計算されていてモーメント法によつて求められた μ の値と最小 x^2 の方法で求められた μ の値が示されている。既に説明したようにこの方法は長たらしいようにみえるが、補整は事実上 1801 から 1926 まで 5 年毎に特別の国勢調査から作られており、また 32.5 から 87.5 まで男女両性に対して 5 年毎に与えられているからこの方法は組織的解法に役立つであろう。この結果は子死死亡率に対する積死亡率曲面をみるために横に目盛られる。同じような表は出生に対しても作られる。

大きな尺度に属しての補整の例としてこの論文はその実際的重要性と興味ある方法との両方に対して研究されるべきである。

附 録 VII

保険数理の用語と記号の手引

本書に用いられている幾つかの用語と記号の説明は保険数理に関係ない読者に参考にとらう。機能と記号の充分な説明には保険数理士のための次の教科書を参照されたい。

"Interest and Annuities Certain" by R. Toddhunter
 "Life Contingencies" by E. F. Spurgeon. これらの書は

(1) どれも Cambridge University Press から出版されている。

生命保険がかけられている間の経験死亡率の調査をするには各年令で加入者の数と契約期間の推移につれて、(1) "死亡" あるいは (2) "脱退" のために各年令での観察からはづれる数や他の理由からの契約差止する数等が記録される。年令 25 女で死の危険にさらされている数 (E_{25}) とは、25 女と 26 女の間に死の危険を持つ人の数を意味し、危険である一年間の平均値である。

25 女と 26 女の間に死ぬ人の数を 25 女での危険にさらされた数で割ると 25 女の "1 年間に死亡する確率" をあたえる (q_{25}) その結果死の危険にさらされた数が新年令かまたは近似年令に従つて一覽表にされているときには死亡数は満年令に従つて表にされる。更にしばしば必要とせられ若し調査が保険契約が効力を発揮してから年令に従つて与えられているのは死亡の危険にさらされた数は正確な期間に従つて表にされ死亡数は "実施年数" に従つて表形式にされる。

ある年、例へば 1930 年と言う 1 年間の観察を終ったときにそこには危険にさらされた火山の人がいる。しかし彼等の証券はまだ有効である。これらは観察の終りに来ていると呼ばれている。補整が与えられたときに期待死亡数は q_x の補整値を死亡の危険にさらされた数に乘じて得られる。その結果が実際の死亡数と比較される。

$q_{M(5)}$ は 1863 から 1893 の間無利子証券の普通の生命保険を付けた男性から作られた死亡率表に与えられた名称である。しかし最初の 5 年の保証は含んでいない。

q_M は全期間にわたつて利子保証をもつ普通の生命保険のかけられた男性から作られた死亡率表に与えられた名称であり、 $q_M(\text{Healthy males})$ は 1863 年に終った古い経験死亡率表に与えられた名称である。

q_x (上を参照せよ) は 1 年間に死亡する確率である。

p_x は x 女の人が 1 年毎まで生存する確率である。

そうするともし人口等の増減しない社会を考え、その中の人々は死亡まで去つていく。また、 l_x は満年令 x で生きている数である。

とすれば、そのとき $l_{x+1} = p_x \times l_x$ 、 $l_{x+2} = p_{x+1} l_{x+1}$ 、
 ----である。

単位当り；利率の割合で減少になつた人が x 年右に生きていれば
 その1つの支拂手形の総額の値は $\frac{v^x l_{x+1}}{l_x}$ である。ここで $v =$
 $(1+i)^{-1}$ 、また一つの支拂手形の年金の値は

$$\frac{v l_{x+1} + v^2 l_{x+2} + \dots}{l_x}$$

表を作る便宜上この式の分母、分子に v^x を掛ければよい。すると
 年金の値として次式をうる。

$$\frac{v^{x+1} l_{x+1} + v^{x+2} l_{x+2} + \dots}{v^x l_x} = \frac{D_{x+1} + D_{x+2} + \dots}{D_x} = \frac{N_{x+1}}{D_x}$$

ここに $D_x = v^x l_x$ また、 $N_x = D_x + D_{x+1} + \dots$

同様にして $S_x = N_x + N_{x+1} + \dots$

D 、 N 、 S の表は「換算の列」と言われている。

$a_{\overline{n}|}$ はある n 年間に對して1つの支拂手形の年金の値を表わす
 ものである。とするときの寿命にも無関係であるから、その値は

$$v + v^2 + v^3 + \dots + v^n$$

である。

$\bar{a}_{\overline{n}|}$ は、1つの支拂手形の年金を m 回に分割して支拂ふ場合の乗
 極であるが、今この m を ∞ の極限に考へれば年金を毎瞬時に給分し
 て支拂ふ年金になる。このような年金は理論上屢々用いられて、そ
 の値は次のようになる。

$$\int_0^n v^m dm$$

$\text{Colog } p_x$ は p_x の余対数である。Makeham の仮説はその値
 が $A + B C^x$ であると假定する。しかし $\text{Colog } p_x = \log l_x - \log$
 l_{x+1} 故に仮説のもう一つの表現は $l_x = l_0 e^{A x} (1 - B C^x)^{-1}$ である

「死力 (μ_x) は $-\frac{1}{l_x} \frac{dl_x}{dx}$ である」 Makeham の仮説では
 この値が $A + B C^x$ の形をとる。

保険会社で証券を償ふるときに保険教士は手数をはぶくために

事実をまとめて組分けする。保険保証が死亡したときに支拂われる
 ときは、誕生年に従つてグループされる。しかし、保証が一定満期
 年金、あるいは死亡よりも前に（「養老保険」）支拂われるときに
 は、それらは誕生年、あるいは「経年数」に従つてグループされる。
 右者の場合には満期の平均年金を見付ける事によつて普通値が
 定められる。以前はこれは年金の平均をとることによつて定められ
 ていた。しかし G. J. Lielstone は幾何級数に年金の重みをつ
 ける事によつてもっと正確な結果が得られる事を示している。この
 目的のために用いられた定数は α と呼ばれている。

Model office は近似的評価を作るために使つた想像上の見本
 事業所である。

補 注

補整法 Graduation の用語は本書で頻繁に使われているので次に
 説明を述べよう。

(1) 概 説

生命保険における予定死亡率を決定するため実際の資料より得
 た粗成死亡率を平滑化するのに用いる方法にアクチュアリー（保
 険数理士）がつけた用語であるが、彼等の間では一般統計系列に
 對しても次第に用いられる様になつて來た。粗成死亡率の平滑化
 は 1725 年の De Moivre の假定まで遡る事が出来る。最初
 adjustment 或は Smoothing of data とこれと呼んでいた。
 後に graduation が一般的となり補整はこの最後の言葉の訳で
 ある。フランスでは adjustment、ドイツでは Ausgleitung
 と呼んでいるが、フランスやドイツではまだ graduation は一般
 的な用語となっていない。

(2) 定 義

補整法とは、補整可能な統計系列の数値から高次の統計的推理を
 与えんがために系列の数学的期望値の實際的近似を求めようとする
 統計的推理の一過程である。しかして誤に行ふものを棄計、幾

和、定差、集積及び比率等の統計的推理法に将来行なおうとする統計的推理によつて今行ふべき補整手段も種々掣肘を受ける。補整の方法、その結果の良否に絶対的の規準は与へ難いが一般の理論は示されている。

(3) 目的

定義から補整の対称となるのは系列統計である。補整された系列に対して平均、標準偏差、統計等々未補整値(粗値)のそれと一致させようとし、歪度、Kurtosis(尖度)等のパラメーターまで一致を試みる事もある。これらのパラメーターは系列を特徴づけるもので保ちつつ平滑化を試みるのが補整の目的である。此の平滑化は系列に対する数理統計的展開のための仮定を求めようとするため、その結果の補整系列の *efficacy* の検査にも必要と事であつて単に凹凸ある標本値を滑らかにすればよいというだけでは満足でない。いま述べた記述は一般系列統計に対するもので数理統計学における曲線のあてはめとアクチュアリー学(本来は生命保険を対称とする数理に関する学問であるが現代では此を広くに理解し、その中には生命保険の原理及び経営に関するすべての科学を含む)における補整との関係は次第に重複する部分が多くなつてゐる。アクチュアリー方面からの歩みよりに対して数理統計学者より意識的なそれは顕著ではないが実質的には殆ど差異のない処まで来ている。時系列に対する取扱いの進歩(1)、(2)は以上の記述をはっきり裏付けてゐる。

(1) J. E. Kintner: *Variate Difference Methods*, 1940

(2) M. Subaby: *Trend Analysis of statistics*, 1952.

(4) 基礎理論

統計系列に対して我々は次の仮定を設ける。即ち統計系列 $\{U_i\}$ は理想的な統計より得られると推定される真値 $\{E(U_i)\}$ と実際の統計において生ずるこの真の値よりの偏り $\{W_i\}$ とから成るものと仮定する。これは $\{E(U_i)\}$ は母集団に於けるパラメーターを表わし $\{U_i\}$ はその標本値であると言う事で $U_i = E(U_i)$

$+ W_i$; $i=1, 2, \dots$ と表わされる。補整は此の U_i が与えられたとき、これを分離して $E(U_i)$ と W_i とにする事であるが与えられているのは $\{U_i\}$ のみであるから $E(U_i)$ 及び W_i に対して更に仮定を設けて補整を実行するのである。此の仮定の性質が種々の補整方法を生ぜしめるしまた補整結果の良否検定の規準と組み合わせるこの仮定を決めてかかる事も必要である。

附 録 VII

本書の抄録

多数曲線の意味やモーメントの一般的考へ度を知らなくても、曲線の適合方法やその補整法を少し詳しく知り度い読者は I. Ⅱ 章とモーメント法による曲線の適合問題を紹介している第Ⅱ章の始めの 8 つの節を読みたい。(第 9 節はもう一つの算術上の方法と与えており省略してはいい補整に求められた対応しているモーメントに近い結果を与えるモーメントの数値計算法はよく理解すべきである (§§ 15, 16)) しかし §§ 17 ~ 22 は省略してもよい。§ 23 ~ 25 は読者におく方がよいが代数計算に余り時間をかける必要はない。また § 26 の修正法は十分に理解した方がよかつた。§ 27 は一般的の要であるが注意して読みたい。

第 IV 章は § 5 を読むことと 57 頁の表(特に注のラ)を調べることも本書の最後にある見開きの図を調べることによつて要約される。これらの曲線は Pearson の完全系を作る。すなわちこれらは密度関数(3)を表わし微分方程式

$$\frac{d \log y}{dx} = \frac{x+a}{b_0 + b_1 x + b_2 x^2}$$

から展開される。

V 章は各々の曲線型がモーメント法で適合できる方法を示してい

る。数学的な部分は全部飛ばしてもよいがこの章の始めから終りまでにくばられた数値例は曲線型の応用範囲の正しい考え方を知らるために計算すべきである。88頁~92頁は是非読まされたい。ここでは誤差の正規曲線と振っており曲線適合の簡単な例を与えてある。第Ⅶ章も読まねばならない。§5の方法を除けばここでは色々な曲線系の比較がされている。

第Ⅶ章には相関の問題が論じてある。ここで論じられている事は相関係数を計算する一以上の算術的方法を使う必要はない。また§§11~13は除いてもよい。残った章は全部読まねばならない。しかし数学的な方の少い読者は代数的な取扱いを追求することなしに結果の常識的評価に満足して165頁~178頁の代数的な証明は飛ばしてもよい。また附録はⅤ以外は全部省略してもよい。

附 録 Ⅷ

参 考 文 献 . そ の 他

本書にあるような例題を計算するには7冊の対数や三角函数を扱うために *Chamber* の対数表や *Barlow* の二乗表のような表が必要である。計算機は多くの算術計算を助けるだろう。次の表はいたる所で役立つ。

Tables for Statisticians and Biometricians, Parts I, Part II.

Edited by K. Pearson. Camb Univ Press.

Tracts for Computers Camb Univ Press 特にこのⅨ表には $\log \Gamma(x)$ の値が $x=1$ から $x=50.9$ まで 0.01 の間隔で与えられている。Ⅹ表には相関係数の確率誤差の表が与えられており、またⅪ表には乱数表が与えられている。

次の文献目録はこの問題を一般的に、あるいは部分的に取扱って

いるものである。この目録は実際で書かれたものに限っているしまた完全な書籍目録を意図したものではない。始めの書籍目録は色々な観点からこの問題の一部を取扱っているものでありより進んで研究される読者に役立つであろう。論文目録は本書で注意した問題を含んでいるが専門的なものを凡て調べた訳ではない。論文の中にはその内容が書籍目録の中の本文読むのに便利なものを含んでいるとは限らない。(例へば *Fisher* の *Statistical method* あるいは *Introductions to Tables for Statisticians*) 論題の区分は幾分か勝手なものである。

数理統計学に関する英国での最も基本的な研究は *The Philosophical Transactions of the Royal Society of Edinburgh* あるいは、月刊誌として *Biometrika*, *Metron*, *Journal of Royal Statistical Society*, *Annals of Eugenics*, *Annals of Mathematical Statistics* 等に出ており、また保険数理的な問題は *The Journal of the Institute of Actuaries* や *The Faculty of Actuaries* あるいは *Skandinavisk Aktuarietidskrift* 等に掲げている。この問題の過去に展開された理論を追求したい人々もつと進んだ問題を勉強しようとする学生はこれらの文献を参照されたい。

List of Books

- A. L. Bowley. *Elements of Statistics*. P. S. King & Son. F. Y. Edgeworth's *Contributions to Mathematical Statistics*. Royal Statistical Society.
- W. Brown and G. H. Thomson. *The Essentials of Mental Measurement*. Camb. Univ. Press.
- D. Brant. *The Combination of Observations*. Camb. Univ. Press.
- R. A. Fisher. *Statistical Methods for Research Workers*. Oliver & Boyd. *The Design of Experiments*. Oliver & Boyd.
- D. C. Jones. *First Course in Statistics*. G. Bell & Sons.
- T. L. Kelley. *Statistical Method*. The Macmillan Co.
- E. S. Pearson. *Application of Statistical Methods to Industrial Standardisation and Quality Control*. British Standards Institution.
- H. E. Soper. *Frequency Arrays*. L. Reeve & Co., Ltd.
- J. F. Stiffensen. *Some Recent Researches in the Theory of Statistics and Actuarial Science*. Camb. Univ. Press.
- J. N. Thiele. *Theory of Observations*. London 1903.
- G. H. Thomson. *How to Calculate Correlations*. G. Harrap & Co.,
- L. H. C. Tippett. *Methods of Statistics*. Williams & Norgate.
- E. T. Whitaker and G. Robinson. *Calculus of*

observations. Blackie and Co.

- G. V. Yule and M. G. Kendall. *Introduction to the Theory of Statistics*. Charles Griffin & Co.

Lists of Papers
Method of Moments, etc.

- H. Cramér and H. Wold. "Mortality Variations in Sweden." *Skand. Aktuarietidskrift*, 1935, p. 161.
- R. A. Fisher. "On the mathematical foundations of theoretical statistics." *Philos. Trans. A* CCII, p. 309. See also *Proc. Camb. phil. Soc.* XXII, p. 700.
- E. S. Martin. "On corrections for the moment coefficients of frequency distributions when the start of the frequency is one of the characteristics to be determined." *Biometrika*, XXVI, p. 12.
- E. Pearson and K. Pearson. "Corrections for moment coefficients." *Biometrika*, XII, p. 231.
- K. Pearson. "Systematic fitting of curves to observations." *Biometrika*, I p. 265, II, p. 1.
- W. F. Sheppard. "Calculation of the most probable values of frequency constants, etc." *Proc. Lond. Math. Soc.* XXX p. 353.
- K. Smith. "On the 'Best' values of the constants in Frequency Distributions." *Biometrika*, XT, p. 262.
- G. V. L. Charlier. "Researches into the theory of

- probability." *Meddelanden Sveriges Astronomiska observatorium*, 1906. "A new form of the frequency function." *Ibid.*, 1928.
- H. Cramér. "On the composition of elementary errors." *Skand. Aktuarietidskrift*, 1928. pp. 13 and 141.
- F. Y. Edgeworth. "Mathematical representation of statistical data." *J. Roy. Statist. Soc.* LXI, p. 691; LXIX, p. 497; LXX, p. 102; LXXIX, p. 455; LXXX, pp. 65, 266, 411; LXXXVII, p. 571.
- "Generalised Law of Error." *Proc. Camb. Phil. Soc.* XX, pp. 36, 113.
- W. P. Elderton. "An approximate law of survivorship and other notes on the use of frequency curves in actuarial statistics." *J. Inst. Actu.* LXV, p. 1.
- J. C. Kapteyn. *Skew Frequency Curves in Biology and Statistics*. Groningen, 1903 and 1916.
- K. Pearson. "Skew variation in homogeneous material." *Philos. Trans. A*, CLXXXVI, p. 343; CXCII, p. 443; CCXVI, p. 429.
- E. C. Rhodes. "On the generalised law of error." *J. Roy. Statist. Soc.* LXXXVIII, p. 576.

Correlation and Contingency

- E. M. Elderton. "Relative value of factors influencing infant welfare." *Annals of Eugenics*, I.
- F. Galton. "Family likeness in stature." *Proc.*

- Roy. Soc.* XI, p. 42. "Correlations and their measurement." *Proc. Roy. Soc.* XLV, p. 136.
- K. Pearson. "Regression, Heredity and Panmixia." *Philos. Trans. A*, CLXXXVII, p. 253. "Correlation of characters not quantitatively measurable." *Philos. Trans. A*, CXCIV, p. 1. "Theory of Contingency; Theory of Skew Correlation; Further methods of measuring correlation; novel method of regarding the association of two variates." *Drapers' Company Memoirs*, Camb. Univ. Press.
- "New method of determining correlation." *Biometrika*, VII, p. 248. "On a form of spurious correlation." *Proc. Roy. Soc.* LX, p. 489. Also *Biometrika*, VIII, p. 254; IX, p. 116; XIV, p. 281.
- C. Spearman. "Proof and measurement of association between two things." *Amer. J. Psychol.* XV, p. 72; also *Brit. J. Psychol.* II, p. 96. "Student." *Elimination of Spurious Correlation, etc.* *Biometrika*, X, p. 179.
- U. Yule. "Significance of Bravais' Formulae, etc." *Proc. Roy. Soc.* LX, p. 477. "Theory of correlation." *J. Roy. Statist. Soc.* LX, p. 812. "On time correlation problem." *J. Roy. Statist. Soc.* XXXIV, p. 497; also *J. Roy. Statist. Soc.* LXXXIX, p. I.

Standard Errors, Goodness of Fit, etc.

Biometrika, editorial. "Probable errors of frequency constants." Biometrika, II, p. 273.

F. L. Engledow and G. V. Yule. Principles and Practice of Yield Trials. Empire Cotton Growing Corporation, 1926.

R. A. Fisher. Various papers dealing with the subject and especially with χ^2 and small samples. J. Roy. Statist. Soc. LXXXV, p. 597; LXXXVII, p. 442; Economica, 1923, p. 139; Biometrika, X, p. 507; Metron, I, pt. 4, p. 1; V, pt. 3, p. 92.

R. Henderson. "Frequency curves and moments." Trans. Acta. Soc. America, VIII, p. 30 or J. Inst. Acta. XLI, p. 429.

M. Greenwood. "Errors in random sampling." Biometrika, IX, p. 69.

T. Kondo. "Standard error of mean square contingency." Biometrika, XXI, p. 376.

E. S. Pearson. "Some notes on sampling tests with two variables." Biometrika, XXI, p. 337.

K. Pearson. "Criterion that a given system of deviation can have arisen from Random Sampling." Phil. Mag. July 1900. See also various other papers. Biometrika, VII, p. 250; XI, p. 292; XIV, pp. 186, 418; (With L. N. G. Filon.) Philos. Trans. A, CXCI, p. 229. "Student." "Small samples." Biometrika, VI, p. 302; X, p. 384; XV, p. 271.

"Rank and grades correlation method." Biometrika

XIII, p. 264.

J. Wishart and H. G. Sanders. Principles and Practice of Field Experimentation. Empire Cotton Growing Corporation.

G. V. Yule. "Application of χ^2 to contingency tables." J. Roy. Statist. Soc. LXXXV, p. 95.

附 録 IX 数 表

便 利 定 数

$$e = 2.7182818285..$$

$$e^{-1} = 0.3678794417$$

$$\pi = 3.1415926536$$

$$\log_{10} e = 0.4342944820$$

$$\log_{10} e^{-1} = 2.30258509$$

$$\log(\log_{10} e) = 7.6377843114$$

$$\log_{10} \pi = 0.4971498728$$

$$\log_{10} \sqrt{\pi} = 0.2485749364$$

$$\log_{10} \frac{1}{\sqrt{2\pi}} = 7.6009100657$$

$$\log_{10} e^{-\frac{1}{2}} = 7.9638087932$$

横座標で表わした正規曲線の面積と縦軸の値.

X	$\frac{1}{2}(1+Z)$	Z
.0	.5000	.3989
.1	.5398	.3970
.2	.5793	.3910
.3	.6179	.3814

(1/2)π

.4	.6554	.3683
.5	.6915	.3521
.6	.7257	.3332
.7	.7580	.3123
.8	.7881	.2897
.9	.8159	.2661
1.0	.8413	.2420
1.1	.8643	.2179
1.2	.8849	.1942
1.3	.9032	.1714
1.4	.9192	.1497
1.5	.9332	.1295
1.6	.9452	.1109
1.7	.9554	.0940
1.8	.9641	.0790
1.9	.9713	.0656
2.0	.9772	.0540
2.5	.9938	.0175
3.0	.9987	.0044
3.5	.9998	.0009

$z = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2}$ 表には $\sigma = 1$ が仮定されて
いる。

$$\alpha = \int_{-x}^x z dx \quad \text{あるいは} \quad \frac{1}{2}(1+\alpha) = \int_{-\infty}^x z dx$$

適合度検定 P の値

χ^2	$n=2$ $n'=3$	$n=5$ $n'=6$	$n=8$ $n'=9$	$n=11$ $n'=12$	$n=14$ $n'=15$	$n=17$ $n'=18$	$n=20$ $n'=21$	$n=23$ $n'=24$	$n=26$ $n'=27$	$n=29$ $n'=30$	χ^2
3	.223	.700	.934	.991	.999	1.000	1.000	1.000	1.000	1.000	3
6	.050	.306	.647	.873	.966	.993	.999	1.000	1.000	1.000	6
9	.011	.109	.342	.622	.831	.940	.983	.996	.999	1.000	9
12	.002	.035	.151	.363	.606	.800	.916	.970	.991	.998	12
15	.001	.010	.059	.182	.378	.595	.776	.895	.957	.985	15
18	.000	.003	.021	.082	.207	.389	.589	.757	.876	.944	18
21	.000	.001	.007	.033	.102	.226	.397	.581	.742	.859	21
24	.000	.000	.002	.013	.046	.119	.243	.404	.576	.729	24
27	.000	.000	.001	.005	.019	.058	.135	.256	.409	.572	27
30	.000	.000	.000	.002	.008	.026	.070	.149	.268	.414	30

$\chi^2 = 0$ 且 $P = 1$ ならば 概略補正は P の代りに $\log P$ を
使つて縦にも横にも可能であるしかし対角線にはだめである。
 n は自由度の数であり $n' = n + 1$ 。此の表のもとは *Table*
for Statisticians にある。

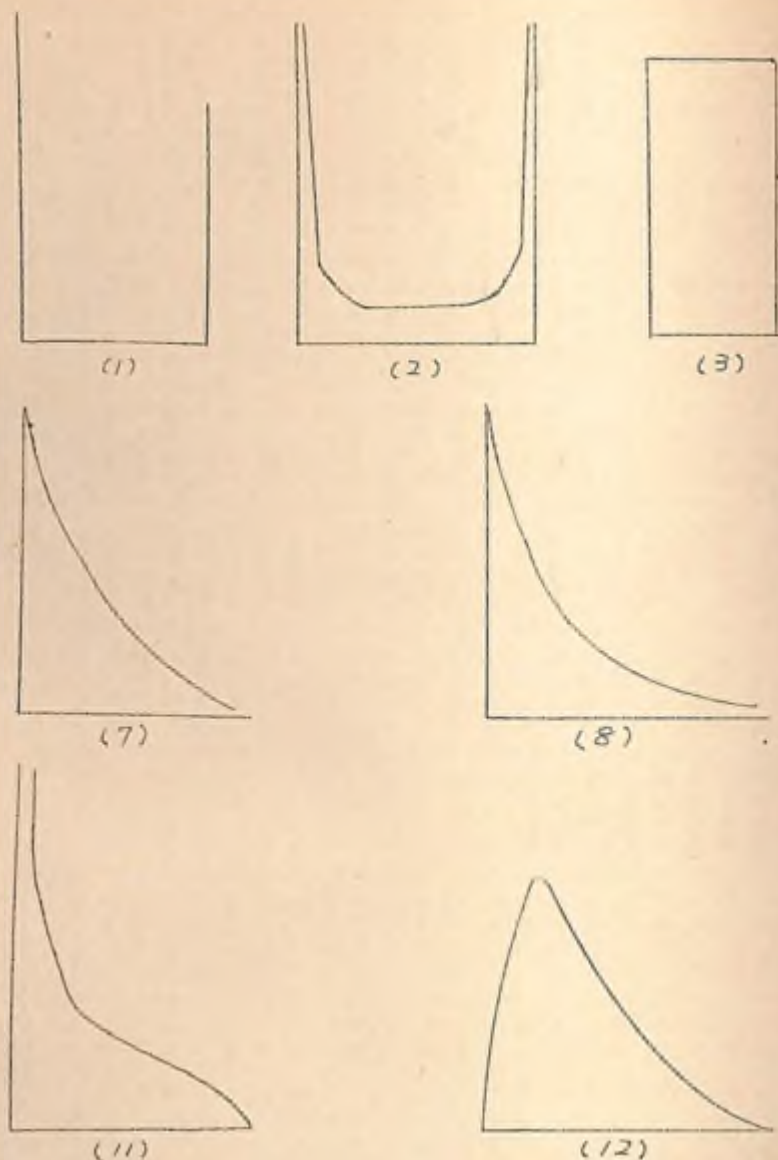
log Γ(p) の表

p	0	1	2	3	4	5	6	7	8	9
100		9950	9500	9251	9003	8755	8509	8263	8017	7773
101	7.99	7285	7043	6801	6560	6320	6080	5841	5602	5365
102	7.99	5128	4656	4421	4187	3953	3721	3489	3257	3026
103	7.99	2796	2338	2110	1883	1656	1430	1205	981	757
104	7.99	0553	0311	0089	9668	9427	9208	8989	8772	8554
105	7.98	8538	8122	7907	7692	7478	7265	7053	6841	6629
106	7.98	6209	6000	5791	5583	5376	5169	4963	4758	4549
107	7.98	4145	3943	3741	3539	3338	3138	2939	2740	2544
108	7.98	2147	1951	1755	1560	1365	1172	978	786	594
109	7.98	0212	0022	9833	9644	9456	9269	9082	8896	8710
110	7.97	8341	8157	7974	7791	7610	7428	7248	7068	6888
111	7.97	6531	6354	6177	6000	5825	5650	5475	5301	5128
112	7.97	4783	4612	4441	4271	4101	3932	3764	3596	3429
113	7.97	3096	2931	2766	2602	2438	2275	2113	1951	1790
114	7.97	1469	1309	1150	992	835	677	521	365	210
115	7.96	9901	9747	9594	9442	9290	9139	8988	8838	8688
116	7.96	8390	8243	8096	7949	7803	7658	7513	7369	7225
117	7.96	6939	6797	6655	6514	6374	6234	6095	5957	5818
118	7.96	5544	5408	5272	5137	5002	4868	4734	4601	4461
119	7.96	4205	4075	3944	3815	3686	3557	3429	3302	3175
120	7.96	2922	2797	2672	2548	2425	2302	2179	2057	1936
121	7.96	1695	1575	1456	1337	1219	1101	984	867	751
122	7.96	0521	0407	0293	0180	0067	955	843	732	621
123	7.95	9401	9292	9184	9076	8968	8861	8755	8649	8544
124	7.95	8335	8231	8128	8025	7923	7821	7720	7620	7520
125	7.95	7321	7223	7125	7027	6930	6834	6738	6642	6547

p	0	1	2	3	4	5	6	7	8	9
126	7.95	6359	6267	6173	6081	5989	5897	5807	5716	5627
127	7.95	5449	5360	5273	5185	5099	4999	4927	4842	4757
128	7.95	4589	4506	4423	4341	4259	4178	4097	4017	3938
129	7.95	3780	3702	3624	3547	3470	3394	3318	3243	3168
130	7.95	3020	2947	2874	2802	2730	2659	2588	2518	2448
131	7.95	2310	2242	2174	2106	2040	1973	1907	1842	1777
132	7.95	1648	1585	1522	1459	1397	1336	1275	1214	1154
133	7.95	1055	0977	0918	0861	0803	0747	0690	0634	0579
134	7.95	0470	0416	0362	0309	0257	0205	0153	0102	0051
135	7.94	9951	9902	9853	9805	9757	9710	9663	9617	9571
136	7.94	9480	9435	9391	9348	9304	9262	9219	9178	9136
137	7.94	9054	9015	8975	8936	8898	8859	8822	8785	8748
138	7.94	8676	8640	8605	8571	8537	8503	8470	8437	8405
139	7.94	8342	8311	8280	8250	8221	8192	8163	8135	8107
140	7.94	8053	8026	8000	7975	7950	7925	7901	7877	7854
141	7.94	7808	7786	7765	7744	7723	7703	7683	7664	7645
142	7.94	7608	7590	7573	7556	7540	7524	7509	7494	7479
143	7.94	7451	7438	7425	7413	7401	7389	7378	7368	7358
144	7.94	7338	7329	7321	7312	7305	7298	7291	7284	7278
145	7.94	7268	7263	7259	7255	7251	7248	7246	7244	7242
146	7.94	7240	7239	7239	7240	7241	7242	7243	7245	7248
147	7.94	7254	7258	7262	7266	7271	7277	7282	7289	7295
148	7.94	7310	7317	7326	7334	7343	7353	7363	7373	7384
149	7.94	7407	7419	7431	7444	7457	7471	7485	7499	7514

P	0	1	2	3	4	5	6	7	8	9
1.50	7.94	7545	7577	7594	7612	7629	7647	7666	7685	7704
1.51	7.94	7724	7764	7785	7806	7828	7850	7873	7896	7919
1.52	7.94	7943	7991	8016	8041	8067	8093	8120	8146	8174
1.53	7.94	8201	8258	8287	8316	8346	8376	8406	8437	8468
1.54	7.94	8500	8564	8597	8630	8664	8698	8732	8767	8802
1.55	7.94	8837	8910	8946	8983	9021	9059	9097	9135	9174
1.56	7.94	9214	9294	9334	9375	9417	9458	9500	9543	9586
1.57	7.94	9629	9716	9761	9806	9851	9896	9942	9989	0035
1.58	7.95	0082	0177	0225	0274	0323	0372	0422	0472	0522
1.59	7.95	0573	0624	0728	0780	0833	0886	0939	0993	1047
1.60	7.95	1102	1157	1268	1324	1380	1437	1494	1552	1610
1.61	7.95	1668	1727	1845	1905	1965	2025	2086	2147	2209
1.62	7.95	2271	2333	2459	2522	2586	2650	2715	2780	2845
1.63	7.95	2911	2977	3110	3177	3244	3312	3380	3449	3517
1.64	7.95	3587	3656	3797	3867	3938	4010	4081	4154	4226
1.65	7.95	4299	4372	4519	4594	4668	4743	4819	4894	4970
1.66	7.95	5047	5124	5278	5356	5434	5513	5592	5671	5740
1.67	7.95	5830	5911	6072	6154	6235	6317	6400	6482	6566
1.68	7.95	6649	6733	6901	6986	7072	7157	7243	7322	7416
1.69	7.95	7503	7590	7766	7854	7943	8032	8122	8211	8301
1.70	7.95	8391	8482	8664	8756	8848	8941	9034	9127	9220
1.71	7.95	9314	9409	9598	9693	9788	9884	9980	0077	0174
1.72	7.96	0271	0369	0565	0664	0763	0862	0961	1061	1162
1.73	7.96	1262	1363	1566	1668	1770	1873	1976	2079	2183
1.74	7.96	2287	2391	2601	2706	2812	2918	3024	3131	3238
1.75	7.96	3345	3459	3669	3778	3887	3996	4105	4215	4326

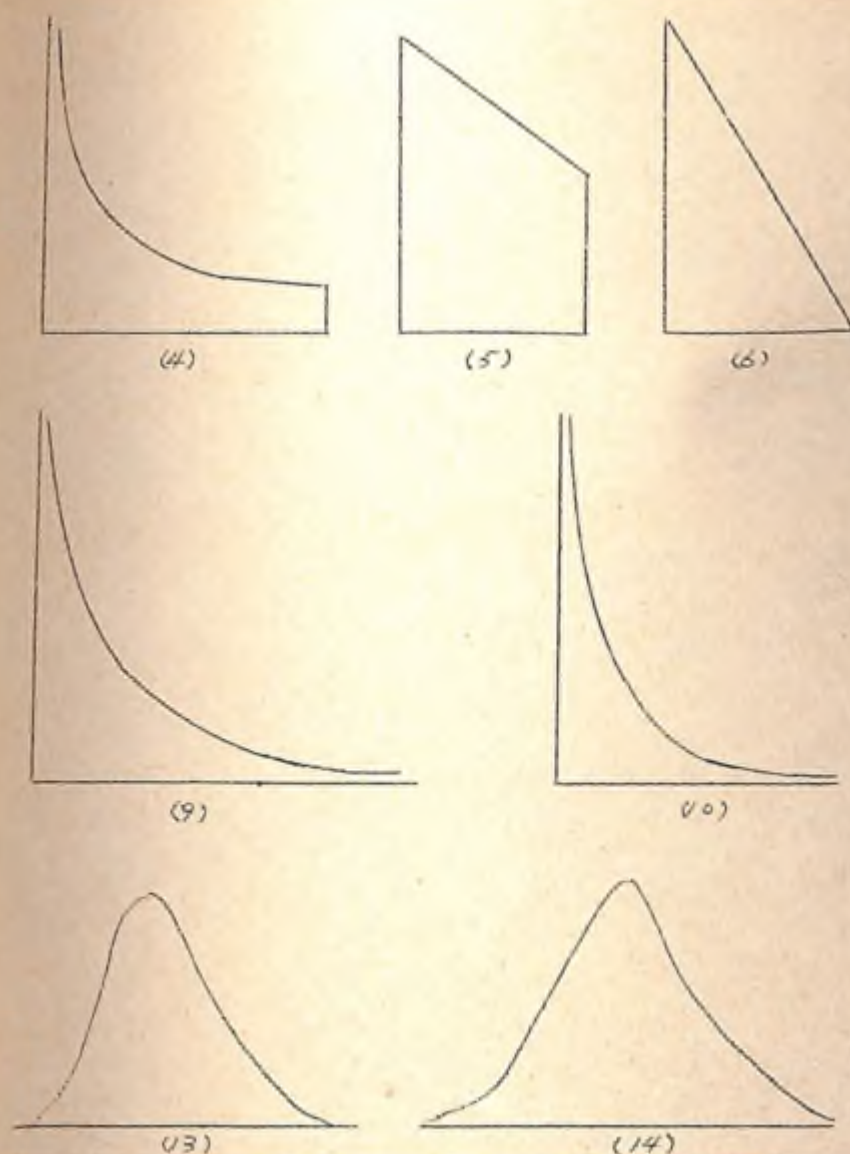
P	0	1	2	3	4	5	6	7	8	9
1.76	7.96	4436	4547	4659	4770	4882	4994	5107	5220	5333
1.77	7.96	5561	5675	5789	5904	6019	6135	6251	6367	6484
1.78	7.96	6718	6835	6953	7071	7189	7308	7427	7547	7666
1.79	7.96	7907	8028	8149	8270	8392	8514	8636	8759	8882
1.80	7.96	9129	9253	9377	9501	9626	9751	9877	0003	0129
1.81	7.97	0383	0509	0637	0765	0893	1021	1150	1279	1408
1.82	7.97	1668	1798	1929	2060	2191	2322	2454	2586	2719
1.83	7.97	2985	3118	3252	3386	3520	3655	3790	3925	4061
1.84	7.97	4333	4470	4606	4744	4881	5019	5157	5295	5434
1.85	7.97	5712	5852	5992	6132	6273	6414	6555	6697	6838
1.86	7.97	7123	7266	7408	7552	7696	7840	7984	8128	8273
1.87	7.97	8564	8710	8856	9002	9149	9296	9443	9591	9739
1.88	7.98	0036	0184	0333	0483	0633	0783	0933	1084	1234
1.89	7.98	1537	1689	1841	1994	2147	2299	2453	2607	2761
1.90	7.98	3069	3224	3379	3535	3690	3846	4003	4159	4316
1.91	7.98	4631	4789	4947	5105	5264	5423	5582	5742	5902
1.92	7.98	6223	6383	6544	6706	6867	7029	7192	7354	7517
1.93	7.98	7844	8007	8171	8336	8500	8665	8830	8996	9161
1.94	7.98	9494	9660	9827	9995	0162	0330	0498	0666	0835
1.95	7.99	1173	1343	1512	1683	1853	2024	2195	2366	2537
1.96	7.99	2881	3054	3227	3399	3573	3746	3920	4094	4269
1.97	7.99	4618	4794	4969	5145	5321	5498	5674	5851	6029
1.98	7.99	6384	6562	6740	6919	7098	7277	7457	7637	7817
1.99	7.99	8178	8359	8540	8722	8903	9085	9268	9450	9633



図は全体を通じて度数をブロックに分けるU型、丁型曲線から共通の度数の山高幅型までの変化を示している。

(1) は度数を二つのブロックに分けたものを表わし、これが発展して(2)にあるU型曲線になる。

(3) 水平線といわば、U型曲線の底部の一部分である。(4)の丁型曲線はU型曲線の一部分の様なものである。(5)と(6)は直線



になったときの限界である。

(7)はIX型であり(8)は指数曲線である。その次の二つの曲線(9)と(10)はII型とI型の丁型曲線であり(11)はXII型の曲線である。此等からI、II、IV、V、VIに進む事が出来、山高幅型の曲線は三つの例(12)、(13)、(14)に与えられているものである。