なる値の變動がある．かくの如き不便があるから，m_1，m_2 を以て相關の程度を決定せんとならば愼重なる注意が必要である．

(ii)　r 及び m_1, m_2 は常に p と同一の附號を取る．

(iii)　變數が互に獨立なる時は,卽ち無相關なる時は，RR′ は M_x 軸に，CC′ は M_y 軸に一致し，$r=0$ 又 $m_1=m_2=0$ となる．又完全なる正の相關にある時は，RR′ と CC′ とは合一し，

$$m_1m_2=+1, \quad r=+1,$$

となるが,この時必ずしも $m_1=m_2$ ならず，

又,完全なる負の相關に於ては，

$$m_1m_2=+1, \quad r=-1 \quad \text{となる．}$$

この時，m_1, m_2 の附號を以て r の附號と定める．故に m_1 も m_2 も共に負であるが,必ずしも等しくない．

(iv)　又完全相關ならざる時は，

$$\alpha+\beta<90^\circ$$

であるから，m_1, m_2 卽ち $\tan\alpha$, $\tan\beta$ は 1 より小である．卽ち

$$|r|<1$$

である．而して，m_1,m_2 の何れか一つは常に 1 より小であるが，他は必ずしも 1 より小とは限らない．

(v)　生物統計學では無相關と完全相關,及びこれに近き場合が稀であつて,普通はその中間にある．讀者は相關表を見ただけで，r の附號は勿論,その値の概略を推算する程にならねばならぬ．

(vi)　ここに一つ注意すべきことは，r が 0 又は極めて小なる

時,直に無相關なりと早合點してはならぬことである．

無相關なる場合は $r\fallingdotseq0$ であるがその逆は常に眞でない．前章,例 8 は,決して無相關ではないが，r を計算すれば極めて小であつた($r=-0{\cdot}014$)．かく回歸が一次でない相關では，r の値は相關の程度を忠實に表はさないから,常に他の研究を必要とする．例 9 の如きはこれを 3 個の相關表に分ち,それぞれの相關表につきて，r_1, m_1, m_2 を計算すべきである．かく,相關が一次でなき時は特種な研究法がいるから,回歸が一次なりや否やを略見定めることは必要なことであるが,大抵の場合,必ずしも各行及び各列の平均を計算するを要せず,單なる視察で特種な傾向が見られなければ一次相關と定めてよい．

(vii)
$$m_1=r\frac{\sigma_x}{\sigma_y},$$
$$m_2=r\frac{\sigma_y}{\sigma_x},$$

卽ち，$\dfrac{m_1}{m_2}=\left(\dfrac{\sigma_x}{\sigma_y}\right)^2$ であるから，σ_x と σ_y との比は m_1, m_2 の大さを左右する．

8.　相關係數及び回歸係數の計算法　I.

相關關係をあらはす最も必要なる數値は相關係數 r で,これに次ぐものは回歸係數 m_1, m_2 である．相關表を作り,回歸が一次なることが視察で,略,明になれば,必ずしも各行及び各列の平均を計算するには及ばず,直ちに r の計算を行つてよし．

r 及び m_1, m_2 の計算には，

$$p=\frac{1}{N}\sum(fxy)$$

を計算せねばならぬ．$\Sigma(fxy)$ は或る欄に於ける度數 f と,その

欄の中央の値の偏差 x 及び y の積 fxy を取り,すべての欄につきこの積の總和を表はすのである.

階級數少ければ, $\Sigma(fxy)$ の計算もさほど困難ではないが,階級多數となり, x, y が簡單な數でない時は,計算は面倒であるから,是非簡便算が必要である.

今,第一變數の階級中央の値 $X_1, X_2, \cdots\cdots, X_K$,
その算術的平均を M_x, 又

$$X_1 - M_x = x_1,\quad X_2 - M_x = x_2,\ \cdots\cdots \text{ とす.}$$

又,任意の原點 A に對して,

$$X_1 - A = \xi_1,\quad X_2 - A = \xi_2,\ \cdots\cdots$$

$$M_x - A = d \quad \text{とすれば,}$$

$$\xi_1 = x_1 + d,\quad \xi_2 = x_2 + d,\ \cdots\cdots$$

又,第二變數に於ては,階級中央の値を $Y_1, Y_2, \cdots\cdots, Y_h$,
その算術的平均を M_y, 任意の原點を B とすれば,

$$Y_1 - M_y = y_1,\quad Y_2 - M_y = y_2,\ \cdots\cdots$$

$$Y_1 - B = \eta_1,\quad Y_2 - B = \eta_2\ \cdots\cdots$$

$$M_y - B = d' \text{ とすれば,}$$

$$\eta_1 = y_1 + d',\quad \eta_2 = y_2 + d',\ \cdots\cdots$$

故に,各欄の度數を f とすれば,

$$\Sigma(f\xi\eta) = \Sigma f(x+d)(y+d'),$$

$$= \Sigma(fxy) + d\,\Sigma(fy) + d'\,\Sigma(fx) + Ndd',$$

しかるに, $\Sigma(fy) = \Sigma(fx) = 0$

なるが故に,

$$\frac{1}{N}\sum(f\xi\eta) = \frac{1}{N}\sum(fxy) + dd', \cdots\cdots\cdots\cdots\cdots\text{(A)}$$

$$\frac{1}{N}\sum(fxy) = p,$$

又, $$\frac{1}{N}(f\xi\eta) = p',$$

とすれば

$$p = p' - dd' \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\text{(B)}$$

ξ, η は A 及び B を適當に選ぶことによつて整數となすことを得るが故に, $\Sigma(f\xi\eta)$ の計算は $\Sigma(fxy)$ の計算に比し遙に容易である. 又, dd' は常數で唯一回の計算である. 故に式(B)は p の計算に代はりて用ひらる.

定理. 變數 X, Y に於て計算の原點を任意の點 A, B に取る時,各階級の中央の値が A 或は B に對する偏差を ξ, η とし,又 A 又は B の M_x 又は M_y との差を d, d' とすれば,

$$\left.\begin{array}{l} r = \dfrac{\Sigma(f\xi\eta) - Ndd'}{N\sigma_x\sigma_y} = \dfrac{p' - dd'}{\sigma_x\sigma_y} \\ \text{但し}\quad p' = \dfrac{1}{N}\sum(f\xi\eta) \end{array}\right\}\cdots\cdots\cdots\text{(VI. 8)}$$

とす.

例 1. 第45表(前章,例 6)の相關表より相關係數を計算せよ.

X も Y も不連續變數ではあるが,連續變數の階級中央の値が與へられたる時と同樣に取扱ふ.

X の原點 A を階級 6, Y の原點 B を階級 7 に選んだ. 次に運算に移るのであるが,先づ運算表(第50表)を作る. 相關表の各欄とその度數を寫し,第一行と第一列とには, $\xi\eta$ を附號を附して記入する. 次で,各欄の $(\xi.\eta)$ を求め,

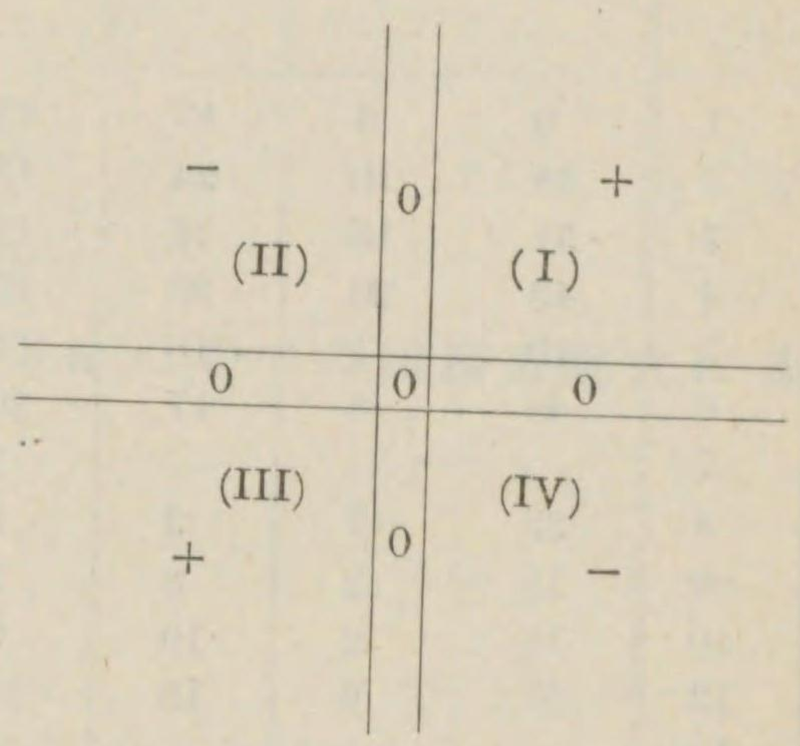
第 50 表

相關係數の運算表. (I) a表.

(母の生める子供數と娘の生める子供數との相關.)

η \ ξ	−6	−5	−4	−3	−2	−1	$\boxed{6}$ 0	1	2	3	4	5	6	7
9			1 (−36)											
8													1 (48)	
7									1 (14)		1 (28)			
6					1 (−12)	2 (−6)	1 (0)	1 (6)	2 (12)				1 (36)	2 (42)
5	3 (−30)	2 (−25)		3 (−15)	5 (−10)	2 (−5)	3 (0)	1 (5)	1 (10)	1 (15)	1 (20)			
4	2 (−24)	2 (−20)	2 (−16)	2 (−12)	1 (−8)	3 (−4)	3 (0)	2 (4)	4 (8)	2 (12)			1 (24)	1 (28)
3	3 (−18)	3 (−15)	4 (−12)	8 (−9)	2 (−6)	8 (−3)	5 (0)	10 (3)	3 (6)	2 (9)	2 (12)	2 (15)		
2	6 (−12)	5 (−10)	2 (−8)	4 (−6)	14 (−4)	14 (−2)	7 (0)	5 (2)	12 (4)	3 (6)	3 (8)	1 (10)		
1	9 (−6)	8 (−5)	3 (−4)	10 (−3)	10 (−2)	13 (−1)	14 (0)	8 (1)	5 (2)	5 (3)	6 (4)		1 (6)	
$\boxed{7}$ 0	8 (0)	9 (0)	9 (0)	13 (0)	18 (0)	12 (0)	15 (0)	9 (0)	8 (0)	7 (0)	4 (0)		1 (0)	
−1	15 (6)	13 (5)	15 (4)	14 (3)	15 (2)	9 (1)	12 (0)	13 (−1)	9 (−2)	4 (−3)	3 (−4)	1 (−5)		1 (−7)
−2	21 (12)	10 (10)	18 (8)	9 (6)	21 (4)	23 (2)	15 (0)	4 (−2)	9 (−4)	3 (−6)	1 (−8)	1 (−10)	3 (−12)	2 (−14)
−3	18 (18)	15 (15)	15 (12)	11 (9)	17 (6)	17 (3)	11 (0)	8 (−3)	12 (−6)	3 (−9)	2 (−12)	1 (−15)	2 (−18)	
−4	11 (24)	14 (20)	10 (16)	16 (12)	19 (8)	7 (4)	8 (0)	3 (−4)	4 (−8)	4 (−12)	1 (−16)	2 (−20)	1 (−24)	
−5	9 (30)	5 (25)	9 (20)	10 (15)	5 (10)	6 (5)	5 (0)	4 (−5)	2 (−10)				2 (−30)	
−6	5 (36)	12 (30)	9 (24)	5 (18)	5 (12)	7 (6)	4 (0)	5 (−6)	1 (−12)					

これを度數の傍に小文字で記入する. 又, A 階級と B 階級とは,表の如く二重線で境界を明にして置く(第50表). A 階級及び B 階級では $(\xi.\eta)$ は 0 となるが,その他では $(\xi.\eta)$ は正又は負である. 第48圖の (I) と (III) とでは $(\xi.\eta)$ は正であるが, (II) と (IV) では負である.



第 48 圖 $\xi.\eta$ の附號を示す圖.
I と III とでは $\xi\eta$ は正,
II と IV とでは負となるを示す.

次に, $(f.\xi.\eta)$なる積を求め,その總和 $\Sigma(f.\xi.\eta)$ を求めるのであるが,その爲には第51表の如き運算表を作るのが安全である.

第51表はその左端の列に, $(\xi.\eta)$の絕對値を順次記入する. 次に第48圖の(III)及び(I)象限の各欄の度數 f を $(\xi.\eta)$に相當する行に記入し,第二列及び第三列を総る.

(II) 及び (IV) 象限に於ける各欄の度數を,第四及び第五列に記入する. この所は共に $(f.\xi.\eta)$ が負である. 次に(III)と (I) の度數の和より (II) と (IV) の度數を減じたる殘りを,正ならば第六列に,負ならば第八列に記入する.

この正負の度數の差と $(\xi.\eta)$ との積を求め,正ならば第七列に,負ならば第九列に記入する. かくて第七列の總和と第九列の總和との差を求めれば,それが $\Sigma(f.\xi.\eta)$ 卽 Np' となる.

p' 以下の計算は,第51表の末尾の運算で明であらう. d 及び d' の正負,從て積 dd' の正負,次には p' 及び p の正負に注意せよ.

第 51 表

相關係數の運算表. (I) b 表.

母の生める子供の數と娘の生める子供の數との相關.

$(\xi.\eta)$	區分III	區分I	區分II	區分IV	+		−	
	+	+	−	−	$f°$ (I)+(III)−(II)−(IV)	$(f°.\xi.\eta)$ f°正ナル場合	$f°$ (I)+(III)−(II)−(IV)	$(f°.\xi.\eta)$ f°負ナル場合
1	9	8	13	13	—	—	9	9
2	38	10	24	13	11	22	—	—
3	31	15	18	12	16	48	—	—
4	43	20	20	15	28	112	—	—
5	19	1	10	5	5	25	—	—
6	48	8	17	20	19	114	—	—
7	—	—	—	1	—	—	1	7
8	37	7	3	5	36	288	—	—
9	11	2	8	3	2	18	—	—
10	15	2	10	3	4	40	—	—
12	57	6	13	10	40	480	—	—
14	—	1	—	2	—	—	1	14
15	25	3	6	1	21	315	—	—
16	10	—	2	1	7	112	—	—
18	23	—	3	2	18	324	—	—
20	23	1	2	2	20	400	—	—
24	20	1	2	1	18	432	—	—
25	5	—	2	—	3	75	—	—
28	—	2	—	—	2	56	—	—
30	21	—	3	2	16	480	—	—
36	5	1	1	—	5	180	—	—
42	—	2	—	—	2	84	—	—
48	—	1	—	—	1	48	—	—
						+ 3653		− 30

$N = 1000$,

$M_x = 4{\cdot}335$,

$M_y = 5{\cdot}898$,

$A = 6$,

$B = 7$,

$\sigma_x = 2{\cdot}984$,

$\sigma_y = 2{\cdot}830$,

$\Sigma(f\xi\eta) = 3653 - 30 = 3623$,

$p' = 3{\cdot}623$,

$d = -1{\cdot}665$,

$d' = -1{\cdot}102$,

$dd' = +1{\cdot}836$,

$p = 3{\cdot}623 - 1{\cdot}836 = 1{\cdot}787$,

$$r = +\frac{1{\cdot}783}{2{\cdot}984 \times 2{\cdot}830} = +0{\cdot}212,$$

$m_1 = +0{\cdot}223 \quad m_2 = +0{\cdot}201$.

p の附號を以て r 及び m_1, m_2 の附號とす.

さて計算により, $r = +0{\cdot}212$,

$$m_1 = +0{\cdot}223,$$

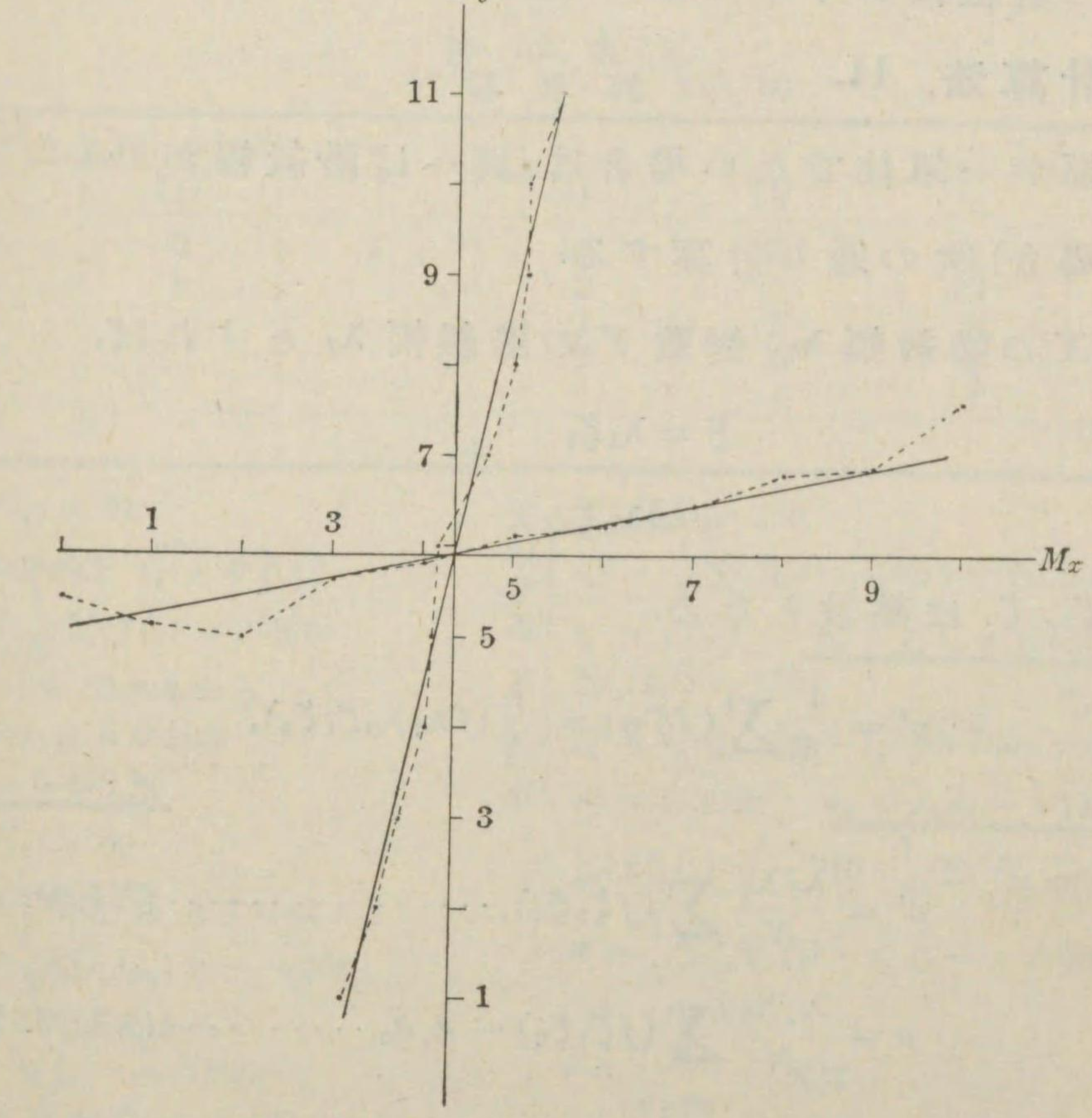
$$m_2 = +0{\cdot}201$$

となつた. 故に回歸方程式は

$$\bar{x} = 0{\cdot}223y,$$

$$\bar{y} = 0{\cdot}201x,$$

又は,偏差 x, y でなく觀測値 X, Y につきての回歸方程式に改むれば,



第 49 圖 母の生める子供の數 (Y) と娘の生める子供の數 (X) との相關(第 45 表).

$$\bar{X}-4{\cdot}335=0{\cdot}223(Y-5{\cdot}898),\ \cdots\cdots\cdots\cdots\cdots\cdots(\mathrm{a})$$

$$\bar{Y}-5{\cdot}898=0{\cdot}201(X-4{\cdot}335),\ \cdots\cdots\cdots\cdots\cdots\cdots(\mathrm{b})$$

(a)式はYの値よりXの平均を計算する式であり,(b)はXよりYの平均を計算する式である.

又,この回歸方程式より,回歸線RR'及びCC'を圖示せんとならば,

$$m_1=\tan\alpha,$$

$$m_2=\tan\beta$$

とし,分度器でα,βを定め,原點を通る直線を畫くのである. 第49圖の二直線はかくして描いたのである.

9. 計算法 II.

階級幅が一單位でない場合は(例へば階級幅が 0.5, 2, 3, 5, 10, 50等の場合),次の通り計算する.

變數Xの階級幅λ_1, 變數Yの階級幅λ_2とすれば,

$$\xi=\lambda_1\zeta_1$$

$$\eta=\lambda_2\zeta_2$$

とす. $\zeta_1,\ \zeta_2$は整數となる.

又　$$p'=\frac{1}{N}\sum(f\xi\eta)=\frac{1}{N}(f\lambda_1\lambda_2\zeta_1\zeta_2),$$

即ち,

$$p'=\frac{\lambda_1\lambda_2}{N}\sum(f\zeta_1\zeta_2),\cdots\cdots\cdots\cdots\cdots\cdots(\mathrm{VI.\ 9.\ a})$$

$$p=\frac{\lambda_1\lambda_2}{N}\sum(f\zeta_1\zeta_2)-d_1d_2,\cdots\cdots\cdots\cdots(\mathrm{VI.\ 9.\ b})$$

例 2. 前章,例 4 に於ける穀粒の重さと,その含有脂肪百分比との相關表より, $r,\ m_1,\ m_2$を計算せよ

第 52 表甲
r の 運 算 表 (II. a)

6·25

ζ_2 \ ζ_1	−3	−2	−1	0	+1	+2	+3	+4	Yの分布
+3			1 (−3)						1
+2		2 (−4)	1 (−2)	1 (0)					7
+1		1 (−2)	12	11 (0)	2 (1)				26
0	1 (0)	2 (0)	10 (0)	48 (0)	37 (0)	8 (0)	1 (0)		107
−1		1 (2)	6 (1)	22 (0)	33 (−1)	10 (−2)	2 (−3)	1 (−4)	75
−2					8 (−2)	2 (−4)	1 (−6)		11
Xの分布	1	6	30	82	80	20	4	1	224

42·5 (left)　42·5 (right)

6·25

第 52 表乙
r の 運 算 表 (II. b)

	+		−		−	
	III	I	II	IV	f_0	$f_0\zeta_1\zeta_2$
1	6	2	12	33	37	37
2	1		2	18	19	38
3			1	2	3	9
4			2	3	5	20
5				1	1	6
						−110

Xの分布より

$\Sigma(f\zeta_1)=91$

ζ_1 の平均を $\bar{\zeta}_1$ とすれば

$\bar{\zeta}_1=\dfrac{1}{N}\Sigma(f\zeta_1)=0{\cdot}406$

$\lambda_1=0{\cdot}5$ であるから

$d_1=\lambda_1\zeta_1=0{\cdot}203$

$\underline{M_x=6{\cdot}453\ \%}$

Yの分布より

$\Sigma(f\zeta_2)=60$

ζ_2 の平均を $\bar{\zeta}_2$ とすれば

$\bar{\zeta}_2=\dfrac{1}{N}\Sigma(f\zeta_2)=-0{\cdot}268$

$\lambda_2=5$ であるから

$d_2=\lambda_2\zeta_2=-1{\cdot}340$

$M_y=B+d_2=41{\cdot}160$ mg.

又 $\bar{\zeta}_1\bar{\zeta}_2=-0{\cdot}406\times0{\cdot}268$

$=-0{\cdot}109$

Xの分布

$\Sigma(f\zeta_1^2)=285$

$s_1^2=\dfrac{1}{N}\Sigma(f\zeta_1^2)-\bar{\zeta}_1^2=1{\cdot}227-0{\cdot}165=1{\cdot}062$

$s_1=1{\cdot}031$　$\underline{\sigma_x=s_1\lambda_1=0{\cdot}516}$

Yの分布

$\Sigma(f\zeta_2^2)=170$

$s_2^2=\dfrac{1}{N}\Sigma(f\zeta_2^2)-\bar{\zeta}_2^2=0{\cdot}759-0{\cdot}072=0{\cdot}687$

$s_2=0{\cdot}8288$　$\underline{\sigma_y=\lambda_2s_2=4{\cdot}145}$

$\Sigma(f\zeta_1\zeta_2)=-110$　$\dfrac{1}{N}\Sigma(f\zeta_1\zeta_2)=-0{\cdot}491$

$\pi=\dfrac{1}{N}\Sigma(f\zeta_1\zeta_2)-\bar{\zeta}_1\bar{\zeta}_2=-0{\cdot}491+0{\cdot}109$

$=-0{\cdot}382$

$r=\dfrac{p}{\sigma_x\sigma_y}$　又は

$r=\dfrac{\pi}{s_1s_2}=\dfrac{-0{\cdot}382}{1{\cdot}031\times0{\cdot}8288}$

$\underline{r=-0{\cdot}447}$

計算の順序として,第52表甲の如く寫し取り,原點 A, B を定める. 今は A を $6.0-6.5$ の階級の中央の値(卽ち6·25)に選び, B を $40-45$ の中央の値(卽ち42·5)に選ぶ.

例1と同樣に $(\zeta_1 \cdot \zeta_2)$ の値と附號を記入する.

この表より乙なる運算表を作る. 第一列には $(\zeta_1 \cdot \zeta_2)$ を1より順次記載したのである. 第二列及び第三列は $(\zeta_1 \cdot \zeta_2)$ の正なる欄の度數を,第四列及び第五列は $(\zeta_1 \cdot \zeta_2)$ の負なる欄の度數を記入する. 第六列では(III)の度數と(I)の度數の和より(II)と(IV)の度數の和を減じたるものを記入する. この差 f_0 はこの例ではすべて負である. 第七列は f_0 に $(\zeta_1 \cdot \zeta_2)$ を乘じたるものを記入し,最後に $(f_0\zeta_1\zeta_2)$ の總和を求めたのである.

かくて, $\Sigma(f\zeta_1\zeta_2)$ を求めることが出來たが,負號を忘れてはならぬ. d_1, d_2 の代りに $\bar{\zeta}_1, \bar{\zeta}_2$ をとるが,これも負であることに注意せよ. 但し $d_1 = \lambda_1\bar{\zeta}_1,\ d_2 = \lambda_2\bar{\zeta}_2$ とす.

これよりの順序は運算表乙の末尾にある通りの順序で明であらう.

$$p = \lambda_1\lambda_2\pi, \quad \sigma_x = \lambda_1 s_1, \quad \sigma_y = \lambda_2 s_2.$$

であるから,

$$r = \frac{p}{\sigma_x\sigma_y} = \frac{\pi}{s_1 s_2}$$

から計算する. (π は圓周率にあらず.)

例 3. 第41表鷄卵の重量と孵化せる雛體重との相關表より, r, m_1, m_2 を計算せよ.

例2と同一であるから運算表のみを掲げる.

第 53 表

r の運算表, III 甲.

						40				
		−4	−3	−2	−1	0	1	2	3	4
	5								1(15)	3(20)
	4								2(12)	3(16)
	3							3 (6)	3 (9)	2(12)
	2						6 (2)	20 (4)	4 (6)	
	1				1(−1)	9	24 (1)	16 (2)		
57	0				1	24	15	2		
	−1				18 (1)	19	5(−1)			
	−2			7 (4)	14 (2)	4				
	−3		1 (9)	12 (6)	4 (3)					
	−4		8(12)	10 (8)	1 (4)					
	−5	4(20)	9(15)							
	−6	5(24)								

第 53 表

r の運算表 III 乙.

$\zeta_1\zeta_2$	+		−		f	$f_0\zeta_1\zeta_2$
	III	I	II	IV		
1	18	24	1	5	36	36
2	14	22			36	72
3	4	—			4	12
4	8	20			28	112
6	12	7			19	114
8	10	—			10	80
9	1	3			4	36
12	8	4			12	114
15	9	1			10	150
16	—	3			3	48
20	4	3			7	140
24	5	—			5	120
計					$\Sigma(f\zeta_1\zeta_2) = 1064$	

$M_x = 40{\cdot}05$

$A = 40$

$d_1 = 0{\cdot}05$

$M_y = 55{\cdot}92$

$B = 57{\cdot}00$

$d_2 = -1{\cdot}08$

$\sigma_x = 3{\cdot}71$

$\sigma_y = 4{\cdot}72$

$\Sigma(f\zeta_1\zeta_2) = 1064$

$$p' = \frac{\lambda_1\lambda_2}{N}\Sigma(f\zeta_1\zeta_2) = \frac{2\times 2}{260}\times 1064$$

$p' = 16{\cdot}38$

$$p = p' - d_1 d_2 = 16{\cdot}38 - 0{\cdot}054 = 16{\cdot}33$$

$$r = \frac{p}{\sigma_x\sigma_y} = \frac{16{\cdot}33}{3{\cdot}71\times 4{\cdot}72} = 0{\cdot}931$$

即ち　$r = 0{\cdot}931$

$m_1 = 0{\cdot}733$

$m_2 = 1{\cdot}186$

回歸方程式

$$\left.\begin{aligned}\bar{x} &= 0{\cdot}733\,y\\ \bar{y} &= 1{\cdot}186\,x\end{aligned}\right\}$$

又は

$$\left.\begin{aligned}X - 40{\cdot}05 &= 0{\cdot}733(Y - 55{\cdot}92)\\ Y - 55{\cdot}92 &= 1{\cdot}186(X - 40{\cdot}05)\end{aligned}\right\}$$

10. 各行及び各列の標準偏差

X變數の標準偏差 σ_x, Y の標準偏差 σ_y は,共に,單一變數の度數分布表から簡單に計算することが出來るが,この σ_x, σ_y は各行及び各列の標準偏差とは必ずしも同一でない.

普通は各行若くは各列の標準偏差は, σ_x 若くは σ_y より小である. 今,これを σ_{ax} 及び σ_{ay} で表はすならば

$$\sigma_{ax} \leqq \sigma_x$$

$$\sigma_{ay} \leqq \sigma_y$$

なるが普通である. 何となれば, σ_{ax} は各行の算術的平均を基準とするが故に,最小なる平均平方根偏差となるに反して,全分布の標準偏差 σ_x は M_x を基準として計算するが故に,各行の最小なる平均平方根偏差を取れないが爲である. σ_{ay}につきても同樣の理由がある. 但し無相關又は殆ど無相關なる時は,各行及び各列の平均は全體の平均M_x及びM_yに殆ど等しく,

$$\sigma_{ax} \fallingdotseq \sigma_x,\qquad \sigma_{ay} \fallingdotseq \sigma_x$$

となる. 故に, σ_{ax} と σ_x, 又 σ_{ay} と σ_y とを比較することによりて,相關を概略推知することが出來る. 但し殆ど無相關なる時は偶然の影響の爲に $\sigma_{ax} > \sigma_x$ 等となることがある.

例 4. 前章,例 1 に於ける各行及び各列の標準偏差を計算せよ(第40表).

各行及び各列の標準偏差は,通例の度數分布表に於けると同樣に計算するのであるが,計算の結果は次の表の通りとなる.

第 54 表

各行及各列の標準偏差

各行の標準偏差.

Y の階級	σ_{ax}
100·45—102·45	3·20
98·45—100·45	2·56
96·45— 98·45	2·69
94·45— 96·45	2·78
92·45— 94·45	2·95
90·45— 92·45	3·14
88·45— 90·45	3·06
86·45— 88·45	2·92
84·45— 86·45	2·84
82·45— 84·45	2·71
80·45— 82·45	3·43
平　均	2·935

各列の標準偏差.

X の階級	σ_{ay}
152·45—154·45	2·33
154·45—156·45	2·53
156·45—158·45	2·00
158·45—160·45	1·76
160·45—162·45	2·19
162·45—164·45	2·01
164·45—166·45	2·19
166·45—168·45	2·13
168·45—170·45	2·18
170·45—172·45	2·09
172·45—174·45	1·99
174·45—176·45	2·69
176·45—178·45	2·14
178·45—180·45	1·57
180·45—182·45	2·13
平　均	2·129

觀測數の少い行もあるから, σ_{ax} は悉くは等しくないが,殆ど一致し,凡そ2·9である. 又これらの σ_{ax} は X の全分布の標準偏差 σ_x に比し著しく小である ($\sigma_x = 5{\cdot}59$). σ_{ay} につきても亦同

様で,各列に於ける標準偏差 σ_{ay} は殆ど相等しく,凡2·1であるが $\sigma_y = 4{\cdot}07$ に比し著しく小である.

例 5. 前章,例4に於ける相關表から,各行及び各列の標準偏差を計算し,これを比較せよ.

計算の結果は第55表の通りとなる. σ_{ax}は概して相等しく,又その平均0·248は σ_x の0·248と等しい. σ_{ay}もまた概ね等しく,その平均3·253は $\sigma_y = 3{\cdot}21$ に殆ど等しい. かく,殆ど無相關なる場合は, σ_{ax} 及び σ_{ay} は σ_x 及びσ_yに殆ど等しくなるのである.

第 55 表

Yの階級	σ_{ax}
29·955	0·225
27·955	0·259
25·955	0·259
23·955	0·232
21·955	0·256
19·955	0·256
17·955	0·244
15·955	0·237
13·955	0·248
11·955	0·258
9·955	0·255
平 均	0·248

Xの階級	σ_{ay}
7·005	3·85
7·105	3·61
7·205	3·21
7·305	3·51
7·405	3·23
7·505	3·22
7·605	3·16
7·705	3·29
7·805	3·23
7·905	2·86
8·005	3·26
8·105	3·31
8·205	3·09
8·305	2·71
平 均	3·253

各行又は各列の標準偏差に代ふるに,回歸方程式より計算せる値, $\bar{X}$又は$\bar{Y}$を基準とする平均平方根偏差を以てすることが出來る. これを S_{ax} 及びS_{ay}とすれば, S_{ax} 又は S_{ay} は σ_{ax}, σ_{ay} より稍大であつて,次の如く簡單に計算し得る.

$$S_{ax} = \sqrt{\sigma_{ax}^2 + (M_{ax} - \bar{X})^2}$$

但しM_{ax}及びM_{ay}は各行及び各列の平均,又 $\bar{X}$ 及び $\bar{Y}$ は(VI. 1. b)より計算せる値である.

11. 回歸線を基準とせる標準偏差(最小標準偏差)

今,階級幅大ならずとし, y_t なる階級につきて(即ち y_t なる行につきて), その偏差 x の平均 $\bar{x}_t$ を求むれば,一次の相關では

$$\bar{x}_t = m_1 y_t$$

で計算することが出來る. 故にこの行に於ける S_{ax}^2 は

$$n_t S_{ax}^2 = \Sigma(x-\bar{x}_t)^2 = \Sigma(x - m_1 y_t)^2,$$

但し,x はその行に於ける觀測値 X の偏差 $(X - M_x)$ である.

$n_t S_{ax}^2$を各行につきて計算すれば,

y_1 行につきては $\Sigma(x - m_1 y_1)^2$,

y_2 行につきては $\Sigma(x - m_1 y_2)^2$,

y_3 行につきては $\Sigma(x - m_1 y_3)^2$,

…………………………

y_t 行につきては $\Sigma(x - m_1 y_t)^2$,

…………………………

但し,Σは各行に於ける總和を意味する. 今これらの式の總計を求むる時は,

$$\Sigma x^2 + m_1^2\Sigma y^2 - 2m_1\Sigma xy \quad\cdots\cdots\cdots\cdots (A)$$

但し(A)式のΣは表全體についてx^2等の總和を意味する. 故に,回歸線 RR′ を基準とする標準偏差をS_xとすれば,

$$S_x^2 = \sigma_x^2 + m_1^2\sigma_y^2 - 2m_1 r\sigma_x\sigma_y,$$

しかるに, $m_1 = \frac{\sigma_x}{\sigma_y} r$ であるから,

$$S_x^2 = \sigma_x^2 + r^2\sigma_x^2 - 2r^2\sigma_x^2,$$

$$\therefore \quad S_x = \sigma_x\sqrt{1-r^2} \quad\cdots\cdots\cdots (VI. 10. a)$$

但し S_x は回歸方程式 $\bar{x} = m_1 y$, または回歸直線 RR′ を基準とせる標準偏差である. 同様に回歸方程式 $\bar{y} = m_2 x$, または回歸

線 CC′ を基準とせる標準偏差 S_y は,

$$S_y=\sigma_y\sqrt{1-r^2}, \cdots\cdots\cdots\cdots\cdots \text{(VI. 10. b)}$$

である. S_x 及び S_y を又**最小標準偏差**と云ふことがある.

12.　S_x 及び S_y の性情

(i)　r の正負にかかはらず, S_x 及び S_y は常に正の値を取らしむ.

(ii)　無相關なる時は, S_x 及び S_y は σ_x 若くは σ_y に等しい(或は殆ど等しい). 何となれば $r=0$ であるから.

(iii)　正又は負の完全相關の時は, S_x 及び S_y は 0 である.

何となれば,　　$r=\pm 1$

なるが爲に.

(iv)　無相關にもあらず,完全相關にもあらざる時は,

$$S_x<\sigma_x$$

$$S_y<\sigma_y$$

である. 故に一般に S_x 及び S_y につきては次の不等式が成立する.

$$\left.\begin{array}{l} 0\leqq S_x\leqq\sigma_x \\ 0\leqq S_y\leqq\sigma_y \end{array}\right\}\cdots\cdots\cdots\cdots\cdots \text{(VI. 11)}$$

(v)　正常二重分布では,各行の標準偏差 σ_{ax}, 及び $\bar{x}$ を基準とする各行の標準偏差 S_{ax} は,最小標準偏差 S_x に等しく,各列の標準偏差 σ_{ay} 及び $\bar{y}$ を基準とする各列の標準偏差 S_{ay} は,最小標準偏差 S_y に等しい. 又,正常に近い二重分布では殆ど等しい.

このことは正常二重分布の重要なる性質の一であつて,たとひ回歸が一次であつて相關面が相稱であつても,この法則に當

辧らないものは正常相關ではない. けれどもその樣な例は自然界には極めて稀に出現するのみであるから,回歸一次なるもの,又は相關面の相稱なるものは正常なりと考へても,事實上差支はないのである.

正常二重分布の定義をすれば次の通りとなる.

一次の相關であつて,各行及び各列の標準偏差が互に等しく且つ最小標準偏差 S_x 及び S_y に等しき時は此の分布を正常二重分布と云ふ.

例 6.　前章,例 1 の相關表より S_x, S_y を計算し,本章例 4 の σ_{ax}, σ_{ay} と比較せよ.

$$\sigma_x=5{\cdot}590,\qquad r=0{\cdot}847,$$

$$S_x=\sigma_x\sqrt{1-r^2}=5{\cdot}590\sqrt{1-0{\cdot}847^2}$$

卽ち,　　$=5{\cdot}590\times 0{\cdot}531=2{\cdot}97$

となり,例 4 に於ける σ_{ax} の平均 2·935 によく一致する. (σ_{ax} の方が S_x より普通小であるのは何に基因するかを考察せよ.)

又,　　$\sigma_y=4{\cdot}073,\quad r=0{\cdot}847,$

$$S_y=\sigma_y\sqrt{1-r^2}=4{\cdot}073\times 0{\cdot}531,$$

$$=2{\cdot}16$$

となり,例 4 に於ける σ_{ay} の平均 2·129 に極めてよく一致して居る. 尙ほ,本章,例 5 に於ては $r\fallingdotseq 0$ であるから, σ_{ax} の平均, σ_{ay} の平均は共に S_x, S_y に殆ど等しい. 序に計算して見られよ.

13.　相關係數の近似計算法

相關係數は $\Sigma(fxy)$, $\Sigma(f\xi\eta)$, $\Sigma(f\zeta_1\zeta_2)$ より公式(VI. 6), (VI. 8), (VI. 9)等により計算するのであるが,若し, m_1, m_2 が與へられて

居る時は,公式(VI. 5)より計算する.

しかしこれらの公式によらず,回歸線が圖示せられて居る時,又は單に相關表のみが與へられて居るだけでも,經驗家は容易に r を概算することが出來る.

(i) 若し m_1, m_2 の代りに,各行及び各列の平均を結ぶ屈折線,(第40圖等の點線)又は回歸線 RR′, CC′ が與へられて居る時は,分度器を用ひて α, β を測り,容易に, $m_1=\tan\alpha$, $m_2=\tan\beta$ を知り, r を

$$r=\pm\sqrt{m_1 m_2}$$

にて計算することが出來る.

又,回歸線 CC′ 及び RR′ 上に各一點をとり,垂線を X 軸及び Y 軸に引きて, x, y 及び x', y' を求め,

$$r=\pm\sqrt{\frac{y}{x}\cdot\frac{x'}{y'}}$$

とし,附號は第 3 條の規定に從ひて決定する.

(ii) 若し單に, Scattered diagram のみが與へられて居る時,極めて概略の値を得んとすれば,各行の分布の凡そ中央と思はるる點を圖上に取り,これを平均と考へ,この點に最も近く平均點の約半數が一側に他の半數が他側にある樣に絲を張り,次に大體この絲の位置に直線を引きて回歸線 RR′ とす. CC′ も同樣に定め(i)の方法により容易に r を概算することが出來る.

(iii) 又,觀測數の少數ならざる二列及び二行につき算術的平均を計算し,この二對の點を結合して RR′, CC′ と定めても概算することは困難でない.

讀者は,それのみならず,相關表を見ただけで r の近似的の値

を知ることの出來る樣に熟練せねばならぬ.

(iv) 各行及び各列の標準偏差 σ_{ax}, σ_{ay}, 又は其の平均が與へられて居る時は,正常分布ではこれを用ひて r を計算することが出來る. 何となれば,各行及び各列の標準偏差は概略回歸線を基準とせる標準偏差に等しいが故に,

$$\sigma_{ax}\fallingdotseq S_x, \qquad \sigma_{ay}\fallingdotseq S_y,$$

$$S_x=\sigma_x\sqrt{1-r^2}$$

$$\left.\begin{aligned} \therefore\quad r&=\pm\sqrt{1-\frac{S_x^2}{\sigma_x^2}}\fallingdotseq\pm\sqrt{1-\frac{\sigma_{ax}^2}{\sigma_x^2}}. \\ \text{又}\quad r&=\pm\sqrt{1-\frac{S_y^2}{\sigma_y^2}}=\pm\sqrt{1-\frac{\sigma_{ay}^2}{\sigma_y^2}} \end{aligned}\right\}\cdots\cdots(\text{VI. 10. c})$$

故に本章,例 4,例 5 に於ける σ_{ax}, σ_{ay} は反對に r の概算をする時に用ふることが出來るが,次に他の例をひく.

例 7. 前章に於ける父とその男兒との身長の相關表から,各行の標準偏差を計算し,これより r を概算せよ.

計算. 分布の兩端にある度數の少い階級は,この計算より省いた方が安全である. さてこの表の Y の各階級につきて σ_{ax} を計算した所,第56表の示す通りであつた. これより σ_{ax} の平均を求めて 2·359, r の概算は公式(VI. 10. c)により,

第 56 表 各行の標準偏差

Y の階級	各行の標準偏差 σ_{ax}
62·5—63·5	2·56
63·5—64·5	2·11
64·5—65·5	2·55
65·5—66·5	2·24
66·5—67·5	2·23
67·5—68·5	2·60
68·5—69·5	2·26
69·5—70·5	2·26
70·5—71·5	2·45
71·5—72·5	2·33
σ_{ax} の平均は	2·359

$$r'=\sqrt{1-\left(\frac{2\cdot 359}{2\cdot 75}\right)^2}$$
$$=0\cdot 514.$$

$$r' \fallingdotseq 0{\cdot}514$$

で眞の r と同一であつた.

14. 相關比

前條(iv)に述べたる r の概算法は,正常相關の場合に適用せらるるものである. 何となれば,各行若くは各列の標準偏差 σ_{ax}, σ_{ay} が近似的に最小標準偏差 S_x, S_y に等しき場合のみに適用するからである. 正常ならざる二重分布,就中,回歸が一次でない時は,この方法で計算せる r' は r に一致しない. 正常相關ならざる時は,相關係數 r は不當に小なることが多い. 例へば前章例 8 に於ては,明なる相關が存するに拘らず $r \fallingdotseq 0{\cdot}14$ であつた. かくの如き時は r は相關の程度を忠實に現はさない憾がある. 然るに,

$$r' = \sqrt{1 - \frac{\sigma_{ay}^2}{\sigma_y^2}}$$

は可なり大であつて,相關を稍忠實に表示する. しかし σ_{ay}^2 に代ふるに,σ_{ay}^2 の重みつけられたる平均を以てするのは一層合理的である.

$$S_{ay}^2 = \frac{1}{N}\sum(n\sigma_{ay}^2),$$

σ_{ay} は各列の標準偏差, n はその列の觀測數, N は觀測總數である. この S_{ay}^2 より,

$$\eta_y = \sqrt{1 - \frac{S_{ay}^2}{\sigma_y^2}}$$

により計算せる η_y を**相關比** (Correlation ratio) と云ふ. 各行につきても,同様の計算を行ふことが出來るから,次の 2 個の相關比が計算せられる.

$$\left.\begin{aligned} \eta_x &= \sqrt{1 - \frac{S_{ax}^2}{\sigma_x^2}}, \\ \eta_y &= \sqrt{1 - \frac{S_{ay}^2}{\sigma_y^2}}, \end{aligned}\right\} \cdots\cdots\cdots\cdots (\text{VI. } 12)$$

さて,各行の標準偏差の平方 σ_{ax}^2 の重みづけられたる平均 S_{ax}^2 は最小標準偏差の平方より小である.

$$S_{ax}^2 < S_x^2,$$

故に
$$\sqrt{1 - \frac{S_{ax}^2}{\sigma_x^2}} > \sqrt{1 - \frac{S_x^2}{\sigma_x^2}},$$

$$\left.\begin{aligned} &\text{卽ち,} \quad |\eta_x| > |r|, \\ &\text{同様に,} \quad |\eta_y| > |r|, \end{aligned}\right\} \cdots\cdots\cdots\cdots\cdots\cdots (\text{VI. } 13)$$

でなければならぬ. 而して η と r との差は,回歸が一次より遠ざかるに從つて益々大となる.

例 8. 前章,例 8 の相關表より相關比 η_y を計算せよ.

計算は第57表の如く行ふことが出來る.

15. 相關指數

散布度を計るに,標準偏差は,各觀測値の偏差 x の平方の算術的平均の平方根を以てした. 斯く x を平方にすることを,餘りに數學的人爲的であつて,不自然又複雜であるとし, Lenz 等は標準偏差に代ふるに平均偏差を以てした. Lenz は同じ批難を相關係數にも向けて居る. 相關係數の計算には σ_x, σ_y を用ひ,又偏差の積 (xy) の總和を求めるのであるが,かくの如き代數的方法は不自然であるから,算術的方法を用ひねばならぬと云ふ. 卽ち,彼は**相關指數** (Der Korrelationsindex), k を以て相關係數 r に代へんとするものである.

第 57 表

η_y の 運 算 表

階級	各列の算術的平均 M_{ay}	各列の總度數	一定の原點よりの Σa^2	nd^2	$(\Sigma a^2 - nd^2)$
466(464·5—467·5)	2·000	1	0		—
469	—	—	—		—
472	2·000	1	0		—
475	—	—	—		—
478	—	—	—		—
481	2·000	2	0		—
484	4·000	2	80	—32	48
487	2·000	2	0		—
490	4·000	4	160	—64	96
493	3·333	6	192	—67	125
496	4·400	15	736	—290	446
499	5·556	18	1760	—556	1204
502	5·913	46	4112	—1608	2504
505	16·327	98	79216	—46080	33136
508	24·784	125	110048	—37019	73029
511	16·326	129	74448	—24097	50351
514	10·318	88	12592	—58	12534
517	8·468	47	4240	—30	4210
520	6·615	26	960	—13	947
523	5·273	11	160	— 0	160
526	5·000	4	144	— 9	135
529	2·000	1	0		—
532	2·000	1	0		—
535	3·333	3	16	— 2	14
538	—	—	—		—
541	2·000	1	0		—
544	2·000	1	0		—
		632		$\Sigma =$	178939
			$S_y^2 = \dfrac{\Sigma}{632} = 283{\cdot}1.$		

$M_y = 14{\cdot}036,$

$\sigma_y^2 = 1608{\cdot}7 - 1293{\cdot}5 = 315{\cdot}2,$

$$\eta_y = \sqrt{1 - \frac{283{\cdot}1}{315{\cdot}2}} = \sqrt{1 - 0{\cdot}898} = \sqrt{0{\cdot}102} = 0{\cdot}319.$$

$$k = \frac{1}{2}\left\{\frac{\Sigma x_1 + \Sigma x_3 - \Sigma x_2 - \Sigma x_4}{\Sigma|x|} + \frac{\Sigma y_1 + \Sigma y_3 - \Sigma y_2 - \Sigma y_4}{\Sigma|y|}\right\}. \qquad \cdots\cdots\cdots\cdots(\text{VI. } 14)$$

この式を r の公式に比較すれば, k は積に代ふるに和を以てし,標準偏差に代ふるに平均偏差を以てして居る點が,相關係數と相異して居ることが明になる.

先づ相關表を M_x, M_y を以て四つの象限に分ち(第50圖), 各觀測値につきて, M_x 及び M_y を基準として x 及び y を計算する. 式の Σx_1 と Σy_1 とは第一象限に於ける x 又は y の總和を意味し, Σx_2 及び Σy_2 は第二象限, Σx_3 及び Σy_3 は第三象限, Σx_4 と Σy_4 とは第四象限に於けるそれぞれの偏差の和である.

これら $+\sum x_1,\ -\sum x_2,\ +\sum x_3,\ -\sum x_4$ 及び,

$$+\sum y_1,\ -\sum y_2,\ +\sum y_3,\ -\sum y_4$$

の和を取り,これより k を計算するので, Σx_1 等の附號につきては第50圖に示された通りである.

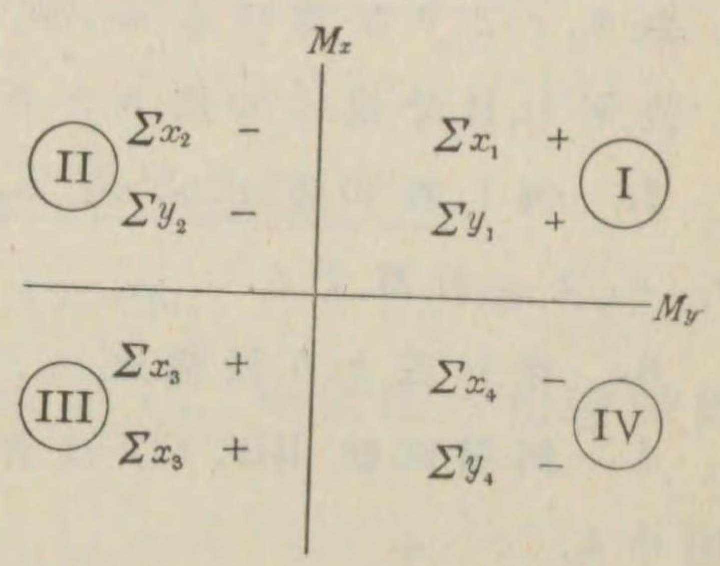
完全なる相關の時は, $k = \pm 1$,

無相關なる時は, $k = 0$

である. その他の場合では一般にその中間にあり. 故に

$$-1 \leqq k \leqq +1$$

である.



第 50 圖

相關指數は,不規則なる觀測値の悪影響を受くること, r の如く甚しくないから, r よりもこの方がよく,又計算の容易なことは同日の比でないと Lenz は云ふ.

相關指數は相關係數に比し,一般には劣つたものではないけれども,代數的取扱の不可能なるが爲に,理論的研究に適しないことは救ふべからざる缺點で,未だ廣く用ひられて居ない. 又計算もLenzの云ふ如く簡單でない.

演習問題

1. 次の相關表より,第40圖にならひ,各行及び各列の平均を結ぶ線を描き,更に回歸直線RR′, CC′をも描き添へよ.

(a) 第41表,鷄卵の重量とその孵化時に於ける雛體重との相關表. 注意! RR′及びCC′が接近し,この二直線の挾む角θの小なることを見よ.

(b) 第43表,英國人男子の頭長徑と視覺に對する反應時間との相關表. 注意! r又はm_1, m_2が與へられず. 從て角α及びβを決定できないから,折線を描きて後,目分量にてRR′, CC′を決定せよ.

(c) 第44表,燕麥の穀粒の重量Yとその含有脂肪百分比との相關表. 注意! 折線を描きたる後目分量にて回歸直線RR′, CC′を描き,$\bar{r}$より計算せるm_1, m_2から角α及びβを計算し, RR′, CC′を決定し,目分量にて描きたるものと比較せよ.

2. 例1,第40表より, m_1, m_2, r, $\bar{r}$, S_x, S_yを計算せよ. 但しS_x, S_yは$\bar{\sigma}_x$, $\bar{\sigma}_y$より計算する.

3. 第44表より相關比, η_x, η_yを計算せよ.

4. 回歸直線, RR′, CC′は共にM_xとM_yの交點を通過することを證明せよ.

本章では,漠然と回歸直線はM_xとM_yとの交點を通過するものとして説明したが,嚴密に云ふならばこの事は證明を要することである.

第七章

相關係數を用ひたる種々の定理

第一項 標準偏差に關する定理

1. 變數の和又は差の標準偏差

單變數の變異統計で,算術的平均と標準偏差とが最も必要であつた如く,相關では相關係數が最も必要な値である. M, σ及びrは何れも代數的取扱が出來,この三つの間には常に種々興味ある又必要なる關係を導くことが出來るのである.

今標準偏差の定理中で,相關係數の關係して居る二三を述べることとする.

定理 I. 二變數XとYとを一對の觀測値とすれば,

$$Z = X \pm Y \quad \cdots\cdots\cdots\cdots\cdots\cdots\text{(A)}$$

に於て,Zの標準偏差σ_zは次の公式で表はされる[1].

$$\sigma_z = \sqrt{\sigma_x^2 + \sigma_y^2 \pm 2r\sigma_x\sigma_y}, \cdots\cdots\cdots\text{(VII. 1)}$$

但し,X及びYの標準偏差をσ_x, σ_yとし,二變數間の相關係數をrとす.

證明. Xの算術的平均をM_x, XのM_xよりの偏差をx,又Yの算術的平均をM_y,偏差をyとす.

Zの算術的平均M_zよりの偏差zは,

---

(1) Zの算術的平均は第二章に説いた.

$$z = x \pm y$$

である．何となれば，(A) より　$M_z + z = M_x + x \pm (M_y + y)$

であるから．

故に　$\sum(z^2) = \sum(x^2) \pm 2\sum(xy) + \sum(y^2)$，

兩邊を觀測値 N で除する時は，

$$\sigma_z{}^2 = \sigma_x{}^2 \pm 2p + \sigma_y{}^2,$$

即ち　$\sigma_z{}^2 = \sigma_x{}^2 \pm 2r.\sigma_x.\sigma_y + \sigma_y{}^2$.········(VII. 1)

若し X と Y との間に相關がなければ，

$r = 0$　　であるから，

$$\sigma_z{}^2 = \sigma_x{}^2 + \sigma_y{}^2. \quad \cdots\cdots\cdots\cdots\text{(VII. 2. a)}$$

この二式は Z が X と Y との和である時にも，差である時にも用ひられる．

若し又，相關が完全なりと見做していい場合は，$r \fallingdotseq \pm 1$ であるから，

$$\sigma_z{}^2 \fallingdotseq (\sigma_x \pm \sigma_y)^2 \quad \cdots\cdots\cdots\cdots\text{(VII. 2. b)}$$

となる．普通は $r = 0$ でもなく，$r = \pm 1$ でもないから，(1) 式を使用する．

人體計測では一對の計測値 X と Y との和又は差を求め，その平均又は標準偏差を求めたいことが甚だ多い．將來，この定理が用ひらるる時期が屢起るであらう．

定理 II.　3 個又はそれ以上の變數 $X_1, X_2, X_3, \cdots\cdots, X_n$ の和又は差の標準偏差 σ は次の式で計算することが出來る．

$$\left.\begin{array}{l}\sigma^2 = \sigma_1{}^2 + \sigma_2{}^2 + \sigma_3{}^2 + \cdots\cdots + \sigma_n{}^2 \pm 2r_{1.2}\sigma_1\sigma_2 \pm 2r_{1.3}\sigma_1\sigma_3 \pm \cdots \\ \quad \pm 2r_{2.3}\sigma_2\sigma_3 \pm 2r_{2.4}\sigma_2\sigma_4 \pm \cdots\cdots\cdots\cdots\cdots\cdots \\ \quad \pm 2r_{3.4}\sigma_3\sigma_4 \pm 2r_{3.5}\sigma_3\sigma_5 \pm \cdots\cdots\cdots\cdots\cdots\cdots \\ \quad \pm \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \quad \pm 2r_{(n-1).n}\sigma_{n-1}\sigma_n.\end{array}\right\} \cdots\text{(VII. 3)}$$

但し，$\sigma_1, \sigma_2, \sigma_3, \cdots\cdots$ は變數 $X_1, X_2, X_3, \cdots\cdots, X_n$ の標準偏差，

又，　$r_{1.2}, r_{1.3}, r_{1.4}, \cdots\cdots, r_{2.3}, r_{2.4}, \cdots\cdots, r_{2.n}$,

$r_{3.4}, r_{3.5}, \cdots\cdots, r_{4.5}, r_{4.6}, \cdots\cdots\cdots\cdots$

$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$

$r_{(n-1)\cdot n}$

はそれぞれ添數に相當する變數間の相關係數である．

2. 觀測誤差が標準偏差に及ぼす影響

今，眞の値を X，誤差を含める計測値を X'，誤差を δ とすれば，

$$X' = X + \delta$$

X', X, δ も共に變數であるが，X と δ との間に相關が無いから[1] (VII. 2. a) を應用して，

$$\sigma_{x'}{}^2 = \sigma_x{}^2 + \sigma_\delta{}^2,$$

即ち，　$\sigma_x = \sqrt{\sigma_{x'}{}^2 - \sigma_\delta{}^2}$,············(VII. 2. c)

但し，$\sigma_{x'}$ は觀測値の標準偏差，σ_x は眞の値の標準偏差，σ_δ は觀測誤差のみの標準偏差である．しかるに，生物統計學では普通 σ_δ は σ_x に比して甚しく小であるから，

$$\sigma_x \fallingdotseq \sigma_{x'}$$

とすることが出來る．

(1) X' と δ との間には大なり小なり僞相關が有るが，X と δ との間には相關が無い．

第二項　重みづけられたる平均

3. 重みづけられたる平均と算術的平均

重みをw,觀測をVとすれば,重みづけられたる平均M'は

$$M' = \frac{\Sigma(wV)}{\Sigma w}, \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \text{(A)}$$

又wの標準偏差σ_w, Vの標準偏差σ_V, wとVとの間の相關係數をrとし,Vの算術的平均M, wの算術的平均$\overline{w}$とすれば,(VI. 8)より,

$$r = \frac{\Sigma(wV) - NM\overline{w}}{N\sigma_w\sigma_V},$$

卽ち,　$$\Sigma(wV) = Nr\sigma_w\sigma_V + NM\overline{w}, \cdots\cdots\cdots\cdots\cdots \text{(B)}$$

又　$$\Sigma w = N\overline{w}, \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \text{(C)}$$

であるから(A),(B)及び(C)より

$$M' = \frac{\Sigma(wV)}{\Sigma w} = \frac{Nr\sigma_w\sigma_V + NM\overline{w}}{N\overline{w}},$$

卽ち　$$M' = M + \frac{r\sigma_w\sigma_V}{\overline{w}}. \cdots\cdots\cdots\cdots\cdots\cdots\cdots \text{(VII. 4)}$$

卽ち若し重みと變數が正相關にある時は,重みづけられたる平均は算術的平均より大である. 之に反して相關が負であれば,算術的平均より小である.

第三項　相關係數に關するもの

4. 材料の異種である爲に起る相關

XとYとが,第一群のN'個の觀測に於ても,第二群のN''の觀測に於ても無相關なりとす. 第一群に於けるXの平均M_x'と

第二群の平均M_x''とが相等しきか,或はYの平均M_y'とM_y''とが相等しきか,若くは二つとも等しき時は,この二群を混じても新しき相關は生じない.

若し,　$$M_x' \neq M_x'',\quad 且つ\quad M_y' \neq M_y'',$$

なる時は,二群を混ずれば此處に相關を生ずるのである. 今二群を混じたる時のX, Yの平均を$\overline{M}_x$, $\overline{M}_y$とすれば,合同群に於ける積の總和$(N'+N'')\,p$は,

$$(N'+N'')p = N'(M_x' - \overline{M}_x)(M_y' - \overline{M}_y) + N''(M_x'' - \overline{M}_x)(M_y'' - \overline{M}_y) \cdot\cdot$$

である. 何となれば,第一群のみに就て見れば,この群は無相關であるから偏差をx', y'とすれば, $\Sigma(x'y') = 0$である. 故に原點をM_x', M_y'より$\overline{M}_x$, $\overline{M}_y$に移したとすれば,積の和は$N'(M_x' - \overline{M}_x)(M_y' - \overline{M}_y)$となる(184頁(A)).

第二群についても同樣であるから,合同群では上式が出來る. 然るに,この式に於て$M_x' = M_x''$なる時は各は亦$\overline{M}_x$に等し. 故に,

$$p = 0$$

となる.

又, $M_y' = M_y''$なる時も亦同樣である.

然るに,若し$M_x' \neq M_x''$, 且つ$M_y'' \neq M_y''$なる時は, $p=0$とならず. 何となれば, $(M_x' - \overline{M}_x)$と$(M_x'' - \overline{M}_x)$とは異附號を所有し, $(M_y' - \overline{M}_y)$と$(M_y'' - \overline{M}_y)$とは異附號を所有す. 故に, $N'(M_x' - \overline{M}_x)(M_y' - \overline{M}_y)$と$N''(M_x'' - \overline{M}_x)(M_y'' - \overline{M}_y)$とは同附號を有し,又0となることが無いからである.

故に, $M_x' \neq M_x''$, 且$M_y' \neq M_y''$なる時は正若くは負の相關を生ず. 今假りに,身長をXとし,頭長幅指數をYとし,二地方に於

けるX及びYの算術的平均が異つて居れば,假令一地方だけではXとYとの間に相關が無くとも,混合群では一定の相關が生ずるのである．人體を計測して,その體勢につき相關を研究せんとする者はこのことを忘れてはならぬ．

5. 階級分類の相關係數に及ぼす影響

階級幅が小さい程,精密なる數値を得らるることは,度數分布のすべての計算に共通である．しかし,それでは計算煩雜となり且概觀困難となるから,階級幅を大に取ることが多い．この時算術的平均は,影響を蒙ること比較的少く,正しい相稱形分布ならば誤差は極めて小であるが；標準偏差はこれに反し,階級幅を大にするに從つて値が大となる．故に,Sheppard の方法によつて修正し,正しい標準偏差に改めなければならぬ(第三章參照)．然らば階級幅を大にする時,相關係數rは如何？階級幅を大にすれば,rは却て小となるのである．先づ例を以て說明する．

例 1. カシの一種(*Quercus pedunculata*)の或る一本から採取した堅果の長徑と短徑とを計測して,次の相關表とした(Raunkiaer)．

初めは階級幅を0·5 mm. とし(第58表A), 次にその二階級づつを合して階級幅を1 mm. とし(同B), 更に2 mm. とし(同C),終りに階級幅を3 mm. とし(同D)表を作つたのである．

表の(C)と(D)とではあまり縮小し過ぎるから,(B),又は精密な表を欲するならば,(A)位がよからう．今は表を縮少することによつて,M,σ,rに如何なる影響を及ぼすかを見よう．第59表はこの四つの相關表から計算した値をまとめたものである．

第 58 表 A

カシの一種の或る一本より取りたる堅果の長徑及び短徑の二重度數分布表．

$\lambda = 0{\cdot}5$ mm.

長徑＼短徑	7	7·5	8	8·5	9	9·5	10	10·5	11	11·5	12	12·5	13	13·5	計
24														1	1
23·5													1	1	2
23														2	2
22·5											1	1	1		3
22									1	1	3	3	2	1	11
21·5										4	8	2	1		15
21									4	5	4	7	1		21
20·5								1	3	8	10	4			26
20							3	1	15	7	3	1			30
19·5						1	3	11	17	7	1				40
19						5	13	16	11	3					48
18·5					1	3	11	6	3	2					26
18					3	10	30	2	2						57
17·5					6	26	10								32
17			1	4	13	9	5	2							34
16·5			2	8	14	7	1								32
16		1	3	12	8	3									27
15·5			5	5	6	1									17
15			11	2	2										15
14·5			2	4	2										8
14	1	3	2												6
13·5	2														2
13															
12·5			1												1
12	1														1
計	4	4	27	35	55	65	76	39	56	37	30	18	6	5	457

第 58 表 B

Aに於けるのと比べる爲に

λ＝1·0 として作つた相關表.

長徑＼短徑	6·75	7·75	8·75	9·75	10·75	11·75	12·75	13·75	計
23·75							1	2	3
22·75						1	2	2	5
21·75					1	16	8	1	26
20·75					8	27	12		47
19·75				7	44	18	1		70
18·75			1	32	36	5			74
17·75			9	76	4				89
16·75		3	39	22	2				66
15·75		9	31	4					44
14·75		13	10						23
13·75	3	5							8
12·75		1							1
11·75	1								1
計	4	31	90	141	95	67	24	5	457

第 58 表 C

同　上.

λ＝2 mm.

長徑＼短徑	6·25	8·25	10·25	12·25	14·25	計
24·25				1	2	3
22·25			1	27	3	31
20·25			59	58		117
18·25		10	148			163
16·25		82	28			110
14·25	3	28				31
12·25	1	1				2
計	4	121	236	91	5	457

第 58 表 D

同　上.

λ＝3·0 mm.

長徑＼短徑	7·75	10·75	13·75	計
24·75			3	3
21·75		53	25	78
18·75	10	222	1	233
15·75	105	28		133
12·75	10			10
計	125	303	29	457

第 59 表

第58表に示せるカシの堅果の長徑及び短徑について

計算せる σ_x, σ_y, r 等.

λ (階級幅)	$M_y \pm m_y$ (長　徑)	σ_y	$\bar{\sigma}_y$	$M_x \pm m_x$	σ_x	$\bar{\sigma}_x$	$r \pm m_r$
0·5 mm.	18·293±0·096	2·043	2·038	10·122±0·062	1·331	1·323	0·922±0·007
1 mm.	18·266±0·096	2·062	2·041	10·094±0·064	1·368	1·337	0·906±0·008
2 mm.	18·294±0·100	2·113	2·053	10·131±0·069	1·468	1·350	0·835±0·014
3 mm.	18·297±0·104	2·231	2·056	10·120±0·076	1·624	1·374	0·770±0·019

算術的平均は, M_x でも M_y でも, λ の大小には殆ど無關係であるが,標準偏差 σ_x, σ_y は λ の大となるに從つて大となつて居る. しかし, Sheppard の修正を施す時は,完全ではないが可なりよく一致した値 $\bar{\sigma}$ を得ることが出來る(第四列及び第七列).

問題は相關係數であるが,λ が大となるに從つて r は反對に小となることは表の第 8 列に明で, λ＝0·5 mm. なる時 0·922 であつたのが漸次小となり λ＝3 mm. となる時 r＝0·770 となつてゐる. 故に,これらを修正しなければ階級幅の大なる表より

計算せる r は使用することが出來ないことになる.

6. 相關係數の修正

平均 M_x 及び M_y よりの,觀測値 X 及び Y の偏差を x, y とせよ. これ眞の偏差であつて,相關係數はこの x, y より計算すべきである. 然るに,階級分類の爲に一定の誤差が出來る. 即ち偏差は

$$x' = x + \varepsilon_1,$$
$$y' = y + \varepsilon_2$$

となる. ε_1 は階級分類の爲に生ずる x の誤差であり,ε_2 は y の誤差である. 又 x', y' は階級中央の値の偏差である. 故に,

$$\Sigma(x'y') = \Sigma(x+\varepsilon_1)(y+\varepsilon_2) = \Sigma(xy) + \Sigma(\varepsilon_1 y) + \Sigma(\varepsilon_2 x) + \Sigma(\varepsilon_1\varepsilon_2).$$

然るに,x の誤差 ε_1 は y とは無相關,又 ε_2 と x と,且つ ε_1 と ε_2 とも互に無相關であるから,

$$\Sigma(\varepsilon_2 x) = \Sigma(\varepsilon_1 y) = \Sigma(\varepsilon_1\varepsilon_2) = 0.$$

故に, $$\Sigma(x'y') = \Sigma(xy). \quad \cdots\cdots\cdots\cdots\cdots\cdots\text{(A)}$$

$\Sigma(x'y')$ から計算せる相關係數を r とし, $\Sigma(xy)$ から計算せる相關係數は正しき相關係數であるから,これを $\bar{r}$ とし, $\bar{\sigma}_x$, $\bar{\sigma}_y$ を修正せる(即ち眞の)標準偏差とすれば, (A) を用ひて

$$\bar{r} = \frac{\Sigma(xy)}{N\bar{\sigma}_x\bar{\sigma}_y} = \frac{\Sigma(x'y')}{N\bar{\sigma}_x\bar{\sigma}_y} \quad \cdots\cdots\cdots\cdots\text{(VII. 5)}$$

即ち,眞の相關係數 r は分母に修正を施せる $\bar{\sigma}_x$, $\bar{\sigma}_y$ を用ふればよろしいので,他の修正を施す必要がないのである. けれども多くの人は修正を施さざる σ_x, σ_y を用ひて居るから,嚴密な比較にはこれ等の r を修正しなければならぬ. (A) を用ひて

$$r\sigma_x\sigma_y = \bar{r}\bar{\sigma}_x\bar{\sigma}_y,$$

$$\bar{r} = \frac{\sigma_x\sigma_y}{\bar{\sigma}_x\bar{\sigma}_y}r, \quad \cdots\cdots\cdots\cdots\cdots\cdots\text{(VII. 6)}$$

對數表を用ふる時はこの式が便利であるが,近似計算には次の式が便利である[1],

$$\bar{r} = r\left\{1 + 0{\cdot}042\left(\frac{\lambda_1^2}{\sigma_x^2} + \frac{\lambda_2^2}{\sigma_y^2}\right)\right\} \cdots\cdots\cdots\cdots\text{(VII. 7)}$$

これは相關係數の修正式であつて,$\bar{r}$ は修正せられたる相關係數, λ_1, λ_2 は X 及び Y の階級幅である.

(VII. 7) を前節のカシの實の相關表に應用し, $\bar{r}$ を計算すれば,

階級幅	0·5 mm.	の時	$\bar{r} = 0{\cdot}931$,
〃	1 mm.	の時	$\bar{r} = 0{\cdot}935$,
〃	2 mm.	の時	$\bar{r} = 0{\cdot}930$,
〃	3 mm.	の時	$\bar{r} = 0{\cdot}946$,

となり,何れもよく一致する. (C) 及 (D) の階級數餘りに少いに拘らず尙ほ可なり正しい値が出て來た.

第四項 指數及び率

7. 指數及び率

觀測の絕對數のみにては眞相を把握するに困難なることがある. 例へば,1 年間の出生實數よりも,出生の千分比とする方が實際を明瞭に知ることが出來る. 又,人體で下肢や上肢の全長等,その絕對の大さも體勢を知ることに必要であること勿論だが,その身長に對する比例を知ることもゆるがせには出來な

[1] (VII. 7) の證明は省略す.

い．かくの如き場合は,普通全體の百分比又は千分比で表はすことになつて居る．例へば,上肢全長の身長に對する比例は,

$$\text{關係的上肢全長} = \frac{\text{上肢全長}}{\text{身長}} \times 100$$

である．

又,必ずしも一部分の全體に對する比例でなく,例へば頭の長徑,幅徑,高徑等の比を計算することも普通の手段であつて,これも百分比であらはされ,古來一般的に使用せられて居る．これ等は率,比例,指數等と名づけらる．又單に

$$\left.\begin{array}{ll} & \dfrac{\mathrm{B}}{\mathrm{A}} \times 100, \\ \text{又は} & \dfrac{\mathrm{B}}{\mathrm{A}} \times 1000, \end{array}\right\} \cdots\cdots\cdots\cdots\cdots\cdots\cdots\text{(A)}$$

と云ふ様な單純な比でなく,一層複雜な計算を行ふことがある．例へば身體充實指數(K.F.I.)は

$$\text{K.F.I.} = \frac{\text{體重}}{(\text{身長})^3} \times 100 \cdots\cdots\cdots\cdots\cdots\cdots\text{(B)}$$

で表はされて居るが,かく分母子に冪や根があつても,亦,屢指數なる名で呼ばれることがある[1]．

本章では最も簡單なる率,比例,指數,即ち (A) 式の如き單純な比を**指數**なる名で總稱し,その性情を研究することとし,身體充

(1) 指數は,分母の Dimension と分子のそれとを同一にするのが合理的である．身體充實指數の場合では體重は 比重×容積 であるが,比重は單なる數(係數)であつて Dimension は 0, 容積は三次の Dimension であるから,分母においても身長を三乘して三次の Dimension としたのである．これとは反對に,分母は身長として置き,分子を三乘根にするのも一の方法であつて, Livi の Index ponderalis と云ふのは $\dfrac{\sqrt[3]{\text{體重}}}{\text{身長}} \times 1000$ を取つて居るが,同じ用意のあることが注意すべきである．

實指數の如き複雜なる指數は,研究出來ないわけでないが,今は除外して置く．指數は二つの數又は量を取扱ふものであるから,二變數の間に相關が有ると無いとで種々複雜なことが生ずるのである．

8. 指數の算術的平均

指數の算術的平均は,各觀測對につきて計算せる指數の算術的平均であること勿論である．即ち觀測對が,

$$(X_1,\ Y_1),\ (X_2,\ Y_2),\ (X_3,\ Y_3),\ \cdots\cdots,\ (X_N,\ Y_N)$$

である時,指數の算術的平均 I は,

$$I = \frac{100}{N}\left\{\frac{X_1}{Y_1} + \frac{X_2}{Y_2} + \frac{X_3}{Y_3} + \cdots\cdot + \frac{X_N}{Y_N}\right\}$$

である．Y の平均 M_y を以て X の平均 M_x を除し 100 を乘じたるもの,

$$I' = \frac{M_x}{M_y} \times 100$$

を以て指數の平均に代用するのは,以前行はれたることがあるが,嚴密にはよくない．しかし指數の平均 I が與へられて居らず, M_x と M_y と標準偏差 $\sigma_x,\ \sigma_y$ とが與へられて居る時に,近似的に I を計算することが出來る．

定理． 指數平均を I とし, r を X, Y 間の相關係數とし,

$$v_x = \frac{\sigma_x}{M_x}, \quad v_y = \frac{\sigma_y}{M_y} \quad \text{とすれば,}$$

$$I = \frac{M_x}{M_y}(1 - r v_x v_y + v_y^2) \times 100 \cdots\cdots\cdots\text{(VII. 8)}$$

である．

證明． 指數の平均 I は定義により,

$$I=\frac{1}{N}\sum\frac{X}{Y},$$

然るに, $X-M_x=x$, $Y-M_y=y$.

$$\therefore\ NI=\sum\left(\frac{M_x+x}{M_y+y}\right),$$

$$=\frac{M_x}{M_y}\sum\left(\frac{1+\frac{x}{M_x}}{1+\frac{x}{M_y}}\right)=\frac{M_x}{M_y}\sum\left(1+\frac{x}{M_x}\right)\left(1+\frac{y}{M_y}\right)^{-1}$$

$\frac{M_x}{M_y}$ は常數なるが故である. しかるに $\frac{y}{M_y}<1$ なる時は,

$$\left(1+\frac{y}{M_y}\right)^{-1}=1-\frac{y}{M_y}+\left(\frac{y}{M_y}\right)^2-\left(\frac{y}{M_y}\right)^3+\cdots\cdot$$

であるから,

$$NI=\frac{M_x}{M_y}\sum\left\{1-\frac{y}{M_y}+\left(\frac{y}{M_y}\right)^2-\left(\frac{y}{M_y}\right)^3+\cdots\cdot\right.$$

$$\left.+\frac{x}{M_x}-\frac{x}{M_x}\cdot\frac{y}{M_y}+\frac{x}{M_x}\cdot\frac{y^2}{M_y^2}+\cdots\cdots\right\}$$

しかるに, $$\sum\frac{y}{M_y}=\frac{1}{M_y}\sum y=0,$$

$$\sum\frac{x}{M_x}=\frac{1}{M_x}\sum x=0,$$

又, x 及び y が M_x 及び M_y に比し甚しく小なる時は, $\frac{x}{M_x}$ 及び $\frac{y}{M_y}$ に關する三次以上の項を棄てても,近似的には差支なきが故に,

$$NI\fallingdotseq\frac{M_x}{M_y}\sum\left\{1-\frac{xy}{M_xM_y}+\left(\frac{y}{M_y}\right)^2\right\}.$$

而して, $\Sigma xy=Np=Nr\,\sigma_x\sigma_y$, $\Sigma y^2=N\sigma_y^2$

であるから,

$$NI\fallingdotseq\frac{M_x}{M_y}\left(N-\frac{Nr\,\sigma_x\sigma_y}{M_xM_y}+\frac{N\sigma_y^2}{M_y^2}\right),$$

今 $\frac{\sigma_x}{M_x}=v_x$, $\frac{\sigma_y}{M_y}=v_y$, とすれば(v_x, v_y はおのおの變異係數の百分の一である),

$$I\fallingdotseq\frac{M_x}{M_y}(1-r\,v_xv_y+v_y^2)$$

となる. 指數には普通 100 を乘ずるが故に

$$I=100\frac{M_x}{M_y}(1-r\,v_xv_y+v_y^2)\quad\cdots\cdots\cdots(\text{VII. }8)$$

なる式が成立する.

若し,相關係數が小である時,卽ち X と Y とが殆ど無相關なる時は,

$$I\fallingdotseq 100\frac{M_x}{M_y}(1+v_y^2),\ \cdots\cdots\cdots\cdots\cdots\cdots(\text{VII. }9)$$

又, M_y が甚しく大であつて, σ_y 甚しく小なる時は,

$\frac{\sigma_y^2}{M_y^2}=0$ とすることが出來るから,この時は

$$I\fallingdotseq 100\frac{M_x}{M_y}.\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots(\text{VII. }10)$$

さて, I の近似値に(VII. 8)式を用ふるか(VII. 9)又は(VII. 10)を用ふるかは M_y と σ_y 及び r の大さによる.

若し, M_x, M_y が餘りに大ならず $\left(\frac{x}{M_x}\right)$, $\left(\frac{y}{M_y}\right)$ に關して三次以上の冪が大にして無視することが出來ない時は,是等の式は何れも用ひることが出來ないのである. 幸にして人體計測上の多くの指數は, M_x 及び M_y は可なり大であつて,(VII. 8)式又は r の小なる時(VII. 9)式を用ひて充分信ずるに足る數を得るのである.

例 2. 第60表は日本人男女學生の計測の結果である. その最大長徑及び最大幅徑の平均より,指數の平均を計算せよ.

第 60 表

計測項目	♂				♀			
	N	M	σ	v	N	M	σ	v
頭最大長徑	6000	188.55	6.17	3.27	2000	179.31	5.71	3.18
頭最大幅徑	6000	152.18	4.98	3.27	2000	146.73	4.86	3.31
頭長幅指數	6000	80.81	3.67	4.54	2000	81.92	3.70	4.52

頭長幅指數の平均が計算されて居るから,公式(VII. 9)と(VII. 10)とを比較するに便である.

(VII. 10)式　$I = 100\dfrac{M_x}{M_y}$　より計算すれば,

$M_x = 152{\cdot}18(♂)$, $M_y = 188{\cdot}55(♂)$　であるから,男では　$I=80{\cdot}71$,

同様に女では, $I = 81{\cdot}83$　極めて僅かではあるが眞の値より小である. 又,(VII. 9)　$I = 100\dfrac{M_x}{M_y}(1+v_y^2)$　より計算すれば,

男では,　$v_y = 0{\cdot}0327$　なるが故に,　$I = 80{\cdot}79$,

女では,　$v_y = 0{\cdot}0318$　なるが故に,　$I = 81{\cdot}91$,

第60表の指數に極めて近い數字が得られる. 僅かの樣であるが6,000人と云ふ多數の個體の計測では,指數の小數點以下一位の數字は重要であるから,(VII. 9)式を使用した方がいい. 又 r は可なり小で有らうと思ふが,日本人につきて計算せられて居ないから(VII. 8)式の適用は今は控へて置く.

9. 指數の標準偏差

指數の算術的平均と同樣に,指數の標準偏差も,亦 M_x, M_y, σ_x, σ_yから計算することが出來る. 指數の標準偏差を σ_I とすれば,

$$\sigma_I = \sqrt{\frac{1}{N}\sum\left(\frac{X}{Y} - I\right)^2}$$

又,原點を 0 にとりて計算すれば,

$$\sigma_I{}^2 = \frac{1}{N}\sum\left(\frac{X}{Y}\right)^2 - I^2$$

である. 故に,

$$\sigma_I{}^2 + I^2 = \frac{1}{N}\sum\left(\frac{X}{Y}\right)^2 = \frac{1}{N}\frac{M_x{}^2}{M_y{}^2}\sum\left(1+\frac{x}{M_x}\right)^2\left(1+\frac{y}{M_y}\right)^{-2},$$

しかるに, $y < M_y$ なる時,

$$\left(1+\frac{y}{M_y}\right)^{-2} = 1 - 2\frac{y}{M_y} + 3\left(\frac{y}{M_y}\right)^2 - 4\left(\frac{y}{M_y}\right)^3 + \cdots\cdots$$

であるから,

$$N(\sigma_I{}^2 + I^2) = \frac{M_x{}^2}{M_y{}^2}\sum\left(1+2\frac{x}{M_x}+\frac{x^2}{M_x{}^2}\right)\left(1-2\frac{y}{M_y}+3\left(\frac{y}{M_y}\right)^2-\cdots\right)$$

$$= \frac{M_x{}^2}{M_y{}^2}\sum\left\{1+2\frac{x}{M_x}-2\frac{y}{M_y}+\left(\frac{x}{M_x}\right)^2-4\frac{x\,y}{M_x M_y}\right.$$

$$\left. + 3\left(\frac{y}{M_y}\right)^2 + \varepsilon^3\right\},$$

ε^3 は $\left(\dfrac{x}{M_x}\right)$ 及び $\left(\dfrac{y}{M_y}\right)$ に關して三次以上の項のみを含む級數の和である. 而して, $\dfrac{x}{M_x}$ 及び $\dfrac{y}{M_y}$ が甚しく小なる時は, ε^3 はこれを無視してもいい.

又, $\sum\dfrac{x}{M_x} = \sum\dfrac{y}{M_y} = 0$ であるから,

$$\sigma_I{}^2 + I^2 = \frac{M_x{}^2}{M_y{}^2}\left(1+\frac{\sigma_x{}^2}{M_x{}^2} - 4\frac{r\sigma_x\sigma_y}{M_x M_y} + 3\frac{\sigma_y{}^2}{M_y{}^2}\right),$$

$\dfrac{\sigma_y}{M_y} = v_y$, $\dfrac{\sigma_x}{M_x} = v_x$　とすれば,

$$\sigma_I{}^2 + I^2 = \frac{M_x{}^2}{M_y{}^2}(1 - 4r\,v_x v_y + v_x{}^2 + 3\,v_y{}^2),$$

この式の I^2 に公式(VII. 8)を代入すれば

$$\sigma_I{}^2 = \frac{M_x{}^2}{M_y{}^2}(1 - 4r\,v_x v_y + v_x{}^2 + 3\,v_y{}^2) - \frac{M_x{}^2}{M_y{}^2}(1 - r\,v_x v_y + v_y{}^2)^2,$$

$$\sigma_I = \frac{M_x}{M_y}\sqrt{1-4r\,v_x v_y + v_x^2 + 3v_y^2 - (1-2r\,v_x v_y + 2\,v_y^2 + \varepsilon'^3)},$$

故に $$\sigma_I \fallingdotseq \frac{M_x}{M_y}\sqrt{v_x^2 - 2r\,v_x v_y + v_y^2},$$

若し, X と Y とが無相關なる時は,

$$\sigma_I = \frac{M_x}{M_y}\sqrt{v_x^2 + v_y^2},$$

但し,指數には普通 100 を乘ずるが故に,

$$\sigma_I \fallingdotseq 100\frac{M_x}{M_y}\sqrt{v_x^2 - 2r\,v_x v_y + v_y^2}, \quad \cdots\cdots\cdots\text{(VII. 11)}$$

又は $$\sigma_I = 100\frac{M_x}{M_y}\sqrt{v_x^2 + v_y}\,. \cdots\cdots\cdots\cdots\cdots\cdots\text{(VII. 12)}$$

例 3. 例2,第60表に於ける頭長徑と頭幅徑の M_x, M_y, σ_x, σ_y を用ひ, σ_I を計算せよ.

男子に於ては,	女子に於ては,
$M_x = 152{\cdot}18$,	$M_x = 146{\cdot}73$,
$M_y = 188{\cdot}55$,	$M_y = 179{\cdot}31$,
$v_x = 0{\cdot}0327$,	$v_x = 0{\cdot}0331$,
$v_y = 0{\cdot}0327$,	$v_y = 0{\cdot}0318$,

(VII. 12)式を用ひて,

$$♂\quad \sigma_I = 100\frac{152{\cdot}18}{188{\cdot}55}\sqrt{2\times 0{\cdot}0327^2} = 100\cdot\frac{152{\cdot}18}{188{\cdot}55}\times 0{\cdot}04626$$
$$= 3{\cdot}73,$$

$$♀\quad \sigma_I = 100\cdot\frac{146{\cdot}73}{179{\cdot}31}\sqrt{0{\cdot}0331^2 + 0{\cdot}0318^2} = 100\frac{146{\cdot}73}{179{\cdot}31}\times 0{\cdot}04590$$
$$= 3{\cdot}75,$$

第60表の値に比し少しく大に過ぎるが,用ふるに足りる. (VII. 11)式を用ふれば一層精確であらう.

10. 擬相關

X, Y, Z は互に相關がないとし,これより二つの指數 $\frac{X}{Z}\times 100$, $\frac{Y}{Z}\times 100$ を作るに,この二つの指數の間には一定の相關が出來る. これ Z なる共通の分母を有するが爲である.

指數の平均は,

$$I_1 = \frac{100}{N}\sum\frac{X}{Z}, \qquad I_2 = \frac{100}{N}\sum\frac{Y}{Z},$$

但し,今は係數 100 を除いて研究する.

又指數 $\frac{X}{Z}$ の標準偏差 σ_1, $\frac{Y}{Z}$ の標準偏差 σ_2 とし, $\frac{X}{Z}$ と $\frac{Y}{Z}$ との間の相關係數を ρ とすれば,

$$\rho = \frac{1}{N}\cdot\frac{\sum\left(\frac{X}{Z}-I_1\right)\left(\frac{Y}{Z}-I_2\right)}{\sigma_1\sigma_2},$$

$$\therefore\quad N\sigma_1\sigma_2\rho = \sum\left(\frac{XY}{Z^2}\right) - I_1\sum\frac{Y}{Z} - I_2\sum\frac{X}{Z} + NI_1I_2,$$

しかるに, $$I_1\sum\frac{Y}{Z} = NI_1I_2,$$

$$I_2\sum\frac{X}{Z} = NI_2I_1,$$

$$\therefore\quad N\sigma_1\sigma_2\rho = \sum\frac{XY}{Z^2} - NI_1I_2.$$

卽ち, $$N\sigma_1\sigma_2\rho + NI_1I_2 = \frac{M_xM_y}{M_z^2}\sum\left(1+\frac{y}{M_y}\right)\left(1+\frac{x}{M_x}\right)\left(1+\frac{z}{M_z}\right)^{-2},$$

最後の因數を無限級數に展開して

$$N\sigma_1\sigma_2\rho + NI_1I_2 = \frac{M_xM_y}{M_z^2}\sum\left(1+\frac{y}{M_y}\right)\left(1+\frac{x}{M_x}\right)\left(1-2\frac{z}{M_z}+3\left(\frac{z}{M_z}\right)^2-\cdots\right)$$
$$= \frac{M_xM_y}{M_z^2}\sum\left(1+\frac{y}{M_y}+\frac{x}{M_x}-2\frac{z}{M_z}+\frac{yx}{M_yM_x}-2\frac{y}{M_y}\cdot\frac{z}{M_z}\right.$$
$$\left.-2\frac{x}{M_x}\cdot\frac{z}{M_z}+3\left(\frac{z}{M_z}\right)^2+\varepsilon^3\right),$$

$\mathcal{E}^3$ は $\frac{x}{M_x}, \frac{y}{M_y}, \frac{z}{M_z}$ に關して三次以上の項よりなり, $\frac{x}{M_x}, \frac{y}{M_y}, \frac{z}{M_z}$ が甚小である時は, $\mathcal{E}^3$ は無視することが出來る. 又 $\sum\frac{y}{M_y} = \sum\frac{x}{M_x} = \sum\frac{z}{M_z} = 0$, 且つ假定により X, Y, Z が無相關であるから, $\Sigma yz = \Sigma xy = \Sigma xz = 0$.

故に, $$N\sigma_1\sigma_2\rho + NI_1I_2 = \frac{M_xM_y}{M_z^2}\left(N + 3N\frac{\sigma_z^2}{M_z^2}\right),$$

$$\sigma_1\sigma_2\rho = \frac{M_xM_y}{M_z^2}(1+3v_z^2) - I_1I_2, \cdots\cdots\cdots\cdots(A)$$

而して, $I_1 = \frac{M_x}{M_z}(1+v_z^2)$, $I_2 = \frac{M_y}{M_z}(1+v_z^2)$ であるから,

$$I_1I_2 = \frac{M_xM_y}{M_z^2}(1+2v_z^2+v_z^4).$$

之を(A)式に代入すれば, v_z^4 は 0 に等しと考へられるから,

$$\sigma_1\sigma_2\rho = \frac{M_xM_y}{M_z^2}(v_z^2),$$

又 $\sigma_1 = \frac{M_x}{M_z}\sqrt{v_y^2+v_z^2}$, $\sigma_2 = \frac{M_y}{M_z}\sqrt{v_y^2+v_z^2}$.

であるから(公式 VII. 12),

$$\rho = \frac{v_z^2}{\sqrt{(v_x^2+v_z^2)(v_y^2+v_z^2)}}, \cdots\cdots\cdots\cdots(\text{VII. 13})$$

これ指數 $\frac{X}{Z}$ と $\frac{Y}{Z}$ との間の相關係數である.

v_x, v_y が v_z に比し甚しく小なる時は,

$$\rho \fallingdotseq 1.$$

となる. 又 v_x, v_y が v_z に比して甚だ大なる時は,

$$\rho \fallingdotseq 0$$

となる. 普通はその中間にあり,且つ分母が共通なる時は常に正の相關としてあらはれるものである.

定理. 分母を同じくする二つの指數間には,たとひその指數を構成する三變數互に無相關であつても,正の相關が成立する. その相關係數 ρ は,

$$\rho = \frac{v_z^2}{\sqrt{(v_x^2+v_z^2)(v_y^2+v_z^2)}} \cdots\cdots\cdots\cdots(\text{VII. 13})$$

である.

かく變數に一定の演算を施したるが爲に新變數間に生じたる相關を**擬相關** (Spurious correlation) と云ふ.

今,假りに X, Y, Z 何れも無相關とし, $v_x = v_y = v_z$ とすれば, $\rho = 0{\cdot}5$ となり可なり,高度の相關が出て來るのである. 人類學的方面にはこの擬相關が可なり多い.

例 4. Y と指數 $\frac{X}{Y} \times 100$ との間の擬相關を研究し,實例と比較せよ. 但し X と Y とは相關なきものと假定す.

解. 指數 $\frac{X}{Y}$ と Y との間の相關係數を ρ とすれば,

$$N\sigma_y\sigma_I\rho = \sum\left(Y\cdot\frac{X}{Y}\right) - NM_yI,$$

$$= \Sigma X - NM_yI,$$

$$\sigma_y\sigma_I\rho = M_x - M_yI, \cdots\cdots\cdots\cdots\cdots\cdots(A)$$

X と Y とは相關がないから(VII. 9)より,

$$I = \frac{M_x}{M_y}(1+v_y^2). \cdots\cdots\cdots\cdots\cdots\cdots(B)$$

又(VII. 12)により, $\sigma_I = \frac{M_x}{M_y}\sqrt{v_x^2+v_y^2}$. $\cdots\cdots\cdots\cdots\cdots\cdots$(C)

これを(A)式に代入して,

$$\rho = \frac{M_x - M_y\frac{M_x}{M_y}(1+v_y^2)}{\sigma_y\frac{M_x}{M_y}\sqrt{v_x^2+v_y^2}},$$

$$\rho = \frac{-v_y{}^2}{\dfrac{\sigma_y}{M_y}\sqrt{v_x{}^2+v_y{}^2}}.$$

即ち $$= \frac{-v_y}{\sqrt{v_x{}^2+v_y{}^2}}, \quad \cdots\cdots\cdots\cdots\cdots\cdots\text{(VII. 14)}$$

この式を實際の計測値に當篏めて見る.

William Rowan 等はアジサシ (Common tern) の卵殼の長徑と幅徑を計測した. 長徑を Y, 幅徑を X, その長幅指數を $100\frac{X}{Y}$ とし,第 61 表の如き數値を得た.

第 61 表
アジサシの卵殼の長徑と幅徑 (cm.)

	M	σ	v (100 を乘せず)
長 徑 (Y)	4·14	0·180	·0434
幅 徑 (X)	2·98	0·099	·0333
$100\frac{X}{Y}$	72·04	3·449	·0479

この數を上式に當篏めて見る.

$$v_x = 0{\cdot}0333, \quad v_y = 0{\cdot}0434$$

であるから,長徑 Y と指數 $100\frac{X}{Y}$ との間の相關係數, ρ は (14) 式より

$$\rho = \frac{-0{\cdot}0434}{\sqrt{0{\cdot}0434^2+0{\cdot}0333^2}} = \frac{-0{\cdot}0434}{0{\cdot}05470} = -0{\cdot}79$$

となるべきである. 然るに, Rowan は計測値から計算して, Y と $100\frac{X}{Y}$ との間の相關係數 ρ' は $-0{\cdot}728$ としてゐる. 上式(14)による値 ρ は可なり信ずるに足る. しかし嚴密に一致したとは云へないのは, X と Y との間に相關が無いとした假定に多少無理がある爲である. (實際 $r_{xy} = +0{\cdot}222$ であつた.)

11. 指數分類の價値[1]

人類學方面では指數を分類することが行はれて居る. 指數が人類學に應用せられたのは, 1842 年 Anders Retzius に始まる. 彼は頭骨の短徑を長徑で除したものを頭骨指數とし,これを以て頭骨を長頭と短頭に分つたのであるが,至極便利であり,又形態學上の事實に立脚せる科學的の分類であつたが爲に,人類學者,解剖學者等によりあまねく採用せられるに至つた.

今日では頭骨その他の骨の計測に於ても,或は生體計測に於ても,指數を計算しないものは寧ろ稀で; 中に Mollison の如きは,人種的差異を立て得るのは指數と角度とに據るべきで,長さとか容積とかは人種的特徵を指示する尺度とはならぬとまで極言して居る. これは極端で,指數や角度は人種的差異のあること勿論であるが,絕對の大さも亦人種的に重要なる尺度である. 指數の內,特に重要視せられて居るのは,頭骨又は生體頭部の長幅指數であるが,是等,重要な指數には分類法が行はれて居る. Retzius は頭骨を指數で長頭と短頭に分類したのであるが,多數の頭骨を研究するにはこれでは不充分だから,中頭なる階級が設けられ (1861, Broca.), 次で,種々なる分類法が多くの人により提唱せられた. 或は四階級に,或は五階級,七階級,九階級等に分類する樣になつた. 今日頭骨の分類に最も普通に行はれて居るのは,七階級に分類することである.

過 長 頭 (ultradolichokran) × —64·9,
超 長 頭 (hyperdolichokran) 65·0—69·9,

(1) このことは相關とは關係が無いけれども指數の序に述べて置く.

長 頭 (dolichokran) 70·0—74·9,

中 頭 (mesokran) 75·0—79·9,

短 頭 (brachykran) 80·0—84·9,

超 短 頭 (hyperbrachykran) 85·0—89·9,

過 短 頭 (ultrabrachykran) 90·0—×

かく多くの階級に分類したのでは煩瑣に過ぎ,又階級の少い分類では用をなさない. Pearsonは51人種の平均頭指數を度數分布表に作り,その中央値(Med.)が78であつたから,これを短頭と中頭との境界とし,第一四分の一限界, 75·38を長頭の境界とし,第三四分の一限界80·54を超短頭の境界とせよと云ふ. 四分の一限界や,中央値を以て區切るのは一進步であるが,それでも絕對的の區分で無く,且つ75·38, 80·54の如き數は簡單に記憶出來ない不便がある.

又,生體に於ては頭長幅指數は頭骨のそれに比して約1單位大であるとし,これを考慮に入れて次の如く分類する人もある.

生體に於ける頭指數分類法

長 頭 (dolichocephal) ×—75·9,

中 頭 (mesocephal) 76·0—80·9,

短 頭 (brachycephal) 81·0—85·4,

超 短 頭 (hyperbrachycephal) 85·5—×,

頭(及び頭骨)の長幅指數の分類法は他に幾らもあり,且又他の指數の分類法を加へると,その名稱と區分の仕方とを記憶するだけでも相當な勞力である. 反對に二階級乃至四階級に分類

したのでは,餘りに簡單で分布を充分に表はさないと云ふ憾がある. 元來,指數分類は,今日の如く度數分布の數學的硏究が充分で無かつた時代に,指數の分類を詳しく表示する目的で廣く行はれたのであるから,分布の硏究の行き屆いた今日,これに代はる散布度の明になつて居る今日,これらの煩雜な分類法を用ひずとも宜いのである.

指數の分類は主として二つの目的に用ひられて居る.

(1) 度數分布を簡單に表はす爲に用ひらる.

指數の分類を度數分布を表示する爲に用ひた一例をあげると,第62表がそれである.

第 62 表

日本人(金澤地方人)の頭骨長幅指數の分類とその理論的度數.

階級	實測の度數	實測百分比度數	理論的計算の百分比度數
過長頭	—	—	
超長頭	2	1·3 %	1·6 %
長頭	34	21·1 %	21·1 %
中頭	83	52·9 %	51·3 %
短頭	33	21·0 %	23·9 %
超短頭	5	3·2 %	2·1 %
過短頭	—	—	
計	157	100·1	100·0

$N=157,\quad M=77·69,\quad \sigma=3·594.$

しかし,この表示法は唯度數分布表の階級を大に取つたに過ぎない. これでは,度數分布を表はすには餘りに簡單すぎ,又分布の代表値としては複雜すぎる. 今日の生物統計學では,代表値を表はすに一層簡單で,又正確な表示法がある. 卽ち,正常分布の場合(人體計測の場合は大多數は正常分布である.), 計測總

数Nと,算術的平均Mと,標準偏差σとであつて,この三つが與へられて居れば,度數分布のすべてが與へられて居るのと同様である. すべての散布度,すべての精確度がこれから計算せられるのみならず,度數分布表までも,近似的ではあるが理論的に計算しうるのである. 第62表の理論數(第4列)は,M及び$\bar{\sigma}$から計算したものであるが,其のよく第三列に一致せるを見よ. 即ちN, M, σの三つが第一義的に重要なもので,第62表三列又は四列の分布の如きは第二義的のもので,熟練せる者にはN, M, σを知つただけで,第62表の分布は略推知することが出來る位である.

かくの如き理由で, Poniatowski は指數分類は意義が無いと云ふ. 肯綮に當つた言である.

(2) 第二には指數の平均を簡單にあらはすに,長頭とか短頭とかの名稱を用ふる. 例へば朝鮮人は短頭であると云ひ(頭指數平均約 83·5 頭骨指數約 81·5), 又日本人は概略中頭と短頭の間にあると(頭指數約 81 頭骨指數約 79·5)云ふ様なものである. 正確に算術的平均を擧げる代りに,指數分類の名を用ふれば,極めて印象深く響くから使用するのみ. 又二つの人種を比較するに,朝鮮人は日本人よりも短頭であるとか,アイヌ人は日本人よりは,より長頭であるとか,形容詞的に用ふることもあるが,その意味は,平均して朝鮮人の頭指數は日本人の平均より大であると云ふこと,又アイヌ人頭指數の平均は日本人のそれよりも小であることを表はすものである. かく用ふるならば今日でも,その名を保存しても宜し. しかし簡單にする爲に重要なる指數に限るべきである (Poniatowski).

演習問題

1. 指數, $100\dfrac{X \pm C}{Z}$ の算術的平均と,標準偏差とを M_x, M_z, v_x, v_z 及び r より計算する式を誘導せよ. 但し C は常數とす.

答[1] $$I = 100\frac{M_x}{M_z}(1 - r v_x v_z + v_z^2) \pm 100\frac{C}{M_z}(1 + v_z^2),$$

$$\sigma = \frac{100}{M_z}\sqrt{M_x^2(v_x^2 - 2r v_x v_z + v_z^2) \pm 2CM_x(v_x^2 - r v_z v_x) + C^2 v_x^2}.$$

2. 指數 $\dfrac{X}{Z}$ と $\dfrac{X}{Y}$ との間に如何なる擬相關を生ずるかを,研究せよ. 但し X, Y, Z 間に元來相關の無きものとす.

3. 朝鮮人の頭骨を計測して,第63表甲を得た. この値を利用して,種々の指數平均と標準偏差とを間接に計算し,乙表の平均及び標準偏差と比較せよ.

第 63 表 甲

		♂			♀		
		M	σ	v	M	σ	v
(1)	頭最大長徑	176·7	6·75	3·82	167·8	6·64	3·95
(2)	頭最大幅徑	142·6	6·23	4·37	136·9	6·03	4·40
(3)	バシオン-ブレグマ高徑	138·4	4·90	3·54	132·0	4·80	3·64
(4)	前頭最小幅徑	91·4	4·90	5·36	87·7	4·00	4·56
(5)	上顏面高徑	74·0	4·69	6·34	68·3	4·76	9·97
(6)	顴骨弓幅徑	134·7	5·70	3·23	126·6	5·24	4·14

第 63 表 乙 (直接計算)

		♂		♀	
		M_I	σ_I	M_I	σ_I
頭長幅指數	$\frac{(2)}{(1)}\times 100$	80·7	5·15	82·0	5·32
頭長高指數	$\frac{(3)}{(1)}\times 100$	78·5	4·13	79·2	3·88

[1] 文獻(20)に證明あり.

幅高指數	$\frac{(3)}{(2)}\times 100$	97·3	4·79	96·4	5·46
上顎面指數	$\frac{(5)}{(6)}\times 100$	55·2	3·52	54·4	3·47
Jugofrontal-Index	$\frac{(4)}{(6)}\times 100$	67·8	3·46	69·3	3·11
Transversaler Craniofacial-I.	$\frac{(6)}{(2)}\times 100$	94·8	4·34	92·2	5·00

4. X と $\frac{X}{Y}$ との間の擬相關を求めよ. 但し X, Y 間には相關なきものとす.

参考文獻 第五章—第七章

Yule, Johannsen の外次のものあり:

1. Galton, Francis :- Regression towards mediocrity in hereditary stature. *Journ. Anthr. Inst., Vol. 15, 1886.*

2. Galton, Francis :- Family likeness in stature. *Proc. Roy. Soc., Vol. 40, 1886.*

3. Galton, Francis :- Correlation and their Measurement. *Proc. Roy. Soc., Vol. 45, 1888.*

4. Pearson, Karl :- Grammar of Science.

5. Pearson, Karl :- Regression, Heredity, and Panmixia. *Phil. Trans. Roy. Soc., Series A, Vol. 187, 1896.*

6. Yule, G. U. :- On the significance of Bravais' formulae for regression, etc. in the case of skew correlation. *Proc. Roy. Soc., Vol. 60, 1897.*

7. Yule, G. U. :- On the theory of correlation. *Journ. Roy. Stat. Soc., Vol. 60, 1897.*

8. Ezekiel, M.:- Methods of correlation Analysis, New York, 1930.

計算簡便法に關するもの:-

9. Harris, J. Arthur :- A short method of calculating the coefficient of correlation in the case of integral variates. *Biometr., Vol. 7, 1909.*

10. Harris, Arthur :- On the calculation of intra-class and interclass coefficients of correlation from class-moments when the number of possible combi-

nations is large. *Biometr., Vol. 9, 1914.*

回歸一次ならざる相關に關するもの:-

11. Pearson, Karl :- On the systematic fitting of curves to observations and measurements. *Biometrica, Vol. I/II, 1902.*

12. Pearson, Karl :- On the general theory of skew correlation and non-linear regression. *Drapers' Co. Research Memoirs : Biometric Series, II, London, 1905.*

13. Pearson, Karl :- On a Correlation to be made to the correlation ratio. *Biometr., Vol. 8, 1911.*

14. Blakeman, J. :- On tests for linearity of regression in frequency-distributions. *Biomtr., Vol. 4, 1905.*

15. Slutsky, E. :- On the criterion of goodness of fit of regression lines and the best method of fitting them to the data. *Journ. Roy. Stat. Soc., Vol. 77, 1913.*

相關係數に及ぼす觀測誤差の影響について:-

16. Spearman, C. :- Demonstration of formulae for true measurement of correlation. *Amer. Journ. of Psych., Vol. 18, 1907.*

Index に關するもの:-

17. Pearson. Karl :- On a form of spurious correlation which may arise when indices are used in the measurement of organs. *Proc. Roy. Soc., Vol, 60, 1897.*

18. Galton, Francis :- Note to the memoir by Prof. Pearson on spurious correlation. *ibid.*

19. Yule, G. U. :- On the interpretation of correlations between indices or ratios.

20. 荒瀨進外三名:- 朝鮮人の體質人類學的研究,朝鮮醫學會雜誌 昭和九年一月

21. Brown, J. W., etc. :- A study of Index-Correlations. *Journ. Roy. Stat. Soc., Vol. 77, 1914.*

階級分類の影響に關するもの:-

22. Sheppard, W. F. :- On the calculation of the average square, cube etc., of a large number of magnitudes. *Journ. Roy. Stat. Soc., Vol. 60, 1897.*

23. Sheppard, W. F. :- The calculation of moments of a frequency-distribution. *Biometr., Vol, 5, 1907.*

24. Pearson, Karl & others :- On an elementary proof of Sheppard's formulae for correcting Raw Moments, and on other allied points. *Biometr., Vol. 3, 1904.*

25. 上田常吉:- 相關係數の修正法に就て,解剖學雜誌第7巻昭和九年五月.

26. Poniatowski, S. :- Über den Wert der Indexklassifikationen, *Arch. f. Anthr. N. F. X., 1911.*

重みづけられたる平均:-

27. Pearson, K. :- Note on Reproductive Selection. *Proc. Roy. Soc., Vol. 59, 1896.*

第三編

確率論及びその應用

第八章

確率の定義

1. 確率論

今一つの貨幣を投ずるに,其の表が出るか裏が出るかは全く偶然の事に屬し,一回投げただけでは,表が出るとも裏が出るとも斷言することは出來ない. 併し,これを度々繰返して行ふときは,表の出る回數と裏の出る回數との間に,一定の割合のあることがわかる. 卽ち,貨幣が完全な圓板ならば,表と裏とは略同じ回數だけ出現する. 又雙六の賽を取つて振るに,六つの面の內何れが出るかは全く偶然であつて,昔から山法師·鴨川の水と共に意のままにならない者の代表となつて居るが,回を重ねて振つて居れば, 1 の面の出る回數と 2, 3, 4, 5, 6 の面の出る回數との間に,一定の割合があることが明になる. 若し,この賽が完全に正六面體で,且つこの賽を作る物質が完全に等質ならば,振つた回數の約六分の一だけの目が出るのである. かく,偶然の現象にも一定の法則があるのである.

偶然の事象の法則を研究する數學の一分科を**確率論**と云ふ. 又,**公算論**,**確らしさの理論** (Theory of probability, Die Wahrscheinlichkeitsrechnung)とも云ふ. 貨幣や賽を投じたる場合の如く,偶然の現象に歸すべき事はこの外にまだ甚だ多い. 袋の中の球の何れを抽き當てるかと云ふ問題や,引いた籤の當るか否か,又は分配したトランプの札の中で或特定の札が特定の人に來るか否かと云ふ様な問題は,皆偶然の現象に歸し,確率論の題目になる. 學術上に取扱ふ事項にも偶然の影響を蒙るものが多く,ことに生物統計學には今日では確率論の知識が缺くべからざるものとなつた.

本編においては,生物統計學への應用を主眼として確率論を初めからわかり易く解說することにし,特に確率論と統計學との聯絡に力め,統計學に比較的緣の遠い幾何學的確率の如きは省略し,又,原因確率や未來事象の確率の如きは,思ふ所ありて續編に讓ることとした.

2. 偶然

確率論は偶然の事象の數學的研究を目的とすると云つたが,偶然とは何かとは當然起る問題である. 銅貨を投じて假りに表が出たとすれば,これは偶然だと云ふが,偶然とは原因がないと云ふことではない. 如何なる事象でも,原因がなく起るものとは考へられない. 然らば,銅貨の表が出た時,表を出した原因を檢べるに,投げる高さや,投ずる時の銅貨の持ち方や,加へる力の强さ及び方向,又投じたる底面の滑さ,硬さ,彈性,傾斜,銅貨の大さ,厚さ,硬さ,彈性等,種々の原因が錯綜して表が出されたのである. しかも,諸種の原因の綜合たるや,極めて複雜微妙であつて,表が出るとも裏が出るとも人智で決定することが出來ない.

かく,結果を決定する原因が極めて複雜微妙で,或る結果を必然的に起す如く原因を定められない様な事象を偶然の事象と云ふのである. もし投ずる時の運動を簡單とし,必ず表又は必ず裏を出さしむることをうる様にすれば,もはや**偶然**でなくして**必然**となるのであつて,確率論の問題を離れる.

而して,偶然の事象の法則は多數の觀測によつて初めて明にすることが出來る.

3. 確率の經驗的定義(統計的定義)

N 個の實驗又は觀測を行へる時に,一事象が P 回生起したり. N を無限に多くせる時, $\dfrac{P}{N}$ なる比の極限値を其の事象生起の確率と云ひ, p を以て表はす. 即ち

$$p = \lim_{N\to\infty}\frac{P}{N} \quad\cdots\cdots\cdots\cdots\cdots (\mathrm{VIII.\ 1})$$

である.

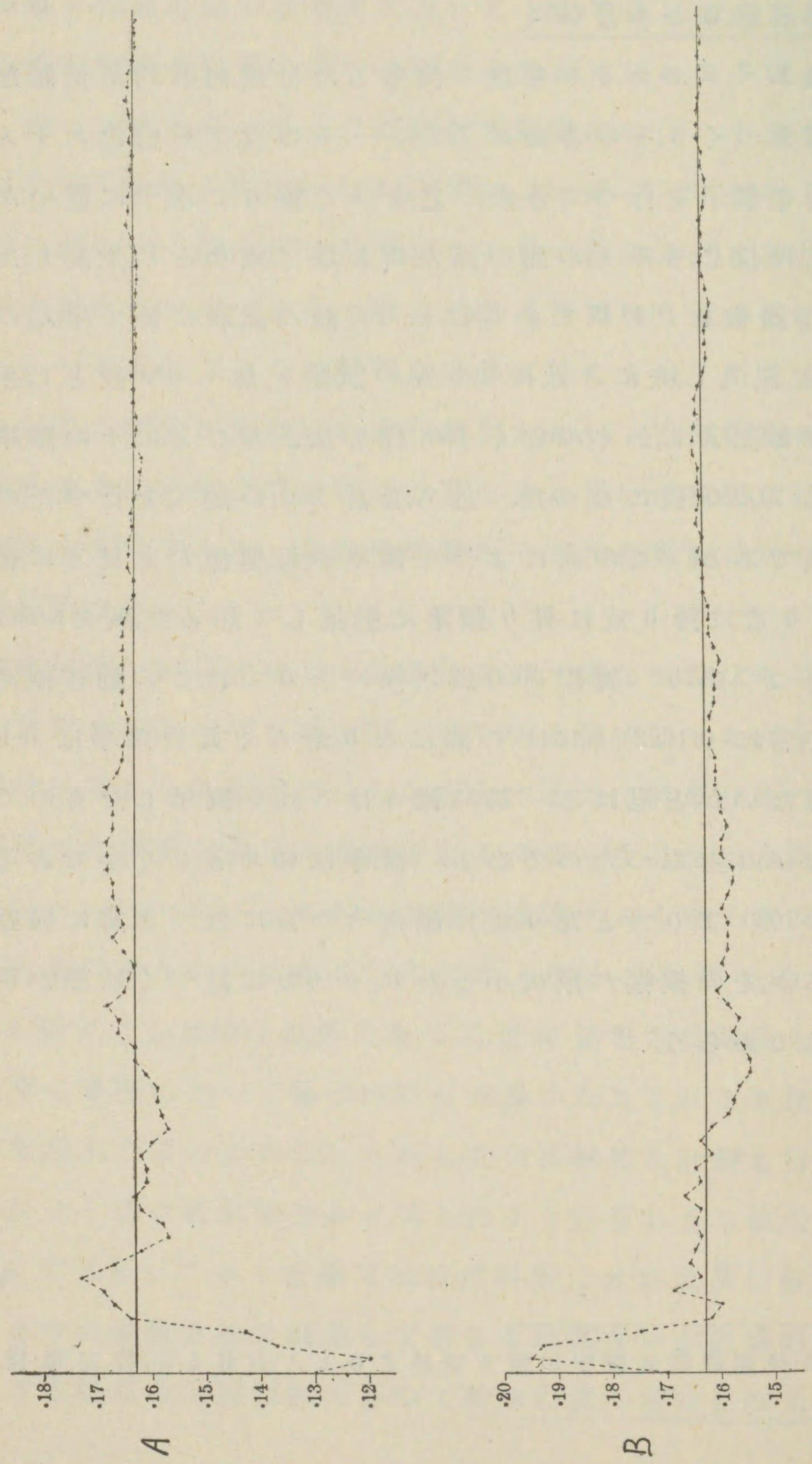
この定義に從へば,確率は實驗または觀察の結果を統計して初めて決定せらるべきものである. 故に,斯く定義せる p を**經驗的確率**または**統計的確率** (a posteriori probability or statistical probability) と云ふ. しかるに,實驗又は觀察回數 N を無限に大ならしむることは,實際に出來ないことであるから,この定義による確率の眞の値は計算出來ないのである. しかし,初めより N を漸次大とすれば $\dfrac{P}{N}$ は或は大となり或は小となり,種々動搖するが,漸次確率 p に接近し,終に非常に大なる N に於ては N 無限大ならずとも, $\dfrac{P}{N}$ を以て實際 p と考へて差支なきものとなる. 即ち,**確率 p の近似値**が得られ,これを以て滿足することが出來るのである.

確率 p の眞の値の計算出來ないことを不滿足に感ずる人もあらうが,數學や自然科學には斯の如き例は極めて多いのである.例へば圓周率 π は一の超越數で無限につづく小數である,これを 707 桁まで算出した人 (William Shanks. 1783) が有るが,其れでも眞の π の値が出たのではなく,眞の値は永久に算出することが出來ないのである.然るに,吾々は 3·1416 なる近似値で滿足して居るのであつて,精密な計算に於ても尙 7 桁位までで,其れ以上の精度は實際はあまり必要がない.自然對數の底 e の如きもまた同樣で,眞の價は計算しきれないが,實際は近似値 2·71828 で充分である.其の他對數表や,三角函數表の數は大部分近似値である.

自然科學の領域に於ても,近似値を以て滿足せねばならぬ場合が甚だ多く,例へば水素の原子量 1·008 と稱するも,これ今日の化學で計算しうる範圍內での近似値であつて,將來化學の發達と共に更に精密な數字で表はされる必要を生じ,又は左樣なるかもしれない.しかし如何に化學が發達するも原子量の眞の値は計算できないのである.又日本人男子の身長平均 161·0 cm. と稱するも勿論近似値であつて,更に精密な計測法と精密な計算を多數について施せば,これと異つた又これより精密な數字を出しうるのである.しかし如何に精密な計測を行ふも,實測による値は近似値であり,又これより計算しうる値も亦近似値にすぎない.かく數學又は自然科學に於ては,眞の値は計算できず近似値を以て滿足して居るもの甚多いが,經驗的に定義しうる確率も亦近似値であつて,確率の眞の値は存在するのであるが計算できない[1].

實驗 I. エボナイトを以て出來るだけ幾何學的に正確な正六面體をつくり,その各面に 1,2,3,‥‥6 の文字を白色エナメルにて書き骰子を作つてみた.之をよく振りて水平に置いた硝子板上に投じ,その 1 の面の出た時を以て成功とし(生起したとし),その回數を P を以てあらはし,100 回の試驗に於て生起の確率 $\frac{P}{N}$ を記入し,次にさらに 100 投の試驗を加へ 200 投とし生起の確率を記入し,かくの如く 100 投を加ふるごとにその確率を記入し,10000 投に至つた.即ち,合計 100 の記入を行つたのが第64表である.この表によつて確率の近似値 P を見るに初め 0·13 より或は降り或は昇り,頻繁に動搖して居るが,漸次 0·163 に近づきつつあり.實際 6000 投以後に至りてはこの動搖極めて少なく,常に 0·1622—0·1641 の間にあり,恐らく眞の確率は 0·1630 位ではないかと思はる.第51圖 A はこれを圖示したもので,橫線に 0·1630 を取つたのである.確率は初め著しく小であるが,次で 0·1630 より大となり,更に漸次その値に近づき,常に振動して居るが,その振幅が漸次小となり,0·1630 に近づく狀態が明瞭になるであらう.

[1] 今後屢經驗的確率の値を計算することが有るが,皆近似値 $\frac{P}{N}$ である.



第51圖　實驗 I 及び II にて其の 1 及び 6 の目の出る確率が漸次一定の値に收斂するを示す.

第 64 表

1 の目の出た確率

·13 → ·12 → ·137 → ·143 → ·164 → ·167 → ·169 → ·173 → ·167 → ·162 → ·152 → ·158 →
·162 → ·163 → ·161 → ·161 → ·162 → ·160 → ·157 → ·158 → ·159 → ·159 → ·161 → ·163 →
·163 → ·166 → ·166 → ·168 → ·165 → ·164 → ·163 → ·165 → ·167 → ·165 → ·166 → ·167 →
·167 → ·168 → ·168 → ·168 → ·167 → ·167 → ·167 → ·168 → ·166 → ·165 → ·165 → ·164 →
·164 → ·165 → ·165 → ·165 → ·165 → ·165 → ·165 → ·165 → ·164 → ·163 → ·163 → ·164 →
·1637 → ·1637 → ·1632 → ·1636 → ·1629 → ·1620 → ·1618 → ·1616 → ·1631 → ·1624 →
·1623 → ·1631 → ·1631 → ·1632 → ·1632 → ·1626 → ·1626 → ·1627 → ·1624 → ·1625 →
·1623 → ·1622 → ·1626 → ·1631 → ·1639 → ·1637 → ·1637 → ·1631 → ·1633 → ·1630 →
·1635 → ·1632 → ·1634 → ·1641 → ·1637 → ·1634 → ·1636 → ·1628 → ·1627 → ·1625

實驗 II. 同じ骰子で 6 の面の出る確率を求め,實驗 I と同樣に 100 投ごとに確率の記入を行つたのが第65表である.

第 65 表

6 の目の出た確率

·17 → ·195 → ·193 → ·175 → ·162 → ·160 → ·169 → ·164 → ·166 → ·165 → ·164 → ·165 →
·164 → ·167 → ·164 → ·165 → ·162 → ·164 → ·162 → ·159 → ·158 → ·157 → ·155 → ·155 →
·156 → ·158 → ·157 → ·159 → ·161 → ·159 → ·160 → ·159 → ·159 → ·159 → ·160 → ·159 →
·158 → ·158 → ·160 → ·160 → ·159 → ·160 → ·160 → ·161 → ·161 → ·161 → ·162 → ·162 →
·162 → ·161 → ·161 → ·161 → ·160 → ·161 → ·162 → ·162 → ·162 → ·163 → ·163 → ·163 →
·1637 → ·1629 → ·1641 → ·1638 → ·1636 → ·1638 → ·1634 → ·1645 → ·1644 → ·1643 →
·1640 → ·1642 → ·1634 → ·1626 → ·1621 → ·1622 → ·1623 → ·1619 → ·1613 → ·1619 →
·1623 → ·1629 → ·1626 → ·1628 → ·1629 → ·1623 → ·1622 → ·1621 → ·1627 → ·1626 →
·1633 → ·1632 → ·1632 → ·1631 → ·1626 → ·1626 → ·1631 → ·1625 → ·1626 → ·1631

この場合實驗 I と反對に,初め確率が著しく高く,二三百投の試驗だけでは 0·19 か 0·18 位であらうと豫想せられたのに,七八百投から先きは漸次確率小となつて, 6000 投位からは 0·1613 と 0·1645 との間を上下して居るが恐らく確率はここでも 0·163 あたりに在るのであらう. 偶然にも 1 の目の出る確率と同一であつたが,かくの如き一致は寧ろ例外である. 第51圖 B はこ

れを圖示したものである.

統計的定義によればそれぞれの面の出る確率は骰子によつて異なり,余の使用した骰子では1の面の出る確率は0·163であつた. これを其の**骰子の特性**(又は個性)とする事が出來る. 若し又,他の骰子を使用すればそれぞれ特有の確率を示すであらう. 個性として確率を固有することは,骰子に限らず他の場合でも同一であつて,銀貨を投じた時,表又は裏の出る確率,一袋中に一定數の赤球及び黒球を入れて置き,これより赤球(又は黒球)を取り出す確率等にも有るので,生物の變異と同様に見做すことが出來る. かく,確率が骰子,貨幣等使用する個體によつて異なるが爲に,確率を定めんとならば,その個體に多數の實驗を施して決定しなければならぬ. これは不便な様ではあるが正しい事實上の問題である. 自然科學の範圍内に於ける確率も多くは實驗と觀察とによつて決定せらるべきものであるが,統計を取る地方や,人種別や,時代や,其の他の條件で其の確率がちがふのである. かく實驗の場合々々に特性があるから觀測又は實驗は異種のものが混合して居てはならない. 骰子ならば必ず一個の骰子を連續多數回同様なる動作をなし,同様なる板上に落下せしめるのが原則である. 初め3000回は甲の骰子を以て試驗し,後の7000回は乙の骰子を以てする如きは,全く沒意義であつて,これを分ち3000回の實驗によりて甲の骰子の確率と,次に7000回の實驗によりて乙の骰子の確率とを,別々に計算すべきである. 常に同種の事象を集めてその觀測を基とせねばならぬ. 統計學に於ても,なるべく純粹なる條件内で觀測を行ふのもこの理由によるのである.

次に起る問題は,確率は何回の觀察又は實驗を基礎として近似値を計算すべきやと云ふことであるが,これは全く規定することが出來ない. 觀測回數が多い程,その近似値は信頼することが出來ること勿論であつて,第一及第二實驗にて實驗回數の多くなるに從つて,近似値動搖の振幅が小となつて居る事實を見ても明瞭である. しかし,實驗の容易なる事象にありては數萬回の觀測必ずしも困難でないが,事象によつてはかく多數に同種なる觀測を集めることが出來ない者があり,僅に100回の觀測を以て滿足しなければならぬものもあれば,或は數十回の觀測すら不可能なるものもある. 又目的によりては多數の觀測を根據としなければならぬものもあるが,目的如何によりては粗略なる觀測で滿足できることもある. 故に一般的に規定する必要もないのである. 確率論は多數の觀測を行ひうる事象に就て研究を行ひ,斯くして知り得たる事を,少數觀測の事象に應用せねばならぬ事が多い. 之を要するに,

(1) 經驗的定義では確率の眞の値は存在するのであるが計算できない. 唯近似値を計算しうるのみである.

(2) 觀測總數Nが小なる時は其の近似値も深く信用することが出來ないが,觀測數が増加する時は信用度も漸次増し,遂に無限に多數の觀測を行ひたりとすれば,其の時は眞の確率となるのである.

(3) この確率(若くは其の近似値)は使用する骰子(又は貨幣等)によりて異ふのであつて,何れも個性が有る.

例 1.　第66表より朝鮮人における男児出生の確率を求む.

第 66 表

明治 44 年より昭和 3 年に至る朝鮮における
朝鮮人出生の性別による統計表.

	♂	♀	♂＋♀	男兒出生の確率	確率の標準誤差 m
明治 44 年	148,447	129,026	277,473	0·535	±0·0010
大正 1 年	223,507	197,993	421,500	0·530	±0·0008
大正 2 年	239,203	213,522	452,725	0·528	±0·0007
大正 3 年	231,292	208,321	439,613	0·526	±0·0008
大正 4 年	233,320	203,083	436,403	0·535	±0·0008
大正 5 年	292,377	260,443	552,820	0·529	±0·0007
大正 6 年	297,304	266,468	563,772	0·527	±0·0007
大正 7 年	299,657	270,538	570,195	0·526	±0·0007
大正 8 年	248,123	218,152	466,275	0·532	±0·0007
大正 9 年	251,286	217,435	468,721	0·536	±0·0007
大正 10 年	274,498	234.624	509,908	0·539	±0·0007
大正 11 年	308,894	276,525	585,419	0·528	±0·0007
大正 12 年	370,411	339,497	709,908	0·522	±0·0006
大正 13 年	363,478	317,350	680,828	0·534	±0·0006
大正 14 年	376,620	335,658	712,278	0·529	±0·0006
昭和 1 年	355,530	310,074	665,604	0·534	±0·0006
昭和 2 年	365,918	321,224	687,142	0·533	±0·0006
昭和 3 年	377,613	332,945	710,558	0·531	±0·0006

出生總數は第四列に,男兒出生數は第二列に記してある. 故に其の確率の近似値は直ちに計算することが出來る. 例へば明治四十四年では $\dfrac{148,447}{277,473}=0{\cdot}5349$ が求むる確率で,第五列はかくの如く計算した確率の近似値である. 第六列につきては第十四章を見よ. 尙ほかくの如き統計は人種又は民族,年代等を混

合してはならぬ. 時代と民族で大なる變化がある爲である[1].

例 2.　Mendel は豌豆の雜種を作り,さらに此の雜種を自花受粉で結實せしめ,之を播種した所,形質が最初の二種の階級に分離するを觀た. 表では A を優性 α を劣性とし二階級をあらはす. 其の觀測せる數は次表の通りであるが,これより分離の比を確率にてあらはせ.

第 67 表

	形　質	A	α	A の確率	α の確率
1	種子の形 (丸形又は丸味を有するもの A / 角張つて皺のあるもの α)	5474	1850	0·747	0·253
2	子葉の色 (黃色子葉を有する種子 A / 綠色子葉を有する種子 α)	6022	2011	0·750	0·250
3	種皮の色 (灰褐色 A / 白 色 α)	705	224	0·759	0·241
4	莢の形狀 (單純にして膨れた莢のもの A / クビレタ莢を有するもの α)	882	299	0·747	0·253
4	未熟なる時の莢の色 (綠色莢 A / 黃色莢 α)	428	152	0·738	0·262
6	花の位置 (腋生花 A / 頂生花 α)	651	207	0·759	0·241
7	莖の丈け (長莖 A / 短莖 α)	787	277	0·740	0·260

(1) 男兒出生數の女兒出生數より大なることは Süssmilch 以來知られてゐるが,その精密な數字に至つては時代により,殊に民族により,大なる差異がある. 問題 (11) と比較して,朝鮮人に於て男兒出生の確率の大なることに注意せよ. 但し,屆け出られたる男兒出生の確率大なりと雖も實際男兒出生の確率が大であるとは限らない. 一般に男兒出生の確率の大きいのは,ブルガリア,エヂプトの如くマホメット教徒が國民の多數を占むる國であつて,これらの教徒では女兒の出生を重んぜざる結果,女兒出生の屆出を怠る等の二次的,人意的の影響大なるが爲であるとせられてるが,朝鮮人に於ても類似の原因が考へうるのである.

表には七つの形質のみを掲げたのであるが,各形質いづれも A と α とに分離したのである. 觀測總數は $A+\alpha$, A の確率 p は

$$p=\frac{A}{A+\alpha}\quad である.$$

第五列には A の確率,第六列には α の確率を記入した.

4. 確率の數學的定義

統計的定義に從へば,近似値ではあるが,必要な場合は何時でも確率を計算することが出來る. けれども必ず統計によらなければならぬから,數學的研究には不便此の上もない. 若し統計によらず,先驗的に確率を計算することが出來れば,至極便利であらう.

今等質にて偏重なき物質を以て,幾何學的に完全なる圓板を作り,これを投ずれば,表が出ることと裏が出ることとは同程度に期待せられ,裏が多く出るとも限らず,表が多く出るとも考へられない. このことを裏の出ることと表の出ることとは**同程度に確らしい**,又各面の出る難易は同程度なりと云ふ. 而して表裏二つの面が出ることが同程度に確らしい時に,其の一面なる表の出る確率は $\frac{1}{2}$ なり,裏の面の出る確率も亦 $\frac{1}{2}$ なりと云ふ. 一般には次の如く定義する.

定義 II. (**確率の數學的定義**) 一の原因の結果として, n 個の事象のうち何れか一つが起るものとす. 今此等の事象のうち A なる事象の起る確率とは,事象の總數 n を以て A 事象の起る數 a を除したる商 $\frac{a}{n}$ を云ふ. 但し各事象の起ることは同程度に確らしきものとす.

此の定義は複雜であるが,上例にあてはめれば,一の原因とは貨幣を振ることである. n 個の事象とは表の出ること裏の出ることの合計二つの事象, A なる事象とは表の出ることである. 表の出るのはただ一通りの場合であるから $a=1$ である. 故に A 事象(表の出ること)の起る確率 p は $\frac{a}{n}=\frac{1}{2}$ となる. 上例に於て,若し表の出ることと裏の出ることとが同程度に確らしくない時は,確率は此の定義では計算が出來ない. 圓板の面が正圓でなかつたり,圓板の厚さに不同があつたり,稜角の一部が磨滅して居たり,一面に深い彫刻が有つたり,物質が等質で無かつたりすれば,兩面の出ることは同程度に確らしくないから,表の出る確率は $\frac{1}{2}$ でない. 故に以下數學的定義で確率を定むる時は,常に「事象の起ること同程度に確らしい」なる假定があるものと考へなければならぬ. たとひ一々斷らずともそれは省略してあるものである. 數學的定義で計算せる確率を**先驗的確率**又は**數學的確率** (a priori probability, mathematical probability) と云ひ經驗的確率或は統計學的確率と區別して居る. 數學で取扱ふのは前者の方が多いが,統計學で取扱ふのは後者の方が多い. かくの如く,數學的に確率を定義する場合は,貨幣を投ずる時の外,尚一般に次の様なものがある.

例 3. **雙六の賽(骰子)** 等質なる物質で幾何學的に完全なる正六面體を作り,六つの面にそれぞれ 1 より 6 までの文字を記したりとせよ. この骰子を投じ, (i) 1の目の出る確率, (ii) 1の目の出ない確率,及び (iii) 奇數の目の出る確率を求む.

解. (i) 面は六個であり,而して上記の如く完全なる骰子な

らば,六つの面の各の出ることが同程度に確らしい. 而して

$n=6 \quad a=1$ であるから,

1の目の出る確率 $p_1=\dfrac{1}{6}$ である.

2, 3, 4, 5, 6 の何れかの面の出る確率も亦同一である.

(ii) 1の目の出ないこと即ち1以外の目の出ることは2か, 3か, 4か, 5か, 6か,何れかの出ることである. 即ち, $a=5$ である. 故に1の目の出ない確率 (2, 3, 4, 5, 6 の目の何れかが出る確率) p は,

$$p=\frac{a}{n}=\frac{5}{6}$$

である.

(iii) 奇數の目は 1, 3, 5 の三個がある. 故に奇數の出る場合は6個の場合のうち3個の場合である. $n=6, \ a=3$

即ち,奇數の目の出る確率は

$$p=\frac{a}{n}=\frac{3}{6}=\frac{1}{2}$$

である.

例 4. 壺中の球 一壺中に白球3, 赤球2を入れて置き,今目隱をしてこの壺中より一球を取り出すに,

(i) 其の球が白球なる確率,

(ii) 其の球が赤球なる確率を求む.

(iii) 又,壺中に白球 a, 赤球 b, 黒球 c 個,・・・・ ありとし壺中に手を入れて取りたる球の,白球,赤球,黒球なる確率を求む. 但し球は皆同じ形狀と,同じ大さと,同一の比重と,同一の表面の滑さとを備へ,其の他硬さ彈性等皆同一であつて,唯色のみ識別せらるるものとす. 即ち,壺中の五球の何れが取り出されることも同

程度に確らしい.

(i) 取り出さるるのは五球の内何れかであるから,事象の總數(取り出される場合の數)は5である. 而して白球の取出される數は其の中の3, 即ち $n=5, \ a=3$. 故に其の確率は,

$$p=\frac{3}{5}$$

である.

(ii) 又白球ならざる確率,即ち赤球の取り出される確率を q とすれば,同樣に球の合計の數を以て赤球の數を除したるもの,

$$q=\frac{2}{5}$$

である.

(iii) 一般に一壺中に白球 a, 赤球 b, 黒球 c, ・・・・ ありとし,其の壺中に手を入れて,白球,赤球・・・・ を取り出す確率を求むるに,次の通りである.

白球を取り出す確率, $p_1=\dfrac{a}{a+b+c+\cdots\cdots}$

赤球を取り出す確率, $p_2=\dfrac{b}{a+b+c+\cdots\cdots}$

黒球を取り出す確率, $p_3=\dfrac{c}{a+b+c+\cdots\cdots}$

・・・・・・・・・・・・・・・・・・・・・・, $p_4=\cdots\cdots\cdots\cdots\cdots\cdots$

何となれば,壺中の球の總數は $a+b+c+\cdots\cdots$ であり,其の中の白球は a であるからである. p_2 以下も同樣の理である. 又,

白球ならざる確率 q_1 は, $q_1=\dfrac{b+c+d+\cdots\cdots}{a+b+c+d+\cdots\cdots}$,

赤球ならざる確率 q_2 は, $q_2=\dfrac{a+c+d+\cdots\cdots}{a+b+c+d+\cdots\cdots}$,

黒球ならざる確率 q_3 は, $q_3=\dfrac{a+b+d+\cdots\cdots}{a+b+c+d+\cdots\cdots}$,

例 5.　トランプの札　52枚のトランプの札を四人に13枚づつ配り,四人の内或一人甲にクラブのAが分配せられる確率を求む.

52枚のトランプの札中クラブのAは唯一枚あるのみ,全體の札を平等に四人に分配するのであるが,所定の札は甲乙丙丁四人に渡ることは同程度に確らしい. 故に甲にこの札が來る確率は$\frac{1}{4}$である. トランプの札の問題は複雜で難解なのがあるが,ここでは其の最單純な一例をあげるに止めておく.

例 6.　抽籤　一萬本の籤あり,その中一等の當り籤1本,二等が5本,三等が20本ありとす. 今無心に1本の籤を引きたる時,各等當籤の確率をとふ.

籤の總數 10000 本であるから $n=10000$, この 10000 本の中何れを引くかは全く偶然で同程度に確らしい. 而して, a は一等では 1, 二等では 5, 三等では 20 であるから,確率は

一等當籤の確率　$p_1=\frac{1}{10000}$,

二等當籤の確率　$p_2=\frac{5}{10000}$,

三等當籤の確率　$p_3=\frac{20}{10000}$,

である.

5.　經驗的定義と數學的定義とについて

各事象の起ることは同程度に確らしいものとすなる假定が眞ならざる場合は,數學的定義で確率を定めることが出來ない. しかし同程度に確らしいことを如何にして知ることが出來るかは,數學者間に屢論議せらるる所であつて,雙六の賽では完全に等質で偏重なく,又形も完全に正六面體で稜も磨滅して居ない時に,各面の出ることは同程度に確らしいと考へるのである. 其の他未だ吾々の知らぬことで,「各面の出ること同程度に確らしい」なることに反する事實があるかも知れぬが,若しあれば其の原因を悉く除いた理想的の骰子を想像して,此の骰子について確率を計算するのである. 貨幣の場合,抽籤の場合,壺中の球の場合等皆,同程度の確らしさを妨害する事實はことごとく除去した時の確率である.

かくの如き完全なる理想的の骰子,又は貨幣等は人工では製作し得ない. 唯觀念的に作りうるのみであるから,定義IIは極めて抽象的の性質を有する. しかしこの抽象的な數學的定義が,確率論に極めて重要な意義を有するものである. 經驗的定義に從ふ時はこれと全く反對に極めて具象的の性質を有するので,骰子の場合では一般的骰子でなく,實驗に使用したる或一個の骰子についての確率を定めるのであつて,他の骰子は亦別の確率を其の個性として所有して居るのである. 故に,經驗的確率は必ずしも數學的確率に一致しない. 然し,數學的確率と經驗的確率とが,全く異なるものかと云ふに決して然らず. 實驗Iに使用した予の骰子は, 1の目を出す確率 0·1625 を個性として所有して居たのであるが,若し,これと同一條件でなるべく完全な正六面體を多數に作れば(或は無數に多く作れば),この多數の骰子は皆1の目を出す確率がちがふ. これを經驗的に算出すれば,其の確率の平均は恐らく $0{\cdot}1666\cdots=\frac{1}{6}$ となるのであ

らう．かくする時初めて,經驗的確率は先驗的確率に一致するのである．

例 7．出產する子供の男兒なる確率を求む． 簡單に考へれば事象の總數は男子の生れることと女子の生れることの二つであつて,この二つの起ること同程度に確らしい．故に,男子の生れる確率は $\frac{1}{2}$ である．

以上の解答は正しくない．例 1 及問題11に示す通り,男兒出生が女兒出生を超過して居ることは,古今,東西一致せる結論である．然らば先驗的確率と經驗的確率とが一致しないのは,何に原因するのであるか．何れか一方が誤りであるかを考へねばならぬ．

(i) 經驗的確率は近似值であるから,屢偶然の爲に影響せられて,誤差をともなふものである．故に男兒出生の確率が 0·5 より大であるのは偶然に起つた誤差であるか否かを考へねばならぬ．しかるに第十四章に解說する通り誤差では說明しきれない．卽ち,誤差の影響を除去しても,尙ほ,男兒出生の確率は 0·5 よりも大である．

(ii) 然らば,統計的確率に男子出生を不當に大ならしむる原因が有るのではないかと考へねばならぬ．これは考へらることであつて,例 1 の註に述べた如く朝鮮人やマホメット教徒では民族の習慣思想等が女兒出生の實數を歪める結果,不當に男兒出生の確率が大となつたのであらう．斯の如き特別なる場合もあるが,其の他の民族に於ける男兒出生過剩の狀況はよく一致し,女兒 100 に對し男兒 105—106 であつて,單に誤差や習慣

のみで說明することは不可能である．

(iii) 經驗的確率に誤りがないとすれば(特種な民族は除く),數學的定義で男兒出生確率を $\frac{1}{2}$ とするのが誤りでなければならぬ．男子が生れることと女子が生れることが同程度に確らしいとした此の假定に誤りがあるのである．少なくとも人類の場合ではこの二つの事象は同程度に確らしくないのである．前記の理想的,抽象的の銅貨や骰子を考へた場合はいいが,現實の問題ではこの假定を置くことは極めて危險である．若しこの假定を置かんとならば,次の事實を確めてからでなければならぬ．

(a) 今日の細胞學說に由れば,人類の精子には X-染色體を有するものと Y-染色體を有するものとが殆ど同數に出來,この X を有するものが卵子を受精せしむれば女兒生れ,Y を有するもの卵子を受精せしむれば男子が生れると云ふ．Painter 等の此の說が正しいと假定して,X を有する精子と Y を有する精子とが全く同數に出來,又同數だけ射出せられること．

(b) 子宮及輸卵管における二種の精子の遊走速度や受精率が同一であること．

(c) 胎兒時代における死亡の割合(流產の割合)が兩性一致せること．

(d) 死產の兩性比率も同一であること．

(e) 胎兒の全經過に於て內分泌等の爲に性が變更せらるること絕對になきこと．

これらの何處かに假定に反する所あらば,兩性の生れること

同程度に確らしくないこととなる．然るにこれらの諸事項の大部分が未だ明でない今日,上記の假定を置くは誤謬の基であつて愼むべきである．故にこの問題の如きは,數學的定義を設けることが出來ず,單に統計的定義による外はないのである．男女出生の場合に限らず,現實の問題になれば,統計的觀察によつて確率を定めねばならぬこと,實驗I及實驗IIの如く賽を投ずる場合ですらその通りである．

6. 基本的定理

定理 I. (i) 一事象が必ず生起する時は其の事象生起の確率は1である．(ii) 生起せざること明白なるものの確率は0である．(iii) 而して確率は常に0或は1に等しいか,或はこの間にあり卽ち確率をpとすれば

$$0 \leqq p \leqq 1 \cdots\cdots\cdots\cdots\cdots\cdots \text{(VIII. 3)}$$

である．

證明．

(i) 觀測の總數をNとすれば,或る事象が必ず生起するのであるから事象の起る回數も亦Nである．故に[1]

$$p = \lim_{N \to \infty} \frac{N}{N} = 1$$

(ii) N回の觀測總數に於て,事象は一回も起らないのであるから，$P=0$，故に

[1] 定理の證明はなるべく經驗的定義から行ふこととし,數學的定義での證明は確率論の教科書に普通出て居るからなるべく省略する．しかし,定理は數學的確率にも經驗的確率にも適用せらるるのである．

$$p = \lim_{N \to \infty} \frac{P}{N} = 0$$

(iii) N觀測に於て事象の起る回數をPとすれば,常に,

$$0 \leqq P \leqq N$$

であるから,

$$0 \leqq \lim_{N \to \infty} \frac{P}{N} \leqq \lim_{N \to \infty} \frac{N}{N}$$

$$0 \leqq p \leqq 1.$$

一般に,數學的に定義せられた場合でも,經驗的に定義せられた場合でも，$p=1$なりとは此の事象の生起が確實なるを意味し,俗に100パーセント確かであると云ふのと同義である．1に甚近い時はこの事象の生起は殆ど確かであることを意味する．又pが0·5に近き時は生起と不生起とは殆ど相半することになり,0に近き時は此の事象の生起することが殆どなきことを示すのである．

例 8. 一囊中に白球5を入れて置きこれより一球を取り出すに,其の球の白球なる確率は1であり,赤球を取り出す確率は0である．何となれば,囊中には白球のみ入れてあるから,白球を取り出すことは確かなるが爲である．卽ち

$$P = \frac{5}{5} = 1 \quad \text{で計算しうる.}$$

又,この囊より赤球を取り出す確率は0である．卽ち

$$p = \frac{0}{5} = 0$$

で計算出來る．これは何れも數學的定義の場合である．

定理 II. 一事象の生起する確率をp生起せざる確率をqとすれば

$$p+q=1 \cdots\cdots\cdots\cdots\cdots\cdots (\text{VIII}, 4)$$

である.

證明. N 回の觀測に於て一事象の生起する回數を P とし,生起せざる回數を Q とすれば,

$$P+Q=N$$

$$\frac{P}{N}+\frac{Q}{N}=1$$

故に $$\lim_{N\to\infty}\frac{P}{N}+\lim_{N\to\infty}\frac{Q}{N}=1$$

即ち $$p+q=1$$

この公式は極めて必要なものである.

定義. 或る事象と反對の事象を**餘事象**と云ふ. 例へば銅貨を投じたる時表の出ることと裏の出ることは互に餘事象である. 又骰子を投じ,1の出ることと,1の出でざることと; 又奇數の出ることと,偶數の出る事とは互に餘事象である. 出生の場合では男兒出生と女兒出生とは,亦互に餘事象である. 但し兩性兒は生れざるものと假定す.

系 1. 一事象の確率を p とすれば其の餘事象の確率は $1-p$ である. 即ち,餘事象の確率を q とすれば,

$$q=1-p$$

演習問題

確率 p の計算には,大別すると三種の方法がある.

(i) **定義IIにより,先驗的確率の直接計算を行ふ場合**

これには複雜な難問題が多く,生起する事象の總數の計算や,其の中A事象の生起する數の計算は,誤り易くして愼重な注意を要することが多い. この種の問題を解くには,順列及組合せの定理や,二項定理等を應用することが多いが,其の比較的容易の二三の例のみを本章の問題として揭げることとし,幾多面白い問題を割愛した. 難問のため道草を食はない爲に,又これらの難問の多くは,第九章の定理を應用すれば,容易に解くことが出來るものが多いからである. 興味のある諸君は文獻(5)(3)又は(4)を讀まれよ.

(ii) **定義Iにより統計的に計算する場合**

即ち經驗的確率である. この種の問題には,難かしいのが無く,注意すべき點は寧ろ統計の取り方にあるのである. 故に,本章の問題には一問を附したのみである.

(iii) **i又はiiにより計算せられたる確率を用ひ,第九章以下の定理により間接に計算しうる確率**

數學に於ても自然科學に於ても必要なものである. 且つ(i)の場合の難問は大抵この方法でときうる. 第九章以下の問題を見よ.

1. 二個の骰子を投じて其の和10なる確率を求む.

解. 一個の骰子を投じたる時出る面は六通りあり. 故に二個を投じたる時二つの骰子の出る面の組合せは6×6通りあり(事象の總數). 而して骰子の目の和10なる組合せは次の3通りあり.

第一の骰子		第二の骰子
(6	+	4)
(5	+	5)
(4	+	6)

故に,二つの骰子の出る目の和10なる確率, p は

$$p=\frac{3}{36}=\frac{1}{12}$$

となる.

2. 二個の骰子を投じ一の骰子は6他の骰子は5の出る確率を求む.

解 I. 事象の總數は問題1により6×6通りあり. 其の中第一の骰子6第二の骰子5の出る場合は唯一通りあり,又第二の骰子6第

一の骰子5の目の出る場合も一通り,合計二通りあり. 而して一骰子に6,他の骰子に5の出る場合は,此の二通より外に無い. 故に,求むる確率は $\dfrac{2}{36}=\dfrac{1}{18}$

解 II. 複確率の定理を用ひてこれを解くことが出來る.

3. 二個の骰子を投じて,二つとも6の目の出る確率を求む.

(答 $p=\dfrac{1}{36}$)

注意. 前題に準じて說くか,又は複確率の定理を用ふるか,何れでも出來る.

4. 完全なる圓板5枚を投じ,5枚とも表の出る確率を求む.

解. 一枚の圓板を投ずれば表が出るか裏が出るかで,此の二通りより外にない. 今5枚同時に投ずれば表裏の出方は 2^5 通りあり. 而して全部表の出る仕方は其の內唯一通りあるのみ,故に $\dfrac{1}{32}$ が求むる確率である.

注意. これも複確率の定理を用ひた方が容易である.

5. 一嚢中に n 個の同樣なる球あり, n の各が取出さるることは同程度に確らしいとする. 又この球は1より n 番まで番號が附しあり. これより同時に二球を取出すに,それが1番と2番の球なる確率を求む.

解. n 個の球より2個の球を取出す方法は,組合せの法則より ${}_nC_2$ 通りあり. 而して1番と2番とを同時に取出す方法は一通りである. 故に求むる確率, p は

$$p=\frac{1}{{}_nC_2}=\frac{2}{n(n-1)}$$

である.

注意. 複確率及全確率の定理を用ふれば,一層容易にこの問題を解くことが出來る. 又曰く,若し二球を同時に取出さず,初めに1番の球,第二回目に2番の球を取出す確率は,上記と異つて

$$p=\frac{1}{n(n-1)}\quad となる.$$

6. 前問の袋中より5個の球を取出すとき, 1, 2, 3の番號の球が,何れか唯一個だけ入り居る確率を求む.

解. 事象の總數は n 個の物から5個づつ取りたる組合せの數であるから, ${}_nC_5$ 通りある. しかるに1球を含み2, 3の球を含まざる組合せの數は ${}_{n-3}C_4$ 通りである. 何となれば1の球を含み2, 3の球を含まざる組合せの數は, n よりこの三つの球をのぞき,他の $n-3$ 個の球より4個づつの組合せを作り,これに1の球を加へたるに等しいからである. 故に5個の球をとり1, 2, 3の球の中何れか1個のみを含む組合せの數は $3{}_{n-3}C_4$ となる. 故に求むる確率 p は

$$p=\frac{3{}_{n-3}C_4}{{}_nC_5}=\frac{3(n-3)(n-4)(n-5)(n-6)5!}{4!n(n-1)(n-2)(n-3)(n-4)}$$

$$=\frac{15(n-5)(n-6)}{n(n-1)(n-2)}$$

注意. 複確率及全確率の定理を用ひても出來る.

7. 一袋中に白球 a 個,赤球 b 個,黑球 c 個あり,これより三個の球を取出し,二つが白球,一つが赤球なる確率を求む.

解. $(a+b+c)$ 個の球を入れたる袋より,三個の球を取出す仕方は ${}_{a+b+c}C_3$ 通りあり. 又白球二個を取出す仕方は ${}_aC_2$ 通りあり,又赤球一個を取出す仕方は ${}_bC_1$ 通りあり. 故に白球二個赤球一個を取出す仕方は ${}_aC_2\times{}_bC_1$ 通りあり. されば求むる確率, p は

$$p=\frac{{}_aC_2\times{}_bC_1}{{}_{a+b+c}C_3}=\frac{a(a-1)b\times 3!}{2!(a+b+c)(a+b+c-1)(a+b+c-2)}$$

$$=\frac{3.a.b.(a-1)}{(a+b+c)(a+b+c-1)(a+b+c-2)}$$

尙この問題も複確率の定理を用ひれば容易である.

8. 100本の籤あり,この中當り籤13本あり,今三本を引きて二本當り籤のある確率をとふ. (答 $p=\dfrac{2\times 13\times 87}{11\times 49\times 100}$)

これも第八章の定理を用ひた方が容易である.

9. 8人の騎士圓卓を圍む,其の中 A, B 二人が相離れずに並ぶ確率をとふ. 又この二人が並ばざる確率をとふ.

解. 圓順列なるが故に八人の人が並ぶに $7!$ 通りあり. A, B 二人が常に離れざる場合の順列の數は $2!\times 6!$ 通りあり. 故に二人がならぶ確率 p は

$$p = \frac{2 \times 6!}{7!} = \frac{2}{7}$$

並ばざる確率 q は

$$q = 1 - \frac{2}{7} = \frac{5}{7}$$

10. 十人の人が圓卓を圍み會議を開く時,順序に意を用ひざれば A, B, C, D 四人が相離れざる確率を求む. $\left(答\ \ p = \frac{1}{21}\right)$

11. 第68表は歐洲諸國の女兒出生100に對する男兒出生數である. 今此の表より男兒出生の確率を計算せよ.

第 68 表

女兒出生 100 に對する男兒出生數.

國名 \ 年號	1876—85	1886—95	1896—1905	1906—1914	$\pm 3m$
獨乙	106·1	106·1	106·0	106·1	±0·2
オーストリア	106·5	106·7	106·3	106·3	±0·2
スウィス	106·3	105·6	105·1	105·6	±0·7
オランダ	106·5	106·3	106·2	105·8	±0·5
ベルギー	105·8	105·6	105·7	105·3	±0·5
フランス	106·2	105·8	105·3	105·6	±0·2
デンマーク	105·8	105·7	105·8	105·4	±0·8
ノールウェイ	106·5	106·9	106·3	106·3	±0·8
スウェーデン	106·2	106·0	106·2	106·5	±0·6
ハンガリー	105·6	105·7	106·1	106·1	±0·3
イタリア	106·9	106·6	106·5	106·2	±0·2
フィンランド		105·9	(1886—1905)		±0·7
ブルガリア		109·1	(1881—1905)		±0·4
ルーマニア		108·6	(1876—1895)		±0·3
エヂプト		109·1	(1918—1923)		±0·3
日本		105·2	(1891—1922)		±0·1

括弧の中にある數は年號である. 又第6列は $\pm 3m$ 即ち標準誤差の三倍であつて各列の比率數は $\pm 3m$ の變動は偶然起りうることを示す. 詳しくは第十四章を見よ.

確率論に關する參考文獻

以下數章は專ら確率論に關する記述であるが其の參考書をあげると:

1. Bruns, H.:- Wahrscheinlichkeitsrechnung u. Kollektivmasslehre. Leipzig 1906.

2. Burnside, William :- Theory of probability. Cambridge, 1928.

3. Czuber, Emanuel :- Wahrscheinlichkeitsrechnung u. ihre Anwendung auf Fehlerausgleichung, Statistik u. Lebensversicherung, I. u. II. Bd. Leipzig u. Berlin, 1928.

4. Fisher, Arne :- The mathematical theory of probabilities 2nd. ed. New-York, 1930.

5. 林鶴一,刈屋他人次郎:- 公算論(確らしさの理論) 東京,大倉書店 7版, 大正15年

6. 龜田豐治朗:- 確率論及其應用,東京 昭和四年.

7. Keynes, John Maynard :- A treatise on pro bility, London, 1929.

8. 渡邊孫一郎:- 確率論,東京,文政社

第九章
確率の加法及び乘法

1. 排反事象

今銅貨を投じて,表が出るならば決して裏は出ない. 又裏が出る時は決して表が出ない. 即ち表が出ることと裏が出ることとは互に兩立しない. この二つを互に排反すると云ひ,又,二つの事象を排反事象と云ふ. かくの如き事象は他に幾らもある. 出生が男兒であることと女兒であることは兩立しないから,この二つの事象は排反事象である. 抽籤が當ることと當らないこと,骰子を投じ1の目が出ることと1の目が出ないこと,或る事が成功することと成功しないこと等は,又互に排反である[1]. 又骰子を投じて,1の目が出れば 2, 3, 4, 5, 6 の何の目も出ない. 六つの目のうち何れか一つが出れば,他の五つは出ないのである. かくの如き場合には,六つの目の各の出ることは互に排反であると云ふ. 詳しく云へば,1の目の出ることと 2, 3, 4, 5, 6 の目の出ることとは互に排反である. 2 以下の場合も同樣である. かく同時に生起する事の出來ない事象を排反事象と云ふ. 一般に次の如く定義する.

定義 I. 事象 $E_1, E_2, E_3, \cdots E_n$ あり; 其の中の一事象 E_k が起れば,他は生起すること絶對にない場合に,これらの事象は互

(1) 遺傳學で相對形質と云ふも,二つの排反なる形質のことである.

に排反なりと云ひ,又此の事象を排反事象と云ふ.

上記の例では排反事象は簡單であるが,次の例ではやや複雜になる.

例 1. 二個の骰子を投ずれば,出る目の組合せは次の三十六通りある. 羅馬數字は第一骰子の出る目の數,アラビア數字は第二骰子の目の數とす.

(I 1), (I 2), (I 3), (I 4), (I 5), (I 6), (II 1), (II 2), (II 3), (II 4), (II 5), (II 6), (III 1), (III 2), (III 3), (III 4), (III 5), (III 6), (IV 1), (IV 2), (IV 3), (IV 4), (IV 5), (IV 6), (V 1), (V 2), (V 3), (V 4), (V 5), (V 6), (VI 1), (VI 2), (VI 3), (VI 4), (VI 5), (VI 6).

この組合せは皆異つて居るから或る組合せが出れば,他の35通りの何れの組合せも出ない. 故にこの組合せは何れも互に排反である.

注意: 一骰子を投じ4の目の出ることと奇數の目の出ることとは排反であると云へるが,しかし3で割り切れる數の出ることと偶數の出ることとは排反でない. 何となれば若し6が出れば3で割り切れると同時に偶數であるからである. 又二骰子を投じ,第一の骰子に1の出ることと第二の骰子に1の出ることとは排反でない. 何となれば,第一の骰子に1が出ても,第二の骰子には1が出ることもあり,又1が出ないこともあり,不定であるからである.

定理 I. 全確率の定理(又確率の加法) n 個の排反事象あり; その事象の何れかが生起する確率は各事象生起の確率の和である. これを**全確率の定理**,又は**確率の加法**と云ふ. これを式で表はせば次の通りとなる.

排反事象の確率をそれぞれ $p_1, p_2, \cdots p_n$ とする時,これら事

象の何れかが生起する確率 p は

$$p = p_1 + p_2 + p_3 + \cdots p_n \cdots\cdots\cdots\cdots (\mathrm{IX}, 1)$$

である.

證明. N を極めて大なる數とし,N回の試驗に於て第一事象は P_1 回,第二事象は P_2 回,$\cdots\cdots$ 起りたりとせよ; しかる時は

$$p_1 = \lim_{N\to\infty} \frac{P_1}{N},$$

$$p_2 = \lim_{N\to\infty} \frac{P_2}{N},$$

$$p_3 = \cdots\cdots,$$

$$p_4 = \cdots\cdots,$$

$$\cdots\cdots\cdots\cdots$$

而して,$E_1, E_2, E_3, \cdots E_n$ は排反事象なるが故に,これら二つ以上の事象が同時に起ることなし. 故に,何れか一つ事象の起る回數 P は

$$P = P_1 + P_2 + \cdots P_n,$$

故に,これら事象の何れか一つが起る確率 p はこの式の兩邊を N で除したものから計算しうる. 即ち,

$$p = \lim_{N\to\infty} \frac{P_1 + P_2 + \cdots\cdots}{N} = p_1 + p_2 + \cdots p_n$$

系. 二つの排反事象があつて,其の確率の和が 1 なる時は,この事象は互に**餘事象**である.

例 2. 骰子を投じたる時,奇數の目の出る確率をとふ. 骰子の奇數の目の出ることは 1, 3, 5 の目の何れかが出ることである. 且つ三つの事象は互に排反であり,それぞれの確率は $\frac{1}{6}$ であるが故に,奇數の目の出る確率 p は上の定理に從ひ,

$$p = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}$$

である. 偶數の目の出る確率も亦同樣である.

全確率の定理が應用出來るのは,二個又は二個以上の互に排反なる事象なることと,此の n 個の排反事象の何れかが起る確率を求むる場合であることに注意せよ.

2. 獨立事象

今銀貨を二回連續して投ずるに,第一回に裏が出るも表が出るも,その事象は第二回に表又は裏の出る確率に影響しない. 又二個の骰子を同時に投じた時も同樣で,第一の骰子に 1 (或は他の目)の出ることは第二の骰子の或目の出る確率に影響せず,反對に第二の骰子に何の目が出ても第一の骰子の出る目の確率に影響することもない. 甲なる女が男又は女兒を生むことは,乙なる女の男兒(又は女兒)を生む確率に影響を與へず,乙が甲に影響することもない. かくの如き二つの事象を互に獨立なりと稱し,又略して獨立事象と云ふ.

定義 II. 二個又は二個以上の事象あり,其の內或事象の生起すると否とに拘らず,他の事象の起る確率が不變なる時は,是等の事象は**互に獨立**なりと云ひ,又,簡單に**獨立事象**なりと云ふ.

定理 II. n 個の獨立事象あり,其の事象が悉く起る確率は,各事象生起の確率の積である.

又次の如く現はすことが出來る.

$E_1, E_2, E_3, \cdots E_n$ を互に獨立なる事象とし,其の確率を夫々 p_1, $p_2, \cdots p_n$ とす; 然るときは,是等の事象が悉く起る確率 p は

$$p = p_1 \cdot p_2 \cdot p_3 \cdot \cdots \cdot p_n \text{ である.} \quad \cdots\cdots\cdots\cdots (\mathrm{IX}, 2)$$

例 3. 二個の骰子を同時に投じた際,甲に 1 の目が出で且つ

乙に6の目が出る確率を問ふ.

解. 甲なる骰子の1の目の出ることと,乙なる骰子の6の目の出ることとは互に獨立である. 故に,甲の骰子に1が出で,且つ乙の骰子に6が出る確率は各の確率の積である. 即ち

$$p=\frac{1}{6}\times\frac{1}{6}=\frac{1}{36},$$

これは第八章問題にある通り計算の結果は同一である.

定理IIの證明. N 甚だ大なる時, N 回の觀測に於て,第一事象が N_1 回起り,又第一の事象が起り且つ第二の事象も起りたる回數を P とす.

$$\frac{P}{N}=\frac{N_1}{N}\cdot\frac{P}{N_1}$$

なるが故に

$$\lim_{N\to\infty}\frac{P}{N}=\lim_{N\to\infty}\frac{N_1}{N}\cdot\lim_{N\to\infty}\frac{P}{N_1}.$$

而して $\lim\limits_{N\to\infty}\frac{P}{N}$ は第一及第二の事象の共に起る確率,即ち p,

$\lim\limits_{N\to\infty}\frac{N_1}{N}$ は第一の事象の起る確率,即ち p_1,

$\lim\limits_{N\to\infty}\frac{P}{N_1}$ は第一事象の起りたる場合に於ける第二事象の確率

である. 然るに二つの事象は獨立であるから, $\lim\limits_{N\to\infty}\frac{P}{N_1}$ は一般の場合における第二事象の確率 p_2 に等しい. 故に,上式は

$$p=p_1\cdot p_2$$

となる.

二個以上の獨立事象がある場合も,同一方法で證明が出來る.

3. 從屬事象

一壺中に白球5個,赤球3個を入れ置き,先づ一球を取り出し,

次にこの球を壺に還すことなく再び壺中より一球を取り出したとする. 第一回にもし白球を取り出したとすれば,第二回に白球を取り出す確率は $\frac{4}{7}$ となる. しかるに,第一回にもし赤球が出て居たとすれば,第二回に白球を取り出す確率は $\frac{5}{7}$ である. 即ち第一回に白球が出たか,或は赤球が出たかと云ふ事が,第二回目に白球(又は赤球)を取り出す確率に影響するのである.

又二個の壺あり,甲の壺には白球5,赤球3個入れてあり,乙の壺には白球のみ5個入れて居る時,もし甲に手が入れば白球を取り出す確率が $\frac{5}{8}$ であり; 若し乙に手が入れば白球を取り出す確率は1である. この場合には手が甲壺に入るか,又は乙壺に入るかと云ふことが白球を取り出す確率を左右するのである.

かくの如き二つの事象を**從屬事象**と云ふ. 一般に次の如く定義する.

定義 III. 二つ以上の事象ありて,其の内第一の事象の何れが起るかにより,第二の事象の起る確率が變化する時は,**第二の事象は第一の事象に從屬なり**と云ひ,又略して**從屬事象**なりと云ふ. 又第一事象及第二事象の何れが起るかに從つて,第三の事象の起る確率が變化する時は,**第三事象は第一及第二の事象に從屬なり**と云ふ.

獨立事象及從屬事象を總稱して,**複事象**と云ふ.

定理 III. $E_1, E_2, \cdots E_n$ を互に從屬する事象とし, E_1 の確率を p_1, E_1 が起りたる時の E_2 の確率を p_2, E_1, E_2 が共に起りたる時の E_3 の確率を $p_3, \cdots$, 順次斯くの如し; 今 $E_1, E_2, \cdots E_n$

が悉く起る確率を p とすれば,

$$p = p_1 \cdot p_2 \cdot p_3 \cdot \cdots\cdots p_n \cdots\cdots\cdots\cdots (\mathrm{IX}, 3)$$

である.

定理 II と III とを比較すれば,前者は獨立事象,後者は從屬事象に關する定理であるが,何れにしても悉く起る又は並び起る確率を求めるのであつて,常に個々の確率の積であらはされる.之の二つの定理を**複確率の定理** (Theorem of compound probability), 又は**確率乘法**と云ふ.

證明. 今 N 回の試行に於て,第一事象の起る回數 N_1, 其の確率 p_1, 又第一事象の起る N_1 回中第二事象の起る回數を N_2, 又その確率 p_2, 次に第一及第二事象の起る回數 N_2 内で第三事象の起る回數 N_3, 其の確率 p_3, $\cdots\cdots$ 第 1 第 2 $\cdots\cdots$ 第 $(n-1)$ 事象の起りたる回數 N_{n-1} 中第 n 事象の起る回數を N_n, 其の確率 p_n とすれば,

$$\frac{N_n}{N} = \frac{N_1}{N} \cdot \frac{N_2}{N_1} \cdot \frac{N_3}{N_2} \cdots\cdots \frac{N_n}{N_{n-1}},$$

即ち, $$\lim_{N\to\infty} \frac{N_n}{N} = \lim_{N\to\infty} \frac{N_1}{N} \cdot \lim_{N\to\infty} \frac{N_2}{N_1} \cdot \lim_{N\to\infty} \frac{N_3}{N_2} \cdot \cdots\cdots \lim_{N\to\infty} \frac{N_n}{N_{n-1}},$$

故に $$p = p_1 \cdot p_2 \cdot p_3 \cdots\cdots p_n$$

例 4. 一袋中に白球 5 個,黒球 10 個あり; これより四回つゞきて一球づつ取るに四球とも白球なる確率をとふ 又この袋より第一回に黒球,第二,第三及第四回に共に白球を取出す確率をとふ.

解 (i). 第一回に白球をとる確率 $p_1 = \dfrac{5}{15}$,

第一回に白球が出たる時第二回に白球なる確率 $p_2 = \dfrac{4}{14}$,

第一第二回に白球の出たる時第三回に白球なる確率 $p_3 = \dfrac{3}{13}$,

第一乃至第三回に悉く白球の出たる後第四回に白球なる確率 $p_4 = \dfrac{2}{12}$,

而してこれは從屬事象であるから,四回ともに悉くそろつて白球なる確率 p は

$$p = \frac{5}{15} \cdot \frac{4}{14} \cdot \frac{3}{13} \cdot \frac{2}{12} = \frac{1}{273}.$$

(ii). 第一回に黒球なる確率 $p_1 = \dfrac{10}{15}$,

第一回に黒球を取出したる時,第二回に白球を取出す確率 $p_2 = \dfrac{5}{14}$,

第一回に黒球,第二回に白球を取出したる時,第三回に白球を取出す確率 $p_3 = \dfrac{4}{13}$,

第一回に黒球を,第二回に白球,第三回に白球を取出したる時,第四回に白球を取出す確率は $p_4 = \dfrac{3}{12}$,

∴ 以上のことが悉くが起る確率, p は

$$p = \frac{10}{15} \cdot \frac{5}{14} \cdot \frac{4}{13} \cdot \frac{3}{12} = \frac{5}{273}$$

かく從屬事象では第一事象の起る確率を求め,この第一回の事象が起りたる時,第二回の起る確率を計算し,次に第一第二の事象の起りたる時,第三事象の起る確率を計算し,以下かくして最後の回までに至り,これらの確率の積を求むればいいのである.

例 5. 一袋中に白球 5, 赤球 3, 黒球 10 あり,今袋に手を入れて一球を取り,其の色を見てこれを袋にかへし,次によく袋の球を

混じて第二球を取り,其の色を見てこれを袋にかへす. かくの如く六球を取り第一回及第二回に白球,第三回及第四回に赤球,第五及第六回に黒球を取出す確率をとふ.

解. 一球をとるごとにその球をもとに返すのであるから,

白球をとる確率は常に $p_1 = \frac{5}{18}$,

赤球をとる確率は常に $p_2 = \frac{3}{18}$,

黒球をとる確率は常に $p_3 = \frac{10}{18}$.

故に,この際白球を取出すことと赤球を取出すことと黒球を取出すことは互に獨立である. されぱこれら六回の事象が並び起る確率はこれらの積である.

即ち, $$p = \frac{5^2}{18^2} \cdot \frac{3^2}{18^2} \cdot \frac{10^2}{18^2} = 0{\cdot}00066$$

これを例4に比較すれば,取り出された球が一回ごとに元に還されることが異つて居る.

注意 第一の事象 E_1 が起ると否とにかかはらず,第二の事象 E_2 の起る確率に變化がない時は, E_2 は E_1 に對して獨立である. 然る時は E_2 の起つたと否とにかかはらず, E_1 の起る確率にも亦變化がないのである. 故に, E_2 は E_1 に對して獨立であると同時に, E_1 は E_2 に對して獨立なのである. 故に,一の事象が他の事象に對して獨立であれば,これらの事象は互に獨立である. それを序でに證明して置く.

E_1 が先づ起り,次に E_2 が起るとし, E_2 は E_1 に對して獨立なりとす. さて, E_1 が起る確率は p_1, 起らざる確率は $1-p_1$, 又 E_1 の起ると否とにかかはらず E_2 の起る確率を p_2 とすれば,獨立事象の定理により E_1, E_2 の起る確率は $p_1 \cdot p_2$ である,次に E_2 が先づ起つたと考へよ; E_2 の起る確率はこの場合でも p_2 である,而して E_2 の起つたときの E_1 の確率を p_1', 又 E_2 が起らざる時の E_1 の確率を p_1'' とすれば, E_2E_1 共に起る確率は $p_2 p_1'$ である.

故に $$p_1 \cdot p_2 = p_2 \cdot p_1',$$

從て $$p_1 = p_1' \quad \cdots\cdots (A)$$

又兎に角 E_1 の起る確率は $$p_2 p_1' + (1-p_2)p_1''$$

である. 故に, $$p_1 = p_2 p_1' + (1-p_2)p_1'' \quad \cdots\cdots (B)$$

(A) を (B) に代入すれば

$$p_1 = p_1 p_2 + (1-p_2)p_1'' \quad \cdots\cdots (C)$$

又 $$p_1 = p_1 p_2 + p_1(1-p_2) \quad \cdots\cdots (D)$$

なるが故に, (C) と (D) とより $p_1 = p_1''$

即ち, $$p = p_1' = p_1''$$

故に E が E_1 に對して獨立なる時は, E_1 は亦 E_2 に對して獨立なのである. **互に獨立**なる語を用ひたる所以である.

更に,これを3個の事象に擴張する時は, E_2 は E_1 に對して獨立であり,又 E_3 が E_1 及 E_2 に對して獨立なる時には,この三つの事象は互に獨立である. n 個の事象に於ても同様である.

演習問題

1. 骰子を三回連續して投じ,其の内一回は1の目,他は偶數の目の出る確率をとふ.

解. 1の目の出る確率は $\frac{1}{6}$,偶數の目の出る確率は $\frac{1}{2}$ である.

(i) 第一回に1,第二及第三回に偶數の目の出る確率は獨立事象の定理により, $$p_1 = \frac{1}{6} \times \frac{1}{2} \times \frac{1}{2},$$

(ii) 第二回に1の目,他は偶數の出る確率は同様に, $$p_2 = \frac{1}{2} \times \frac{1}{6} \times \frac{1}{2},$$

(iii) 第三回に 1, 他は偶數の出る確率は, $p_3=\frac{1}{2}\times\frac{1}{2}\times\frac{1}{6}$.

而して (i) (ii) (iii) の三つの事象は互に排反であるから, この内何れか一つが起る確率は, 卽ち兎に角一回は 1 の目他は偶數の目の出る確率は

$$p=p_1+p_2+p_3=\frac{1}{8}$$

2. 袋中に白球 a 個, 赤球 b 個, 黑球 c 個あり. 今この袋中より順次に三回球を取出すに, (i) 三個とも赤球なる確率, (ii) 第一回白, 第二回赤, 第三回黑なる確率をとふ.

解. (i) 第一回に赤球を取出す確率 $p_1=\frac{b}{a+b+c}$,

第一回に赤球を取出したる時, 第二回にも赤球を取出す確率

$$p_2=\frac{b-1}{a+b+c-1},$$

第一及第二回に赤球を取出したりとし, 第三回にも赤球を取出す確率 $p_3=\frac{b-2}{a+b+c-2}$,

故に三回とも續けて赤球をとり出す確率は, 從屬事象の複確率の定理により,

$$p=p_1p_2p_3=\frac{b(b-1)(b-2)}{(a+b+c)(a+b+c-1)(a+b+c-2)}$$

(ii) 次に第一回に白球を取出す確率 $p_1=\frac{a}{a+b+c}$,

第一回に白球を取出したりとし, 第二回に赤球を取出す確率

$$p_2=\frac{b}{a+b+c-1},$$

第一回に白球, 第二回に赤球を取出したりとし, 第三回に黑球を取出す確率 $p_3=\frac{c}{a+b+c-2}$,

故に, 第一回に白, 第二回に赤, 且つ第三回に黑球を取出す確率は, 從屬事象の複確率の定理により,

$$p=p_1p_2p_3=\frac{abc}{(a+b+c)(a+b+c-1)(a+b+c-2)}$$

3. 三個の壺あり A なる壺には白球 10, 黑球 5 あり, B なる壺には白球 3, 黑球 4 あり, C なる壺には白球 5 あり, 今, 無心に一つの壺に手を入れて一球を取出すにその白球なる確率を問ふ. 但し三つの壺は手より等距離にあり, 形狀及び大さ同一であつて, 三つの壺の何れに手を入ることも同程度に確らしいとす.

解. 假定により A, B, C の三壺に手の入ることは同程度に確らしい, 故に, 各壺に手が入る確率は何れも $\frac{1}{3}$ である. さて A なる壺に手が入りたりとし, それより白球を取出す確率は $\frac{10}{15}$ である. 而して A なる壺より白球を取出すことは, A なる壺に手の入ることの從屬事象であるから, A に手を入れ且つこれより白球を取出す確率は,

$$p_1=\frac{1}{3}\cdot\frac{10}{15}=\frac{2}{9},$$

同樣に B なる壺に手が入り且つこれより白球を取出す確率は,

$$p_2=\frac{1}{3}\cdot\frac{3}{7}=\frac{1}{7},$$

又, C なる壺に手が入り白球を取り出す確率は $p_3=\frac{1}{3}\cdot 1=\frac{1}{3}$,

而して, A, B, C なる壺の何れかより白球を取出すことは排反事象であるから, この内より白球を取出す確率は, 全確率の定理により

$$p=\frac{2}{9}+\frac{1}{7}+\frac{1}{3}=\frac{44}{63},$$

である.

注意. この問題を次の如く解くは誤謬である.

A の壺より一白球を取出す確率: $\frac{10}{15}$,

B 壺より一白球を取出す確率: $\frac{3}{7}$,

C 壺より一白球を取出す確率: 1,

故に, 三つの壺の何れかより白球を取出す確率は,

$$p=\frac{10}{15}+\frac{3}{7}+1+=\frac{220}{105}$$

一見して其の誤りであることがわかるが, これは, 各壺から一白球を出すことは, 手が A, B, C の何れかの壺に入ると云ふことと, その壺の中から白球を取出すと云ふことの從屬事象であることを忘れたが爲に起る誤りである.

4. 五回連續して銅貨を投じ, 五回とも表の出づる確率をとふ.

第一回に銅貨を投じて表の出る確率は$\frac{1}{2}$,第二回以下でも亦同一である. 而して,各回に表の出ることは何れも互に獨立であるから,五回とも表の出る確率は,

$$p=\left(\frac{1}{2}\right)^5=\frac{1}{32}$$

5. 13本の福引中に當り籤3あり,最初引く人と第二番に引く人との損得如何.

一般に,籤は一番に抽いても二番三番に抽いても損得は無きさうに思はれるが,果して無いのであるか否かは精密な計算によらねばならぬ. 第一の人が籤を引く時,當り籤の確率は明に$\frac{3}{13}$である. 次に,第二の人が當り籤を抽く確率はやや複雜であつて,第一の人が當り籤を抽き且つ第二の人が當り籤を抽く確率は,從屬事象であるから,

$$p_2=\frac{3}{13}\cdot\frac{2}{12}=\frac{1}{26},$$

又,第一の人が空籤を抽いて第二の人が當り籤を抽く確率は,

$$p_2=\frac{10}{13}\cdot\frac{3}{12}=\frac{5}{26},$$

而して,第二の人が當り籤を抽く確率は,この二つの全確率で計算できる.

即ち,　$$p=p_1+p_2=\frac{1}{26}+\frac{5}{26}=\frac{3}{13}$$

即ち,第二の人が當り籤を抽く確率も$\frac{3}{13}$であつて,第一の人が當り籤を抽く確率と同一である. 故に,第一に抽く人と第二に抽く人とが損得がない. 同樣に第三の人にも損得がない. 以下皆同じ. されば,損得は籤を抽く順序には關しないのである.

6. 二個の骰子を投じ,二個とも1の目の出ざる確率は$1-\frac{1}{36}$即ち$\frac{35}{36}$である. 今,n回これをくり返して行ひ,n回とも二個の1を出さざる確率を求め,この確率が$\frac{1}{2}$より小なる爲には,nは如何なる價をとるべきか計算せよ.

二個とも1を出さざる確率は　$1-\frac{1}{36}=\frac{35}{36}$

さて,この實驗をn回くりかへし,n回とも二個の1を出さざる確率は,獨立事象の定理により,$\left(\frac{35}{36}\right)^n$である. 次にこれが$\frac{1}{2}$より小なのであるから,

$$\frac{1}{2}>\left(\frac{35}{36}\right)^n,$$

$$\log\frac{1}{2}>n\log\frac{35}{36},$$

$$-\log 2>n(\log 35-\log 36),$$

$$\frac{\log 2}{\log 36-\log 35}<n,$$

$$n>24{\cdot}6$$

即ち,nが24になる迄は$\left(\frac{35}{36}\right)^n$は$\frac{1}{2}$より大であるが,$n$が25以上になれば,初めて$\frac{1}{2}$より小となる.

7. 一人の女が四人の子供をうみ,(i) それが何れも男なる確率をとふ. 但し,男子出生の確率を0·51とし,且つ各回出產の男兒出生は互に獨立なるものとす.

(ii) 又末の子供のみ男なる確率をとふ.

(iii) 又最初の子供のみ男なる確率をとふ.

解. (i) 或る女が子供を生み,それが男なる確率は0·51である. 而して,假定により各回の男子出生は互に獨立であるから,四人とも男兒をうむ確率は　$p=0{\cdot}51^4=0{\cdot}068,$

(ii) 女兒を生む確率は0·49であるが故に,初め三人の女兒,最後に男兒をうむ確率は　$p=0{\cdot}49^3\times0{\cdot}51=0{\cdot}060,$

(iii) 最初のみ男子なる確率は(ii)と同じく

$$p=0{\cdot}51\times0{\cdot}49^3=0{\cdot}060,$$

各回の男兒出生が互に獨立なりと假定したのであるが,若し各回の出生に何等か關係あらば,上記は修正を要することになる. 又もし,統計の結果上記の通りにならなければ,二つの假定の何れかが誤りなのである. これを研究することは生物統計學上の面白き問題の一である.

第十章
屬性統計法[1]

第一項　二項分類

1. 屬性統計,屬性,二項分類

變數統計法は變異統計法の大部分を占むるものであるが,後者には尙ほ或性質又は狀態が存在するかしないかを,多數の個體につきて觀測する場合が殘つて居る．これを**屬性統計法**と云ふ．又,性質狀態を總稱して**屬性**と云ふ．屬性統計法では單に屬性を取扱ふが故に,一個體を觀測したのみでは,數値を以て表示することが出來ず,多數の觀測を行ふ時,初めて數として表示することが出來るのである．

さて,或屬性はこれを有するか有せざるか二階級に分類する時,この分類を**二項分類**(又は**二級分類**, division by dichotomy)と稱し,基礎的の分類法である．この項以下先づ二項分類につきて詳しく說く．

例へば豌豆の白花のものと赤花のものとの交配による雜種を數代培養するに,出て來る花は白か赤か何れかであつて,他の花色又は中間色が無い．此の時は白花と赤花との二項分類を

(1) この章は變異統計學の最初に置くべきかも知れないが,確率の定義や其の簡單な定理が用ひられて居るから,確率論の應用として此處で說くこととした．

行ひ,統計すればいい．高性豌豆と短矮性豌豆との交配によつて出て來るものも,亦,高性か然らざれば短矮性であるから,二項分類が成り立つ．しかし屬性をかく明確に二階級に分類し得ない場合がある,例へば皮膚の色の明暗の如きは無數の階段があつて,100人を觀察すれば100人その色調が異ふ．これなどは變數統計法が適するのであらう．けれども,或標準を設けて之を境とし,明と暗との二階級に分類すれば二項分類を行ふことが出來る．身長の如きも明に變數統計で取扱ふべき事項であるが,一定の標準を設けて,それより大なる身長と,それより小なる短身とに區別すれば,屬性統計的取扱が可能となる．又血液型には0型, A 型, B 型, AB 型なる四型があるから,多項分類が適するのであるが, 0型と非0型とに,或は A 型と非 A 型,又は $(A+AB)$ 型と其の他の型と云ふ樣に種々分類することが出來る．研究の目的次第で適宜に二項分類に更めうるのである．

二項分類法を一般に研究するには,屬性を文字を以てあらはす．例へば屬性を華文字で $[A]$, $[B]$, $\cdots\cdot[F]\cdots\cdot$ 等であらはす．$[A]$ は一つの屬性例へば豌豆の花色, $[B]$ は第二の屬性例へば種皮の性質, $[C]$ は莖の長さ等の如く定める．

2. 階級,級度數,級確率

二項分類であるから,屬性 $[A]$ は二階級に分たる．之を A, α とす．A とは $[A]$ なる屬性を有する階級[1], α は $[A]$ なる屬性を有せざる階級と定める．例へば $[A]$ は豌豆の花色, A は赤花

(1) 屢ば屬性 A, 屬性 B 等の語を用ふることがあるが,階級 A, B 等を意味する．

とすれば，α は非赤花なる階級をあらはす．又 $[B]$ は莖の長さなる屬性，B は長莖，β は短莖をあらはす．若し又 A が盲者の階級とすれば，α は明視者，A が男子出生の階級とすれば，α は女兒出生の階級となる．一般に，(A, α), (B, β), (C, γ), (D, δ), (E, ε), $\cdots\cdots$ 等を以て,互に排反なる二階級を代表せしむる．華文字と希臘文字とが,何れの階級を代表する樣に定めるかは,任意であるが,一研究に於ては定まつた表示法で終始すべき事勿論である．又遺傳學では,華文字を以て優性を,希臘文字の代りに，a, b 等小文字を以て劣性を代表せしむるのが普通である．

次に N 個の觀測に於て，A なる階級に屬する個體數 r，α なる階級に屬する個體數 $N-r=s$ 個なりとせよ．r は A なる階級の，$N-r$ は α なる階級の**級度數**と云ふ．

$A, B, C, \cdots\cdots \alpha, \beta, \gamma, \cdots\cdots$ なる階級の級度數を $(A), (B), (C), \cdots\cdots (\alpha), (\beta), \cdots\cdots$ 等を以て表はす．

又 $100\dfrac{(A)}{N}$, $1000\dfrac{(\alpha)}{N}$, $10000\dfrac{(B)}{N}$ 等の如く，級度數の 100 分比又は 1000 分比等を以て,級度數に代へ,或は，$\dfrac{(A)}{N}$, $\dfrac{(\alpha)}{N}$, $\dfrac{(B)}{N}$, 等の如く，經驗的の確率の近似値を以て,級度數に代ふることもある．

$(A)_{\%}$, $(B)_{\%}$, $(\alpha)_{\%}$, $\cdots\cdots$ 等を以てそれぞれの階級の**百分比度數**(**級百分比**)を表示し，$(A)_p$, $(B)_p$, $(\alpha)_p$, $\cdots\cdots$ 又は p_A, p_B, $\cdots\cdots$ 等を以てそれぞれの階級の**經驗的級確率の近似値**(略して**級確率**)を表示することと定める．

これらの記號を用ふる時は,次の等式が成り立つ．

$$\left.\begin{array}{l} 100(A)_p = (A)_{\%}, \\ 100(\alpha)_p = (\alpha)_{\%}, \\ \cdots\cdots\cdots\cdots\cdots\cdots \end{array}\right\} \cdots\cdots\cdots\cdots\cdots\cdots (\text{X. } 1)$$

$$\left.\begin{array}{l} N(A)_p = (A), \\ N(\alpha)_p = (\alpha), \\ \cdots\cdots\cdots\cdots \end{array}\right\} \cdots\cdots\cdots\cdots\cdots\cdots (\text{X. } 2)$$

而して,餘事象の全確率の定理により，

$$\left.\begin{array}{l} (A)_p + (\alpha)_p = 1, \\ (B)_p + (\beta)_p = 1, \end{array}\right\} \cdots\cdots\cdots\cdots\cdots\cdots (\text{X. } 3)$$

兩邊に觀測總數を乘ずることにより，

$$\left.\begin{array}{l} (A) + (\alpha) = N, \\ (B) + (\beta) = N, \end{array}\right\} \cdots\cdots\cdots\cdots\cdots\cdots (\text{X. } 3. \text{ b})$$

3. 階級の次數

或る一の屬性が有るか無いかを表示する階級，A, α, B, β, C, γ, $\cdots\cdots$ 等はこれを**一次の階級**と云ふ．即ち一次の階級とは唯一個の文字で表示される階級を云ふ．一次階級の度數 (A), (α), (B), (β), (C), $\cdots\cdots$ 等を**一次の級度數**と云ふ．又,一次階級の級確率，$(A)_p$, $(\alpha)_p$, $(B)_p$, $\cdots\cdots$ 等を一次の級確率と云ふ．二個の屬性を有するか有しないかを表示する階級を**二次の階級**と云ふ．即ち二次の階級は二個の文字を以て表示せらるる階級のことである．例へば AB は屬性 A を有し且つ屬性 B を有する階級のことであり，$A\beta$ は屬性 A を有するが屬性 B を有せざる階級のこと，$\alpha\beta$ は屬性 A も B も有せざる階級のことで,皆二次の階級である．AD, $A\delta$, αD, $\alpha\delta$, AC, αC $\cdots\cdots$ 等も亦二次の階級である．二次の階級の度數を**二次の級度數**,二次の階級の級確率を**二次の級確率**と云ふ．

(AB), $(A\beta)$, (αB), $(\alpha\beta)$, $\cdots\cdots$ (BC), $(B\gamma)$, (βC), $(\beta\gamma)$, $\cdots\cdots$ 等は何

れも二次の級度數,又 $(AB)_p$, $(\alpha B)_p$, $(A\beta)_p$, $(\alpha\beta)_p$, $\cdots\cdots$ 等は何れも二次の級確率である. 三つの屬性 A, B, C を有するか,有せざるかを表示する階級は,之を**三次の階級**と云ひ,其の度數を**三次の級度數**,其の確率を**三次の級確率**と云ふ.

ABC, $A\beta C$, $A\beta\gamma$, $\cdots\cdots$ 等は三次の階級であり, (ABC), $(A\beta C)$, $(A\beta\gamma)$, $(\alpha\beta\gamma)$, $\cdots\cdots$ 等は三次の級度數, $(ABC)_p$, $(A\beta C)_p$, $(A\beta\gamma)_p$, $\cdots\cdots$ 等は三次の級確率である.

一般に k 個の屬性よりなる階級を **k 次の階級**,その度數を **k 次の級度數**その確率を **k 次の級確率**と云ふ.

s 個の屬性を取扱ふ際, s 次の階級を**最終の階級**,その度數を**最終(級)度數**,その確率を**最終(級)確率**と云ふ. 又 **0 次の階級**とは觀測の總ての範圍であつて, **0 次の級度數**は**觀測總數**,又 **0 次の確率**は 1 である.

二次の級確率及級度數につきては,次の式が成立する[1].

$$\left.\begin{aligned}
(A)_p &= (AB)_p + (A\beta)_p = (AC)_p + (A\gamma)_p \\
&= (AD)_p + (A\delta)_p = \cdots\cdots, \\
(\alpha)_p &= (\alpha B)_p + (\alpha\beta)_p = (\alpha C)_p + (\alpha\gamma)_p = \cdots\cdots, \\
1 &= (AB)_p + (A\beta)_p + (\alpha B)_p + (\alpha\beta)_p \\
&= (AC)_p + (A\gamma)_p + (\alpha C)_p + (\alpha\gamma)_p = \cdots\cdots,
\end{aligned}\right\} \cdots\cdots\cdots \text{(X. 4. a)}$$

兩邊に N を乘じて,

$$\left.\begin{aligned}
(A) &= (AB) + (A\beta) = (AC) + (A\gamma) = (AD) + (A\delta) = \cdots\cdots, \\
(\alpha) &= (\alpha B) + (\alpha\beta) = (\alpha C) + (\alpha\gamma) = (\alpha D) + (\alpha\delta) = \cdots\cdots, \\
N &= (AB) + (A\beta) + (\alpha B) + (\alpha\beta) \\
&= (AC) + (A\gamma) + (\alpha C) + (\alpha\gamma) = \cdots\cdots\cdots\cdots\cdots\cdots,
\end{aligned}\right\} \text{(X. 4. b)}$$

[1] AB と $A\beta$, AC と $A\gamma$ 等は排反事象なることに着眼すれば是等の式は容易に理解せらる.

又三次の級確率及級度數については,次の諸式が成立する.

$$\left.\begin{aligned}
(AB)_p &= (ABC)_p + (AB\gamma)_p = (ABD)_p + (AB\delta)_p = \cdots\cdots, \\
(A\beta)_p &= (A\beta C)_p + (A\beta\gamma)_p = (A\beta D)_p + (A\beta\delta)_p = \cdots\cdots, \\
(A)_p &= (ABC)_p + (AB\gamma)_p + (A\beta C)_p + (A\beta\gamma)_p = \cdots\cdots, \\
(\alpha)_p &= (\alpha BC)_p + (\alpha B\gamma)_p + (\alpha\beta C)_p + (\alpha\beta\gamma)_p = \cdots\cdots, \\
1 &= (ABC)_p + (AB\gamma)_p + (A\beta C)_p + (A\beta\gamma)_p \\
&\quad + (\alpha BC)_p + (\alpha B\gamma)_p + (\alpha\beta C)_p + (\alpha\beta\gamma)_p = \cdots\cdots,
\end{aligned}\right\} \text{(X. 5. a)}$$

兩邊に N を乘じ,

$$\left.\begin{aligned}
(AB) &= (ABC) + (AB\gamma) = (ABD) + (AB\delta) = \cdots\cdots\cdots, \\
(A\beta) &= (A\beta C) + (A\beta\gamma) = (A\beta D) + (A\beta\delta) = \cdots\cdots\cdots, \\
(A) &= (ABC) + (AB\gamma) + (A\beta C) + (A\beta\gamma) = \cdots\cdots\cdots, \\
(\alpha) &= (\alpha BC) + (\alpha B\gamma) + (\alpha\beta C) + (\alpha\beta\gamma) = \cdots\cdots\cdots, \\
N &= (ABC) + (AB\gamma) + (A\beta C) + (A\beta\gamma) \\
&\quad + (\alpha BC) + (\alpha B\gamma) + (\alpha\beta C) + (\alpha\beta\gamma) = \cdots\cdots\cdots,
\end{aligned}\right\} \text{(X. 5. b)}$$

三次以上の場合も同樣に等式を作ることが出來る.

4. 階級の數

(i) s 個の屬性がある時,其の一次の階級の總數は $2s$ 個あり.

(ii) 又 s 個の屬性がある時,二次の階級の總數は $2s(s-1)$ 個あり. 何となれば,花文字のみから成る階級の數は, s 個の中から二個づつ取りたる組合せの數であるから $\dfrac{s(s-1)}{2!}$ 個あり; 希臘文字のみからなる階級數も,同じく $\dfrac{s(s-1)}{2!}$ 個あり; 第一の文字が花文字で第二が希臘文字なる階級も,第二が花文字で第一が希臘文字なる級階數も,共に $\dfrac{s(s-1)}{2!}$ である. 故に

二次の階級總數は,

$$4\times\frac{s(s-1)}{2!}=2s(s-1)$$

個あり.

(iii)　又三次の階級の總數は $\frac{4}{3}s(s-1)(s-2)$ あり.

(iv)　一般に, k 次の階級の總數は $2^k\times\frac{s(s-1)\cdots\cdot(s-k+1)}{k!}$ 個あり.

証明. 全部が花文字のみよりなる組合せの數は, $\frac{s(s-1)\cdots\cdot(s-k+1)}{k!}$ 通りあり. 最初の文字のみが希臘文字で他は華文字なる組合せの數も,同數だけあり. 第二番目のもののみが希臘文字なる組合せも同數だけあり,第三番以下亦同様である. 故に, 1 個が希臘文字なる組合せの數は,

$$k\frac{s(s-1)\cdots\cdot(s-k+1)}{k!},$$

次に,最初の二個が希臘文字よりなる組合せの數は,

$$\frac{s(s-1)(s-2)\cdots\cdot(s-k+1)}{k!}\quad あり,$$

故に,二個だけ希臘文字なる組合せの數は,

$${}_kC_2\frac{s(s-1)\cdots\cdot(s-k+1)}{k!}\quad 個あり.$$

…………………………………

而して全部が華文字なる組合せの數,一個が希臘文字なる組合せの數, …… の總和は

$$\frac{s(s-1)(s-2)\cdots\cdot(s-k+1)}{k!}\{1+{}_kC_1+{}_kC_2+{}_kC_3+\cdots\cdot+{}_kC_k\}$$

となる.

而して,此式の括弧の中は,

$$1+k+\frac{k(k+1)}{2!}+\frac{k(k-1)(k-2)}{3!}+\cdots\cdot+k+1=(1+1)^k$$

の左邊に一致するが故に, 2^k となる. されば k 次の階級の數の總和は

$$2^k\frac{s(s-1)(s-2)\cdots\cdot(s-k+1)}{k!}\qquad\cdots\cdots\cdots\cdots(X.\ 6)$$

通りあり.

(v)　最終度數は 2^s 個あり. 何となれば, $k=s$ なる時上式は 2^s となるからである.

5. 最終級度數と積極級度數

かく計算すると, s 個の屬性が有る時階級數は甚しく多數となるが,これらの級度數が悉く必要なわけではない. **最終度數**(又は**最終確率**)を悉く記録するか,或は**積極度數**(又は**積極確率**)を悉く記録するかで充分である.

最終度數(または最終確率)が記録されてあれば,必要な場合には,公式 (5. a) 又は (5. b) により,また,これを應用して他の級度數を計算することが出來るからである. しかし常に 2^s 個の級度數を記録せねばならぬ. 但し,若し最終級確率(または百分比等)を示さんとならば, 2^s 通りの外に觀測總數 N をも記録すべきである.

最終度數の代りに,積極度數が記録せられる事がある. $A, B, C, \cdots\cdot AB, AC, AD, BC, \cdots\cdot ABC, ABD, \cdots\cdot$ 等の如く希臘文字を含まざる階級を**積極階級**と云ひ,其の級度數を**積極級度數**と云ふ. $\alpha, \beta, \gamma, \cdots\cdot \alpha\beta, \beta\gamma, \cdots\cdot \alpha\beta\gamma, \cdots\cdot$ 等の如く,花文字を含まざる階級を**消極階級**,其の度數を**消極度數**と云ふ. 今,屬性が四個ある時,積極度數に次のものがある.

0 次の積極級度數: N,

1 次の積極級度數: (A), (B), (C), (D),

2 次の積極級度數: (AB), (AC), (AD), (BC), (BD), (CD),

3 次の積極級度數: (ABC), (ABD), (ACD), (BCD),

4 次の積極級度數: $(ABCD)$,

故に四個の屬性の時は,積極級度數は

$$1+4+6+4+1=16=2^4$$

あり. 一般に s 個の屬性がある時,極積級度數の數は

$$\begin{aligned}&1+{}_sC_1+{}_sC_2+{}_sC_3+\cdots\cdot+{}_sC_k\cdots\cdot+{}_sC_{s-1}+{}_sC_s\\&=1+s+\frac{s(s-1)}{2!}+\frac{s(s-1)(s-2)}{3!}+\cdots\cdot\\&\quad+\frac{s(s-1)\cdots\cdot(s-k+1)}{k!}+\cdots\cdot s+1=2^s\end{aligned}$$

個あり.

即ち,最終度數と同じく,積極級度數も 2^s 個あり. さて,總ての積極級度數が記錄せられてあれば,他の級度數も,亦,公式 (X.3), (X.4), (X.5) 等を應用して,容易に計算することが出來る. 例へば,三個の屬性を取扱ふ時,積極級度數のみが與へられて居るならば,他の級度數は次の如く計算しうる.

$$\begin{aligned}(\alpha)&=N-(A),\\(\beta)&=N-(B),\\(\gamma)&=N-(C),\\(A\beta)&=(A)-(AB),\\(\alpha B)&=(B)-(AB),\\(A\gamma)&=(A)-(AC),\\(\alpha C)&=(C)-(AC),\end{aligned}$$

$$\left.\begin{aligned}(\beta C)&=(C)-(BC),\\(B\gamma)&=(B)-(BC),\\(\alpha\beta)&=(\alpha)-(\alpha B)=N-(A)-(B)+(AB),\\(\alpha\gamma)&=(\alpha)-(\alpha C)=N-(A)-(C)+(AC),\\(\beta\gamma)&=(\beta)-(\beta C)=N-(B)-(C)+(BC),\\(AB\gamma)&=(AB)-(ABC),\\(A\beta C)&=(AC)-(ABC),\\(\alpha BC)&=(BC)-(ABC),\\(A\beta\gamma)&=(A\beta)-(A\beta C)\\&=(A)-(AB)-(AC)+(ABC),\\(\alpha B\gamma)&=(B\gamma)-(AB\gamma)\\&=(B)-(BC)-(AB)+(ABC),\\(\alpha\beta C)&=(\alpha C)-(\alpha BC)\\&=(C)-(AC)-(BC)+(ABC),\\(\alpha\beta\gamma)&=(\alpha\beta)-(\alpha\beta C)=(\alpha)-(\alpha B)-(\alpha C)+(\alpha BC)\\&=N-(A)-(B)-(C)+(AB)+(AC)+(BC)\\&\quad-(ABC),\end{aligned}\right\}\quad\text{(X. 7)}$$

要するに,s 個の屬性を取扱つた時,記錄すべき度數はすべての最終度數か,若くはすべての積極級度數かである. 何れにしても記錄すべき階級數は 2^s 個あり.

最終度數を記錄することと,積極級度數を記錄することと,何れを選んでもいいが,初學者には最終度數を記錄した方が誤りがない.

第二項 成　　立

6. 成立

若し,最終度數と觀測總數とが與へられてある時,其の觀測に過失が有るか無いかを數字の上から發見するには,最終級度數の總和が觀測總數に一致するか否かを見る. 例へば三つの屬性を考へたる時は,

$$N=(ABC)+(AB\gamma)+(A\beta C)+(A\beta\gamma)$$
$$+(\alpha BC)+(\alpha B\gamma)+\alpha\beta C)+(\alpha\beta\gamma),$$

若し,此の等式が成立しない時は,觀測も亦成立しないのであつて,何處かに過失がある筈である. 若し,最終級度數以外に他の級度數も與へられて居る時は,公式(X. 4)又は(X. 5)等によりてこれを計算し,その一致するか否かを見る事が出來る. 積極度數のみが與へられて居る時は,式(X. 7)によりて最終級度數を計算せよ. 若し最終級度數が一つでも負數となることがあれば,この觀測は成立しない. 過失には單に誤植,誤寫,誤算の外に記錄又は觀測の誤り,時又は處をことにせる觀測を混同し,又は條件を異にせる觀測を混同する等の種々の過失が有り得る.

屬性二個なる時は,最終級度數は

$$\left.\begin{aligned}(A\beta)&=(A)-(AB),\\(\alpha B)&=(B)-(AB),\\(\alpha\beta)&=N-(A)-(B)+(AB),\end{aligned}\right\}\cdots\cdots(\text{X. 7. a})$$

であつて,これらの級度數は何れも0より大ならねばならぬから,成立の條件としては

$$\left.\begin{aligned}(A)-(AB)&\geqq 0,\\(B)-(AB)&\geqq 0,\\N-(A)-(B)+(AB)&\geqq 0,\end{aligned}\right\}\cdots\cdots\cdots\cdots(\text{X. 8. a})$$

或は

$$\left.\begin{aligned}(AB)&\geqq 0,\\(A)&\geqq(AB),\\(B)&\geqq(AB),\\N&\geqq(A)+(B)-(AB),\end{aligned}\right\}\cdots\cdots\cdots\cdots(\text{X. 8. b})$$

屬性三個なる時は,(X. 7)式を用ひて計算せる最終度數何れも正なるが爲に,次の不等式が成立する.

$$\left.\begin{aligned}&(ABC)\geqq 0,\\&(AB)\geqq(ABC),\\&(AC)\geqq(ABC),\\&(BC)\geqq(ABC),\\&(A)+(ABC)\geqq(AB)+(AC),\\&(B)+(ABC)\geqq(BA)+(BC),\\&(C)+(ABC)\geqq(CA)+(CB),\\&N+(AB)+(AC)+(BC)\\&\quad\geqq(A)+(B)+(C)+(ABC),\end{aligned}\right\}\cdots\cdots\cdots(\text{X. 9})$$

若し,積極級度數以外の級度數も與へられて居れば,積極級度數より其の級度數を計算し,一致するか否かを見よ. 一致せなければやはり不成立である.

最後に觀測が成立するとは,與へられた度數表に矛盾のないことを意味するのであつて,誤り又は過失が絕對に無いと云ふのではない. 過失が他に有るかも知れぬが,度數表だけでは其

の誤りを指摘することが出來ない. 斯の如き場合も觀測は成立すると云ふのである.

例 1. 次の報告が實際觀測の結果としての積極度數でありとすれば,何處かに誤りが有る筈である. 若し,それが誤植とすれば如何なる所に誤植が有るか.これを發見せよ.

$$N=10{,}000,\qquad (AB)=338,$$
$$(A)=877,\qquad (AC)=243,$$
$$(B)=1{,}086,\qquad (BC)=135,$$
$$(C)=286,\qquad (ABC)=57.$$

この數を見ただけでは,何の感銘も起らぬが,これは成立しない. これより最終度數を計算すれば次の如くなる.

$$(AB\gamma)=(AB)-(ABC)\qquad =338-57=281,$$
$$(A\beta C)=(AC)-(ABC)\qquad =243-57=186,$$
$$(\alpha BC)=(BC)-(ABC)\qquad =135-57=78,$$
$$(A\beta\gamma)=(A\beta)-(A\beta C)=(A)-(AB)-(AC)+(ABC)$$
$$=877-338-243+57=353,$$
$$(\alpha B\gamma)=(\alpha B)-(\alpha BC)=(B)-(AB)-(BC)+(ABC)$$
$$=1{,}086-338-135+57=570,$$
$$(\alpha\beta C)=(\alpha C)-(\alpha BC)=(C)-(AC)-(BC)+(ABC)$$
$$=286-243-135+57=-35,$$
$$(\alpha\beta\gamma)=N-(A)-(B)-(C)+(AB)+(AC)+(BC)-(ABC)$$
$$=10{,}000-877-1{,}086-286+383+243+135-57=8410,$$

$(\alpha\beta C)$ が負となるから, (C) 或は (ABC) を過少に記錄してあるか,又は, (AC) 若くは (BC) を過大に記錄してあるか,何れかでな

ければならぬ.

四個以上の屬性を取扱ふ場合も,亦同樣に計算することが出來る. もし, N, (A), (B), (C), (D), (AB), (AC), (AD), (BC), (BD), (CD), (ABC), (ABD), (BCD) 及び $(ABCD)$ が與へられて居るときは,

$$\left.\begin{aligned}
&(ABCD)\geqq 0\\
&(ABC)\geqq(ABCD),\\
&(ABD)\geqq(ABCD).\\
&(ACD)\geqq(ABCD),\\
&(BCD)\geqq(ABCD),\\
&(AB)+(ABCD)\geqq(ABC)+(ABD),\\
&(BC)+(ABCD)\geqq(ABC)+(BCD),\\
&(CD)+(ABCD)\geqq(CDA)+(CDB),\\
&(DA)+(ABCD)\geqq(DAB)+(DAC),\\
&(AC)+(ABCD)\geqq(ACB)+(ACD),\\
&(BD)+(ABCD)\geqq(BDA)+(BDC),\\
&(A)+(ABC)+(ABD)+(ACD)\\
&\qquad\geqq(AB)+(AC)+(AD)+(ABCD),\\
&(B)+(BCD)+(BCA)+(BDA)\\
&\qquad\geqq(BC)+(BD)+(BA)+(ABCD),\\
&(C)+(CDA)+(CDB)+(CAB)\\
&\qquad\geqq(CD)+(CA)+(CB)+(ABCD),\\
&(D)+(DAB)+(DAC)+(DCB)\\
&\qquad\geqq(DA)+(DB)+(DC)+(ABCD),\\
&N+(AB)+(AC)+(AD)+(BC)+(BD)\\
&\quad+(CD)+(ABCD)\geqq(A)+(B)+(C)\\
&\quad+(D)+(ABC)+(ABD)+(ACD)+(BCD),
\end{aligned}\right\}\cdots\text{(X. 10)}$$

これらの16個の式は成立の條件であつて,1個でも缺ける時は不成立となる.

第三項　關　聯

7. 關聯表

變數統計で二つの變數を考へた場合,この變數間に相關が有るか無いか,又有りとすれば其の程度如何を研究した. 屬性統計では,二つの屬性に**關聯** (Association) が有るか無いか,有りとすれば其の程度如何が問題となる.

關聯を研究するには第一に**關聯表** (Associationtable) を作るがいい. 第69表の甲及び乙は其の一般表であるが,何れか一方を記入するだけで足りる.

第 69 表

關　聯　表

甲

(度數は絕對數を記入してある)

[A] \ [B]	A	α	計
B	(AB)	(αB)	(B)
β	$(A\beta)$	$(\alpha\beta)$	(β)
計	(A)	(α)	N

乙

(度數は確率を記入してある $N=\cdots\cdots$)

階數	A	α	計
B	$(AB)_p$	$(\alpha B)_p$	$(B)_p$
β	$(A\beta)_p$	$(\alpha\beta)_p$	$(\beta)_p$
計	$(A)_p$	$(\alpha)_p$	1

甲は度數分布表であり,乙は確率の分布表である. 確率の分布表では觀測總數を記入することを忘れてはならぬ. 乙に代ふるに100分比分布表, 1000分比分布表を以て代ふるも,同一の注意を要する. 表甲は觀測の基本的の表であるが,乙は他の觀

測と比較する時に,屢,用ひられる.

例 2. 次表は種痘と,天然痘に罹れるものとの關聯表である.

$[A]$ を罹病, A は罹病せるもの, α は罹病せざるもの,

$[B]$ を種痘, B は種痘せるもの, β は種痘せざるもの

とし關聯表を作れば次の二つの表が得られる.

第 70 表

甲

[A] \ [B]	罹病せるもの A	罹病せざるもの α	計
種痘せるもの B	3	276	279
種痘せざるもの β	66	473	539
計	69	749	818

乙

[A] \ [B]	A	α	$A+\alpha$
B	0·004	0·337	0·341
β	0·081	0·578	0·659
$B+\beta$	0·085	0·915	1·000

しかし,關聯表の代りに他の表を以てすることも出來る.

(i) 最終度數のみで表示すれば,

$$(AB)=3, \qquad (A\beta)=66,$$
$$(\alpha B)=276, \qquad (\alpha\beta)=473,$$

(ii) 積極度數のみを以て表示すれば,

$$N=818, \qquad (B)=279,$$
$$(A)=69, \qquad (AB)=3,$$

これらの表の何れかより關聯表を作ることも容易である.

8. 無關聯

Aの確率とBの確率との複確率がABの確率に等しい時,即ち,

$$(AB)_p=(A)_p(B)_p$$

なる時は, $[A]$ なる屬性(事象)と $[B]$ なる屬性(事象)とは互に無關聯なり,又互に獨立なりと云ふ[1].

上式が成立する時は,

$$\begin{aligned}(A)_p-(AB)_p&=(A)_p-(A)_p(B)_p\\&=(A)_p\{1-(B)_p\}\\&=(A)_p(\beta)_p\end{aligned}$$

故に　$(A\beta)_p=(A)_p(\beta)_p,$

一般に次の四式の內,何れか一つが成立する時は,他の三つは成立するから,何れの一式も無關聯の條件となりうるのである.

$$\left.\begin{aligned}(AB)_p&=(A)_p(B)_p,\\(A\beta)_p&=(A)_p(\beta)_p,\\(\alpha B)_p&=(\alpha)_p(B)_p,\\(\alpha\beta)_p&=(\alpha)_p(\beta)_p,\end{aligned}\right\}\cdots\cdots\cdots\cdots(\text{X. 11. a})$$

而して無關聯を證明するには,四式中何れか一つで充分である. (X. 11. a) 式の兩邊に N を乘ずることによりて,(右邊には $\frac{N^2}{N}$ を乘ず),

$$\left.\begin{aligned}(AB)&=\frac{(A)(B)}{N},\\(A\beta)&=\frac{(A)(\beta)}{N},\\(\alpha B)&=\frac{(\alpha)(B)}{N},\\(\alpha\beta)&=\frac{(\alpha)(\beta)}{N},\end{aligned}\right\}\cdots\cdots\cdots\cdots(\text{X. 11. b})$$

となる. 又この式を變化する時は,

[1] 獨立事象に於ける複確率の定理の逆である.

$$\left.\begin{aligned}\frac{(A)}{N}&=\frac{(AB)}{(B)}=\frac{(A\beta)}{(\beta)},\\\frac{(B)}{N}&=\frac{(AB)}{(A)}=\frac{(\alpha B)}{(\alpha)},\\\frac{(\alpha)}{N}&=\frac{(\alpha B)}{(B)}=\frac{(\alpha\beta)}{(\beta)},\\\frac{(\beta)}{N}&=\frac{(A\beta)}{(A)}=\frac{(\alpha\beta)}{(\alpha)},\end{aligned}\right\}\cdots\cdots\cdots\cdots\cdots\cdots(\text{X. 11. c})$$

又, (X. 11. a) 式より,

$$(AB)_p(\alpha\beta)_p=(A)_p(\alpha)_p(B)_p(\beta)_p,$$

$$(\alpha B)_p(A\beta)_p=(A)_p(\alpha)_p(B)_p(\beta)_p,$$

故に,

$$\left.\begin{aligned}(AB)_p(\alpha\beta)_p&=(\alpha B)_p(A\beta)_p,\\\frac{(AB)_p}{(\alpha B)_p}&=\frac{(A\beta)_p}{(\alpha\beta)_p},\\\frac{(AB)_p}{(A\beta)_p}&=\frac{(\alpha B)_p}{(\alpha\beta)_p},\end{aligned}\right\}\cdots\cdots\cdots(\text{X. 11. d})$$

又,これより

$$\left.\begin{aligned}(AB)(\alpha\beta)&=(A\beta)(\alpha B),\\\frac{(AB)}{(\alpha B)}&=\frac{(A\beta)}{(\alpha\beta)},\\\frac{(AB)}{(A\beta)}&=\frac{(\alpha B)}{(\alpha\beta)},\end{aligned}\right\}\cdots\cdots\cdots\cdots(\text{X. 11. e})$$

$[A]$ と $[B]$ とが互に獨立なる時は,これらの諸式が成立し,又,若しこれらの式の內一つが成立する時は, $[A]$ と $[B]$ とは互に獨立である. 換言すれば, (X. 11. a), (X. 11. b), (X. 11. c), (X. 11. d), (X. 11. e) の中,何れの式も獨立の條件となりうるのである. このうち比較的屢用ひらるる式:

$$\left.\begin{array}{l} (AB) = \dfrac{(A)(B)}{N} \\[2ex] \dfrac{(AB)}{(B)} = \dfrac{(A\beta)}{(\beta)} \\[2ex] \dfrac{(AB)}{(A)} = \dfrac{(\alpha B)}{(\alpha)} \end{array}\right\} \cdots\cdots\cdots\cdots \text{(X. 11)}$$

を記憶すれば充分である. $\dfrac{(AB)}{(B)} = \dfrac{(A\beta)}{(\beta)}$ は $[B]$ を有する度數と,其の階級內で更に A をも有するものの度數との比は,(B) を有せざる度數と,其の階級內で更に A を有するものとの比に等しきことを意味する.

例 3. 甲,乙二個の骰子を同時に投じ,甲に奇數が出た時をA,偶數が出た時を α, 又乙に奇數の出た時を B, 偶數の出た時を βとす. 今1000投の實驗に於て次表の如き結果を得た. A と B との間に關聯ありや.

$$(AB) = 238, \qquad (\alpha B) = 287,$$
$$(A\beta) = 242, \qquad (\alpha\beta) = 233,$$

解. これより關聯表を作る.

第 71 表

[B] \ [A]	A	α	$A+\alpha$
B	238	287	525
β	242	233	475
$B+\beta$	480	520	1000

$$\frac{(A)(B)}{N} = 252,$$
$$\Delta = (AB) - \frac{(A)(B)}{N} = -14,$$

Δ は負であるから,負の關聯が有る樣に見えるけれど,Δ が (AB) に比し小であるから,斯の如き動搖は偶然の爲に常に起るもので,$\Delta = 0$ と假定して差支ないのである.

又 $\dfrac{(AB)}{(A)} = 0{\cdot}496, \qquad \dfrac{(\alpha B)}{(\alpha)} = 0{\cdot}552,$

$$\frac{(AB)}{(B)} = 0{\cdot}453, \qquad \frac{(A\beta)}{(\beta)} = 0{\cdot}461,$$

卽ち, $\dfrac{(AB)}{(A)} \fallingdotseq \dfrac{(\alpha B)}{(\alpha)}, \qquad \dfrac{(AB)}{(B)} \fallingdotseq \dfrac{(A\beta)}{(\beta)},$

卽ち關聯が無いと云ふことが出來る. 但し關聯の有無を精確に檢討するのは以上の等式のみでは不充分であつて關聯係數と其の誤差を計算し(第14章)或は χ^2-試驗によらねばならぬ.

例 4. トウモロコシの種子に,紫色のものと,白色のものとあり. 紫色が優性であるから,之を表はすにW,白色をあらはすに w を以てす. 又,種子に澱粉性のものと,砂糖性のものとが有り. 前者が優性であるから,之を表はすにS, 砂糖性のものを表はすに s を以てす.

この二つの屬性につきて雜種をつくれば WS のみが出來るが,其の自家生殖にて第二代目に生じたる種子は,次の四種あつた.

$$(WS) = 1816, \qquad (Ws) = 614,$$
$$(wS) = 548, \qquad (ws) = 217,$$
$$N = 3195.$$

W と S とに關聯がありや.

解. $\dfrac{(WS)}{(W)} = \dfrac{1816}{2430} = 0{\cdot}747, \qquad \dfrac{(wS)}{(w)} = \dfrac{548}{765} = 0{\cdot}716,$

$$\frac{(WS)}{(S)} = \frac{1816}{2364} = 0{\cdot}768, \qquad \frac{(Ws)}{(s)} = \frac{614}{831} = 0{\cdot}739,$$

$$\frac{(WS)}{(W)} \fallingdotseq \frac{(wS)}{(w)} \qquad \frac{(WS)}{(S)} \fallingdotseq \frac{(Ws)}{(s)}$$

となるから,先づ關聯が無いと云へよう. 併し,正しくは關聯係數と其の誤差を計算するか,或はχ^2-試驗の法によるべきである.

9. 關聯の正負

$$(AB)_p=(A)_p(B)_p$$

である時 $[A]$ と $[B]$ とは無關聯であつた. 然らば,

$$(AB)_p \gtrless (A)_p(B)_p$$

である時は $[A]$ と $[B]$ とは**關聯がある**,卽ち獨立でないと云ふ. 若し, $(AB)_p>(A)_p(B)_p$ である時は $[A]$ と $[B]$ とは**正に關聯**して居ると云ふ. (positively associated). 又 $(AB)_p<(A)_p(B)_p$ なる時は, $[A]$ と $[B]$ とは**負に關聯**せりと云ふ. (negatively associated). この二式より (X. 11. a) (X. 11 b) (X. 11. c) (X. 11. d) (X. 11. e) にならひ多數の不等式を誘導することが出來るが,不必要であるから省略する. 然らば,これらの不等式中で關聯の正負を討檢するに,何の式が最適切であるかの問題が起る. これには,

(i) 第一に不等式が攻究問題を直接に說明して居ることが必要である. 例へば

$$\frac{(\alpha B)}{(\alpha)}<\frac{(B)}{N}$$

の如きは正の關聯を示す式ではあるが攻究の目的に對して直接性に乏しい. A, B 等の花文字は攻究の目的になるものを直接に指示して居るから,上式にかふるに,

$$\left.\begin{array}{ll}\dfrac{(AB)}{(B)}>\dfrac{(A)}{N}, & \dfrac{(AB)}{(A)}>\dfrac{(B)}{N}, \cdots\cdots \\[2ex] \dfrac{(AB)}{(B)}>\dfrac{(A\beta)}{(\beta)}, & \dfrac{(AB)}{(A)}>\dfrac{(\alpha B)}{(\alpha)} \cdots\cdots\end{array}\right\} \cdots\cdots(\text{X. }12)$$

等の式を用ふべきである.

(ii) 又關聯の度を最も明白に決定する式が一番いい. 故に,四個の上式中,第三第四が主として用ひられる. 何となれば,例へば (B) 又は (A) の大なる時は $\frac{(AB)}{(B)}$ と $\frac{(A)}{N}$ と,或は $\frac{(AB)}{(A)}$ と $\frac{(B)}{N}$ との差は甚しく小となり,一見關聯が無きかの如き觀を呈するが,反之

$$\left.\begin{array}{ll} & \dfrac{(AB)}{(B)}>\dfrac{(A\beta)}{(\beta)}, \\[2ex] \text{或は} & \dfrac{(AB)}{(A)}>\dfrac{(\alpha B)}{(\alpha)}\end{array}\right\} \cdots\cdots\cdots\cdots(\text{X. }12)$$

は,比較的正しい觀念を與へるからである.

又,負に關聯せることの條件としては,

$$\left.\begin{array}{l}\dfrac{(AB)}{(B)}<\dfrac{(A\beta)}{(\beta)}, \\[2ex] \dfrac{(AB)}{(A)}<\dfrac{(\alpha B)}{(\alpha)},\end{array}\right\} \cdots\cdots\cdots\cdots(\text{X. }13)$$

を用ふる.

例 5. 大正五年現在の日本人口,男 27,837,000; 女 27,398,000; 計 55,235,000. 又同年の死亡數,男 604,156; 女 583,674; 計 1,187,832. 今 A を男, B を死亡とし, A と B との關聯(男たることと死亡との關係)を求む.

解.

$$\frac{(AB)}{(A)}=\frac{604{,}156}{27{,}837{,}000}=0{\cdot}0217,$$

$$\frac{(\alpha B)}{(\alpha)}=\frac{583{,}674}{27{,}398{,}000}=0{\cdot}0213,$$

卽ち $\frac{(AB)}{(A)}>\frac{(\alpha B)}{(\alpha)}$ であるから,男子と死亡との間に正の關聯がある.

他の言葉で云へば男子の死亡率, 21·7‰

女子の死亡率, 21·3‰

であるから,男子の死亡率は女子のそれより大であると云ふことに歸する. この問題を次の様に計算できないことはない.

$\frac{(AB)}{(A)}$ を $\frac{(\alpha B)}{(\alpha)}$ と比較する代りに $\frac{(B)}{N}$ と比較しても差支ない.

$$\frac{(AB)}{(A)}=0{\cdot}0217,$$

$$\frac{(B)}{N}=0{\cdot}0215,$$

即ち $\frac{(AB)}{(A)}>\frac{(B)}{N}$, 故に正の關聯がある.

けれども,この二數 0·0217 と 0·0215 との間には差違は小であるから,本文で說いた樣に,男子の死亡率と女子の死亡率を比較すべきである. 又 $\frac{(AB)}{(B)}$ と $\frac{(A\beta)}{(\beta)}$ とを比較してもいい.

其の年に死亡せる者のうち男の割合 $\frac{(AB)}{(B)}=0{\cdot}509,$

其の年に死亡せざりし者のうち男の割合 $\frac{(A\beta)}{(\beta)}=0{\cdot}504,$

$$\frac{(AB)}{(B)}>\frac{(A\beta)}{(\beta)}$$

であるから A と B との間に正の關聯がある.

例 6. 本章例 2 に於ける種痘 B と,天然痘罹病 A との間には,負の關聯あることを示せ.

解. $(AB)=3,\quad (A)=69,$

$(A\beta)=66,\quad (B)=279,$

$(\alpha B)=276,\quad (\beta)=539,$

$(\alpha)=749,$

$$\left.\begin{aligned}\frac{(AB)}{(B)}&=0{\cdot}0108\\ \frac{(A\beta)}{(\beta)}&=0{\cdot}1225\end{aligned}\right\}\text{(i)}\qquad \left.\begin{aligned}\frac{(AB)}{(A)}&=0{\cdot}0435\\ \frac{(\alpha B)}{(\alpha)}&=0{\cdot}3685\end{aligned}\right\}\text{(ii)}$$

(i) 式よりは

$$\frac{(AB)}{(B)}<\frac{(A\beta)}{(\beta)},$$

即ち種痘せるものの罹病率は僅に 0·011 であるに拘らず,種痘せざるもの (β) の罹病率は 0·123 であつて,遙に大である. 即ち,負の關聯があることが明である. 又 (ii) 式よりは,

$$\frac{(AB)}{(A)}<\frac{(\alpha B)}{(\alpha)}$$

となり,種痘せるものの中罹病せる者の比は, 0·044 であるに拘らず,種痘せざるものの中罹病せるものの比(確率)は,實に, 0·369 であつた. (i) と (ii) とは同一の事を見方を變へたのみで,何れにしても種痘は罹病と負の關聯にありて,種痘は豫防法として有效なることを證して居る.

尙ほこの問題では關聯係數と,並に其の誤差を考慮せよ.

10. 完全なる關聯

$$(AB)_p=(A)_p(B)_p$$

なる無關聯の狀態から,左邊が右邊より大となれば正の關聯となる. 又兩邊の差漸次大となれば,正關聯亦漸次顯著となる. けれども $(AB)_p$ は $(A)_p$ より大となることなく, $(B)_p$ より大となることもない. 即ち $(AB)_p$ の取りうる最大値は, $(A)_p$, $(B)_p$ のうち小なる方の値である. 何となれば,常に

$$(A)_p=(AB)_p+(A\beta)_p$$

であつて, $(A\beta)_p$ は決して負となることがないから,

$$(A)_p\geqq(AB)_p,$$

同樣に $$(B)_p\geqq(AB)_p,$$

なるが故である. 若し, $(AB)_p=(A)_p$ 若くは $(AB)_p=(B)_p$ なる

何れか一方の式が滿足せらるれば,A と B とは**完全に正に關聯**して居ると云ふ．この時は AB なる階級に所屬するものは悉く A に屬するか,若くは悉く B に所屬するかである．次に,

$$(AB)_p=(A)_p(B)_p$$

なる狀態より,左邊が漸次右邊より小となる時に負の關聯が生じ,かくて $(AB)_p$ 益す小となる時, $(AB)_p$ の取りうる最小値は 0 となる．かくの如き關聯を**完全なる負の關聯**と云ふ．

完全なる正の關聯と,完全なる負の關聯を總稱して,**完全なる關聯**と云ふ．

この事を表にすれば,第72表及び第73表となる．完全に正に關聯するとは $(\alpha B)_p$ と $(A\beta)_p$ との何れか一つ或は共に 0 であることである,即ち第72表の三つの場合の何れかである．

第 72 表

完全なる正の關聯

甲

	A	α	$A+\alpha$
B	$(B)_p$	0	$(B)_p$
β	$(A\beta)_p$	$(\alpha)_p$	$(\beta)_p$
$B+\beta$	$(A)_p$	$(\alpha)_p$	1

乙

	A	α	計
B	$(A)_p$	$(\alpha B)_p$	$(B)_p$
β	0	$(\beta)_p$	$(\beta)_p$
計	$(A)_p$	$(\alpha)_p$	1

丙

	A	α	計
B	$(A)_p=(B)_p$	0	$(B)_p$
β	0	$(\alpha)_p=(\beta)_p$	$(\beta)_p$
計	$(A)_p$	$(\alpha)_p$	1

完全に負に關聯せる場合は,次の三つの場合の何れかである．

第 73 表

完全なる負の關聯

甲

	A	α	計
B	0	$(B)_p$	$(B)_p$
β	$(A)_p$	$(\alpha\beta)_p$	$(\beta)_p$
計	$(A)_p$	$(\alpha)_p$	1

乙

	A	α	計
B	$(AB)_p$	$(\alpha)_p$	$(B)_p$
β	$(\beta)_p$	0	$(\beta)_p$
計	$(A)_p$	$(\alpha)_p$	1

丙

	A	α	計
B	0	$(B)_p=(\alpha)_p$	$(B)_p$
β	$(A)_p=(\beta)_p$	0	$(\beta)_p$
計	$(A)_p$	$(\alpha)_p$	1

11. 關聯係數

完全なる關聯,又は無關聯の場合は,稀で,普通はこの二つの場合の中間の狀態にある．

そこで,**關聯の程度**を表示する係數を計算したいのであるが,この係數としては, (i) 相關係數に於ける如く, -1 と $+1$ との間にあること, (ii) 無關聯の時は 0 となり, (iii) 正關聯では「+」,負關聯では「−」となり, (iv) 又完全關聯の

第 74 表

甲

階級	A	α	計
B	$\frac{(A)(B)}{N}+\Delta$	$\frac{(\alpha)(B)}{N}-\Delta$	(B)
β	$\frac{(A)(\beta)}{N}-\Delta$	$\frac{(\alpha)(\beta)}{N}+\Delta$	(β)
計	(A)	(α)	N

乙

階級	A	α	計
B	$(A)_p(B)_p+\delta$	$(B)_p(\alpha)_p-\delta$	$(B)_p$
β	$(A)_p(\beta)_p-\delta$	$(\alpha)_p(\beta)_p+\delta$	$(\beta)_p$
計	$(A)_p$	$(\alpha)_p$	1

時は, $+1$ 又は -1 となる係數であつて欲しいのである. 今第69表の甲及乙に代ふるに,第74表甲,乙を以てすることが出來る. δ(若くは $\varDelta$) が正なる數である時は, A と B とは正に關聯し, δ (又は $\varDelta$) 負なる時は負の關聯である. 又, δ が大なる程關聯も亦顯著である. 今

$$Q=\frac{(AB)_p(\alpha\beta)_p-(A\beta)_p(\alpha B)_p}{(AB)_p(\alpha\beta)_p+(A\beta)_p(\alpha B)_p}\quad\cdots\cdots\cdots\cdots(\mathrm{A})$$

に於て, Q の絕對値は常に 1 より小,且つ分子は正又は負となるが故に　　$-1<Q<+1$

である. 故に上記の (i) を滿足する. 又 Q の分子を變化させて見る.

$$\begin{aligned}
&(AB)_p(\alpha\beta)_p-(A\beta)_p(\alpha B)_p\\
&\quad=\{(A)_p(B)_p+\delta\}\{(\alpha)_p(\beta)_p+\delta\}-\{(A)_p(\beta)_p-\delta\}\{(\alpha)_p(B)_p-\delta\}\\
&\quad=\delta\{(A)_p(B)_p+(\alpha)_p(\beta)_p+(A)_p(\beta)_p+(\alpha)_p(B)_p\}\\
&\quad=\delta
\end{aligned}$$

括弧の中は 1 に等しいからである.

そこで　　$Q=\dfrac{\delta}{(AB)_p(\alpha\beta)_p+(A\beta)_p(\alpha B)_p}$

分母は常に正なる値を取るが,正關聯なる時は δ は正で,從て Q も亦正となる. 又負關聯なる時は δ は負. Q は從て負となる. 又無關聯である時は $\delta=0$, 又 $Q=0$ となり,完全關聯の時は (A) 式は 1 となり,前記 (ii) (iii) (iv) を滿足する. 故に Q を以て關聯の程度をあらはす一の係數となし得る.

之を**關聯係數** (Coefficient of association) と云ふ.

しかるに, $$Q=\frac{(AB)_p(\alpha\beta)-(A\beta)_p(\alpha B)_p}{(AB)_p(\alpha\beta)_p+(A\beta)_p(\alpha B)_p}\quad\cdots\cdots(\mathrm{X.\ 14.\ a})$$

の分母子に N^2 を乘ずれば,

$$Q=\frac{(AB)(\alpha\beta)-(A\beta)(\alpha B)}{(AB)(\alpha\beta)+(A\beta)(\alpha B)}\quad\cdots\cdots\cdots\cdots(\mathrm{X.\ 14.\ b})$$

となる. この (b) 式を以て, Q の計算式としてもよい. 又,第74表甲の如く,

$$(AB)=\frac{(A)(B)}{N}+\varDelta$$

とおけば, $\varDelta$ と δ は同一附號を有す. 且つ,

$$(AB)_p-(A)_p(B)_p=\delta$$

であるから,兩邊に N を乘じて,

$$(AB)-\frac{(A)(B)}{N}=N\delta,\qquad 故に\qquad \varDelta=N\delta,$$

又 $(AB)-\dfrac{(A)(B)}{N}=\varDelta$ の兩邊に N を乘じて變化させると,

$$\begin{aligned}
&(AB)\{(AB)+(\alpha B)+(A\beta)+(\alpha\beta)\}\\
&\quad-\{(AB)+(A\beta)\}\{(AB)+(\alpha B)\}=N\varDelta\\
&\qquad(AB)(\alpha\beta)-(A\beta)(\alpha B)=N\varDelta\quad\cdots\cdots\cdots\cdots\cdots(\mathrm{C})
\end{aligned}$$

故に,關聯係數 Q の計算は,

$$Q=\frac{(AB)_p(\alpha\beta)_p-(A\beta)_p(\alpha B)_p}{(AB)_p(\alpha\beta)_p-(A\beta)_p(\alpha B)_p}=\frac{\delta}{(AB)_p(\alpha\beta)_p+(A\beta)_p(\alpha B)_p},$$

又は, (X. 14. b) 式に (C) 式を代入して,　　(X. 14)

$$Q=\frac{(AB)(\alpha B)-(A\beta)(\alpha B)}{(AB)(\alpha\beta)+(A\beta)(\alpha B)}=\frac{N\varDelta}{(AB)(\alpha\beta)+(A\beta)(\alpha B)}$$

なる二式の何れかを用ふる.

12. 關聯係數 Q の値の變化

$[A]$ と $[B]$ とが無關聯なる時は,　　$Q=0$

完全なる正關聯にある時は,　$Q=+1$

完全なる負關聯にある時は,　$Q=-1$

$[A]$ と $[B]$ とが一般に正關聯にある時，$0<Q<+1$

〃　負關聯にある時，$-1<Q<0$

一般に Q の絕對値の大なるものは,小なるものより關聯が高度である.

例 6. スヰートピー (Sweet pea, Lathyrus odortus) の花瓣が紫色で花粉が長味を持つた一品種と,花瓣が赤色で花粉が圓き一品種との雜種(雜種第一代 F_1) では,悉く花瓣は紫色,花粉は長味を持つて居た. 雜種第二代 (F_2) では, 11,291 個の觀測で次の結果を得た.

但し　A は花瓣の紫色なる階級,

α は花瓣の赤色なる階級,

B は花粉の長き階級,

β は花粉の短き階級とすれば,

$(AB)=7,897$

$(A\beta)=\ \ 583$

$(\alpha B)=\ \ 614$

$(\alpha\beta)=\ 2,197$

$N=11,291$

これより關聯係數 Q を計算せよ.

解. $$Q=\frac{(AB)(\alpha\beta)-(A\beta)(\alpha B)}{(AB)(\alpha\beta)-(A\beta)(\alpha B)}=\frac{7,897\times 2,197-583\times 614}{7,897\times 2,197+583\times 614}$$

計算には大なる數を取扱ふのは不便であるから分母子を 1000^2 で割る時は

$$Q=\frac{7{\cdot}897\times 2{\cdot}197-{\cdot}583\times{\cdot}614}{7{\cdot}897\times 2{\cdot}197+{\cdot}583\times{\cdot}614}=\frac{17{\cdot}35-{\cdot}358}{17{\cdot}35+{\cdot}358}$$

$$=\frac{16{\cdot}992}{17{\cdot}708}=0{\cdot}959\fallingdotseq 0{\cdot}96$$

高度の正關聯がある. 獨立でない.

13. 相關係數

二屬性間の關聯の程度を表示するに,關聯係數は唯一のものではない. 關聯表に於て, A なる屬性を有するものを 1 なる階級と考へ, α なる階級を 0 なる階級と考ふることが出來る. 卽ち成功 1 と成功 0 との二階級と考へるのである. 又 B なる屬性につきても同樣である. かく考ふる時は變數統計に於けると同樣に,關聯表は最も單純なる相關表であるから,此の表より相關係數 r を計算することが出來る.

第 75 表

[A] / [B]	1	0	計
1	(AB)	(αB)	(B)
0	$(A\beta)$	$(\alpha\beta)$	(β)
計	(A)	(α)	N

第45表は A, α, B, β に代ふるに 1, 0, 1, 0 を以てしたのみである. これより相關係數を計算する公式を導く.

$[A]$ 屬性の算術的平均,

$$M_A=\frac{1}{N}\{1\times(A)+0\times(\alpha)\}=\frac{(A)}{N}=(A)_p,$$

$[B]$ 屬性の算術的平均,

$$M_B=\frac{(B)}{N}=(B)_p,$$

$[A]$ 屬性の標準偏差 σ_A は,

$$\sigma_A{}^2=\frac{1}{N}\{1-(A)_p\}^2(A)+\{0-(A)_p\}^2(\alpha)\},$$

$$=\frac{1}{N}\{(A)(\alpha)_p{}^2+(A)_p{}^2(\alpha)\},$$

$$\sigma_A=\sqrt{(A)_p(\alpha)_p\{(A)_p+(\alpha)_p\}},$$

故に　$$\sigma_A=\sqrt{(A)_p(\alpha)_p},$$

$[B]$ 屬性の標準偏差は同樣に $\sigma_B=\sqrt{(B)_p(\beta)_p}$,

又各欄の偏差の積の和は,

$$\sum(fxy)=(AB)\{1-(A)_p\}\{1-(B)_p\}+(\alpha\beta)\{0-(A)_p\}\{0-(B)_p\}$$
$$+(A\beta)\{1-(A)_p\}\{0-(B)_p\}+(\alpha B)\{0-(A)_p\}\{1-(B)_p\},$$

$$\therefore\ \frac{1}{N}\sum(fxy)=(AB)_p(\alpha)_p(\beta)_p+(\alpha\beta)_p(A)_p(B)_p-(A\beta)_p(\alpha)_p(B)_p$$
$$-(\alpha B)_p(A)_p(\beta)_p,$$
$$=(A_p)(B)_p(\alpha)_p(\beta)_p+\delta(\alpha)_p(\beta)_p+(A)_p(B)_p(\alpha)_p(\beta)_p$$
$$+\delta(A)_p(B)_p-(A)_p(B)_p(\alpha)_p(\beta)_p+\delta(\alpha)_p(B)_p$$
$$-(A)_p(B)_p(\alpha)_p(\beta)_p+\delta(A)_p(\beta)_p,$$
$$=\delta\{(A)_p(B)_p+(\alpha)_p(\beta)_p+(\alpha)_p(B)_p+(A)_p(\beta)_p\},$$
$$=\delta$$

然るに $r=\dfrac{\frac{1}{N}\sum(fxy)}{\sigma_A\sigma_B}$

即ち $r=\dfrac{\delta}{\sqrt{(A)_p(B)_p(\alpha)_p(\beta)_p}}$,

又分母子に N^2 を乘じて ……(X. 15)

又は $r=\dfrac{N\varDelta}{\sqrt{(A)(B)(\alpha)(\beta)}}$,

但し $\varDelta=(AB)-\dfrac{(A)(B)}{N}$ とす,

關聯係數 Q と相關係數 r とを比較するに,

$Q=0$ なる時は, $r=0$,

$Q=+1$ なる時は, $r=+1$,

$Q=-1$ なる時は, $r=-1$

であつて,この點で Q と r とは等しいが,この値以外では Q と r とは附號を同じくするが,値は同一でない.

例 7. 例 6 に於ける Bateson の Sweet pea の例について, r を計算せよ.

解.
$$(AB)=7{,}897,$$
$$(A\beta)=583,$$
$$(\alpha B)=614,$$
$$(\alpha\beta)=2{,}197,$$
$$N=11{,}291,$$
$$(A)=(AB)+(A\beta)=8{,}480,\quad (\alpha)=2{,}811,$$
$$(B)=(AB)+(\alpha B)=8{,}511,\quad (\beta)=2{,}780,$$
$$\varDelta=(AB)-\frac{(A)(B)}{N}=7{,}897-6{,}392=1{,}505,$$
$$r=\frac{1{,}505\times 11{,}291}{\sqrt{8{,}480\times 8{,}511\times 2{,}811\times 2{,}780}},$$

分母子を 1000^2 で除して,

$$r=\frac{1{\cdot}505\times 11{\cdot}291}{\sqrt{8{\cdot}480\times 8{\cdot}511+2{\cdot}811\times 2{\cdot}780}}=0{\cdot}714,$$

高度の相關があるが, $Q=0{\cdot}959$ に比して著しく小である.

第四項 部分關聯

14. 部分關聯及全關聯

前項では二つの屬性 $[A]$ と $[B]$ の關聯を論じたのであるが,もし第三の屬性 $[C]$ をも考慮する時は, $[A]$ と $[B]$ との關聯はかくの如く簡單でない. たとひ $[A]$ と $[B]$ とが關聯して居ても,この關聯が直接でなく, $[C]$ が $[A]$ 及び $[B]$ と關聯して居るが爲に,間接に $[A]$ と $[B]$ の間に關聯が起ることがある. 例をあぐれば,種痘したものが天然痘にかかること少く,種痘しないものが罹

病することが多い. 今種痘せるものを A, せざるものを α, 天然痘にかかれるものを B, 罹病せざるものを β とすれば, A と B とは負に關聯して居る. 此の事實は覆ふことの出來ない事實であるが(第70表), ここに考ふべきことは A と B との負の關聯は直接であるか, 或は間接であるかと云ふことである. 例へばここに不衞生なる生活をなす者と然らざる者とがあり, 前者に種痘せざる者多く, 又罹病者の多い事も考へうるのである. かかる時はたとひ種痘と罹病とが本來無關聯でありとするも, 間接に關聯が現れて來るかも知れないのである. この問題を決定せんとならば, 衞生狀態同一なる者の範圍內に於て, 觀測を行へばいい. 即ち不衞生生活をする者について, $[A]$ と $[B]$ との關聯を求め, 又衞生良好なる者のみについて, $[A]$ と $[B]$ との關聯を求むればいい. 若し何れの條件のもとに於ても, 尙, A と B とが高度の關聯にある時は, 種痘と罹病との關聯は直接であつて, 衞生狀態に關しないことが明になる. 又, 或は, 幾分は衞生狀態に起因するか, 或は A と B との關聯は實は衞生狀態の良, 不良から來るものであるかが明になる. 二屬性 $[A]$ と $[B]$ との關聯は全範圍, 即ち觀測せる總數につきて云ふのであるから, 之を**全關聯** (total association) と云ふ. しかるに今 $[C]$ なる屬性を考へて, C 階級內又は γ 階級內に於て, $[A]$ と $[B]$ との關聯を考へるとき, かく限られた範圍(或る階級)內に於ける $[A]$ と $[B]$ との關聯を**部分關聯** (partial association) と云ふ. 全關聯では

$$(AB)_p \gtreqqless (A)_p(B)_p \quad \text{又は} \quad (AB) \gtreqqless \frac{(A)(B)}{N}$$

なる式は正關聯, 獨立及負關聯の判定式であつた.

C の範圍內に於てならば,

$$\left.\begin{aligned} &(ABC)_p \gtreqqless \frac{(AC)_p(BC)_p}{(C)_p}, \\ \text{或は}\quad &(ABC) \gtreqqless \frac{(AC)(BC)}{(C)}, \end{aligned}\right\} \cdots\cdots\cdots\cdots (\text{X. 16. a})$$

又 γ 範圍でならば

$$\left.\begin{aligned} &(AB\gamma)_p \gtreqqless \frac{(A\gamma)_p(B\gamma)_p}{(\gamma)_p}, \\ &(AB\gamma) \gtreqqless \frac{(A\gamma)(B\gamma)}{(\gamma)}, \end{aligned}\right\} \cdots\cdots\cdots\cdots (\text{X. 16. b})$$

が正關聯, 獨立及び負關聯の判定式となるのである. 部分關聯でも, 以上の式から其の正負又は無關聯を示す多數の不等式及び等式を誘導することが出來るが, 何れも (X. 11) 以下にならふことが出來るから省略する. 唯普通用ひられる二三の式をあぐれば,

$$\left.\begin{aligned} &\frac{(ABC)}{(BC)} > \frac{(AC)}{(C)}, \\ &\frac{(ABC)}{(AC)} > \frac{(BC)}{(C)}, \\ &\frac{(ABC)}{(BC)} > \frac{(A\beta C)}{(\beta C)}, \\ &\frac{(ABC)}{(AC)} > \frac{(\alpha BC)}{(\alpha C)}, \end{aligned}\right\} \cdots\cdots\cdots\cdots\cdots (\text{X. 16. c})$$

は何れも正關聯をあらはす式である.

屬性四個以上の場合も同樣で, これを擴張することが出來る. 若し CD の範圍を選べば,

$$\left.\begin{aligned} &(ABCD) > \frac{(ACD)(BCD)}{(CD)}, \\ &(ABCD)_p > \frac{(ACD)_p(BCD)_p}{(CD)_p}, \end{aligned}\right\} \cdots\cdots (\text{X. 16. d})$$

又 $\gamma\delta$ の範圍を選ぶならば,

$$(AB\gamma\delta) > \frac{(A\gamma\delta)(B\gamma\delta)}{(\gamma\delta)}$$

等が正關聯を示すものである. 他の範圍を選ぶ時も,類似の方法で式を誘導することが出來る.

屬性五個以上の場合も,亦同様に研究することが出來るが,しかし,かく範圍が次第に複雜となれば,級度數が小になるから,偶然の影響が大となる. 故に,なるべく範圍を一次にとり,C 又は γ 範圍内に於ける A と B との關聯を研究し,次で,D 及び δ 範圍内に於ける A と B との關聯を考察する様にするがいい.

例 7. 多數の學童につきて一定の缺陷を觀察し,これを一萬分比として表示したのが,次の通りとなつた. (A) は發育不完,(B) は神經徴候,(D) は遲鈍なる者の數である.

$$N = 10{,}000, \qquad (AB) = 338,$$
$$(A) = 877, \qquad (AD) = 338,$$
$$(B) = 1{,}086, \qquad (BD) = 455,$$
$$(D) = 789, \qquad (ABD) = 153,$$

原報告には,身體の缺陷と遲鈍との間の連鎖は,異常神經徴候として知らるる腦の缺陷に因すると,結論して居るが,この結論を吟味せよ.

連鎖 (connecting link) と云ふ語は,聊か曖昧であるが,多分神經徴候 B が指示する腦の缺陷は,發育不全 A, 及び遲鈍 D を惹起する. 即ち A 及び D は同一の原因 B の結果であつて, A と D とが直接に影響しあつて居るのではないと云ふのであらう. 故に,この問題では A と D との關聯を全範圍と, B 範圍(B階級)

と,β 範圍(β 階級)内で研究しなければならぬ.

全範圍で,

(i) 遲鈍の比 $\dfrac{(D)}{N} = \dfrac{789}{10{,}000} = 0{\cdot}079,$

(ii) 發育不完全なるものの内遲鈍なるものの比 $\dfrac{(AD)}{(A)} = \dfrac{338}{877} = 0{\cdot}385,$

神經徴候を現はすものの範圍(B 範圍)に於て,

(i) 遲鈍の比 $\dfrac{(BD)}{B} = \dfrac{455}{1{,}086} = 0{\cdot}419,$

(ii) 發育不全なるものの内遲鈍なるものの比 $\dfrac{(ABD)}{(AB)} = \dfrac{153}{338} = 0.453,$

神經徴候を現はさない者の範圍(β 範圍)に於て,

(i) 遲鈍の比 $\dfrac{(\beta D)}{(\beta)} = \dfrac{334}{8{,}914} = 0{\cdot}037,$

(ii) 發育不全なるものの内遲鈍なるものの比 $\dfrac{(A\beta D)}{(A\beta)} = \dfrac{185}{539} = 0{\cdot}343.$

計算の結果は極めて顯著で,全範圍に於ける A と D との關聯も,β 範圍に於ける其れも,共に高度であるが, B 範圍内では A と D との關聯が低い. この結果は原報告の結論と一致して居ない. 何となれば,原報告の結論の云ふ如くならば,B 範圍に於ても β 範圍に於ても A と D との關聯は極めて低くなければならない. 事實は之に反するからである.

今假りに全範圍, B 範圍, β 範圍に於ける關聯表を作ると次の通りとなる. 若し,三つの關聯係數を計算するならば,一層この事實は明になるであらう.

甲

全範圍に於ける A と D との關聯表

	A	α	計
D	338	451	789
δ	539	8,672	9,211
計	877	9,123	10,000

乙

B 範圍に於ける A と D との關聯表

	AB	αB	計
BD	153	302	455
$B\delta$	185	446	631
計	338	748	1,086

丙

β 範圍に於ける A と D との關聯表

	$A\beta$	$\alpha\beta$	計
βD	185	149	335
$\beta\delta$	354	8,226	8,580
計	539	8,375	8,914

この三つの關聯表から三つの關聯係數を求むるに,全範圍に於ける A と D との關聯係數 Q_u 其他 Q_B, Q_β は次の通りとなる.

$$Q_u = \frac{\cdot338 \times 8\cdot672 - \cdot539 \times \cdot451}{\cdot338 \times 8\cdot672 + \cdot539 \times \cdot451} = 0\cdot846$$

$$Q_B = \frac{1\cdot53 \times 4\cdot46 - 3\cdot02 \times 1\cdot85}{1\cdot53 \times 4\cdot46 + 3\cdot02 \times 1\cdot85} = 0\cdot100$$

$$Q_\beta = \frac{\cdot185 \times 8\cdot226 - \cdot149 \times \cdot354}{\cdot185 \times 8\cdot226 + \cdot149 \times \cdot354} = 0\cdot933$$

B 範圍に於ける A と D との關聯は極めて微弱であるが, β 範圍に於ける關聯は極めて高度である. 若し,著者の云ふ如くならば B 範圍にも β 範圍にも關聯が微弱でなければならぬが,事實は其の然らざることを證明して居る. 全範圍に於ける高度の關聯は,其の大部分を占むる β 範圍内(神經徴候なきもの)に

於ける顯著な關聯に由來するのである.

15. 關聯係數

二屬性の場合に於ては,唯一個の關聯係數があるのみであるが,三屬性になれば, 9 個の關聯係數が計算しうる. 卽ち,三個の屬性を二つづつ取れば,三つの組合せが出來るが,各の組合せに一の全關聯と二個の部分關聯と,合せて 3 個の係數が出來るから,合計 9 個の關聯係數が計算しうる. 若し四個の屬性を取れば, 54 個,五屬性につきては 270, 六屬性につきては 1,125 の係數が計算しうることとなる.

實際の場合には,かく總ての關聯係數を計算する必要はなく,問題の性質により,目的に叶へるもののみを精査して計算すればいいのである. 例へば,例 7 に於ては三個の係數を計算すればいい.

第五項 多項分類

16. 多項分類

二項分類は屬性統計法の最も簡單なるもので,又最基礎的分類法であるが,統計學では三階級又はそれ以上に分類する必要が屢起る. この分類法を**多項分類**(**多級分類**; Manifold classification)と云ふ.

例へば,毛髮の色を明,暗,赤なる三色に分類するとか,花の色を紫,赤,白に,血液型を O, A, B, AB に分類するとかの類である. 今, [A] 屬性に於て階級をあらはすに,

$$A_1,\ A_2,\ \cdots\cdot A_s,$$

其の度數を表はすに,

$$(A_1),\ (A_2),\ (A_3),\ \cdots\cdots\ (A_s)$$

を以てし,又確率を表はすにはこれに副附號 p を附記す. 其の他, B なる屬性につきても,同様に記號す.

其の二次の階級は　$A_1B_1,\ A_1B_2,\ A_1B_3,\ \cdots\cdots$

度數は　$(A_1B_1),\ (A_1B_2)\ (A_1B_3)\ \cdots\cdots$

等を以て記號する. 但し,何れの階級も,互に排反でなければならぬ. 二項分類で二個の屬性を取扱ふ時は,關聯表を作つた如く,多數の屬性を取扱ふ場合は**配分表** (Contingency-table) を作製する.

例 8. 次表は Ammon 氏の數で,毛髮の色と眼の色との配分表である. (Ammon, Zur Anthropologie der Badener)

第 76 表

Baden に於ける 6,800 人の毛及び眼の色の配分表

眼の色 ＼ 毛の色	淡色	褐	黒	赤	計
青	1,768	807	189	47	2,811
綠又は灰白	946	1,387	746	53	3,132
褐	115	438	288	16	857
計	2,829	2,632	1,223	116	6,800

この表は一寸相關表の如く見えるが,決して相關表の如く,各階級一定の幅を有し,且つ階段的に變化するものと同一に見るべきでない. 故に,配分表では,そのまま相關係數を計算することが出來ない. 又關聯係數も計算できないから,配分係數を計算する. 但し,配分表を關聯表に更めて,これより關聯係數或は相關係數を計算するなら出來ぬことはない.

17. 關聯係數

三階級以上に分類せる配分表は,これを關聯表に改めることが出來る. 例へば $[A]$ に於て, A_1 なる階級と, $A_2+A_3+\cdots\cdots+A_s$ 卽ち (非 A_1) $=(\alpha)$ との二階級に分類することが出來; B につきても亦同様である. 卽ち上例 (例 8) は第77表の如く縮小することが出來る.

かく關聯表となれば,關聯係數を計算することは容易である. しかし,分類のしかたを變更して,他の關聯表を作製することも出來るから,關聯係數も一でない. 例へば毛髮の色は黒と非黒とに分類し,或は赤と非赤とに分類し,眼の色は褐色と非褐色とに,或は綠と非綠とに分類することも出來る. 卽ち例 8 の表からは

第 77 表

眼の色 ＼ 毛の色	淡色	非淡色	計
青	1,768	1,043	2,811
非青	1,061	2,928	3,989
計	2,829	3,971	6,800

$3\times4=12$ 種の異なる關聯表を作製することが出來,從て,それだけの關聯係數を計算することが出來る. しかし,可能なる總ての係數を計算する必要はないので,目的に應じた必要な計算のみを行へばいいのである.

今 $[A]$ では A_m, 又 $[B]$ では B_n を取り,他の階級は一まとめとして 非 A_m, 若くは 非 B_n とすれば, A_m と B_n とが無關聯なる條件は,

$$\left.\begin{aligned}&(A_mB_n)_p=(A_m)_p(B_n)_p,\\ \text{又は}\quad &(A_mB_n)=\frac{(A_m)(B_n)}{N},\end{aligned}\right\}$$

である. A_m と B_n との正なる關聯の條件としては,

$$\left.\begin{array}{ll} & (A_mB_n)_p > (A_m)_p(B_n)_p, \\ \text{若くは,} & (A_mB_n) > \dfrac{(A_m)(B_n)}{N}, \end{array}\right\}$$

負の關聯では,不等附號を反對の方向にむければいいのである. 又,關聯係數も前々項に說けるが如く,簡單に計算することが出來る.

例 9. 第77表の毛髮及び眼の色の關聯表より,關聯係數を計算せよ.

$$N = 6{,}800$$
$$(AB) = 1{,}768$$
$$(A\beta) = 1{,}061$$
$$(\alpha B) = 1{,}043$$
$$(\alpha\beta) = 2{,}928$$

$$Q = \frac{(AB)(\alpha\beta) - (A\beta)(\alpha B)}{(AB)(\alpha\beta) + (A\beta)(\alpha B)}$$

$$= \frac{1{\cdot}768 \times 2{\cdot}928 - 1{\cdot}061 \times 1{\cdot}043}{1{\cdot}768 \times 2{\cdot}928 + 1{\cdot}061 \times 1{\cdot}043}$$

$$= \frac{5{\cdot}18 - 1{\cdot}107}{5{\cdot}18 + 1{\cdot}109} = \frac{4{\cdot}073}{6{\cdot}287} = 0{\cdot}648 \fallingdotseq 0{\cdot}65$$

可なり高度の正關聯がある. 若し,他の組合せを作れば,關聯係數著しく相違することを忘れてはならぬ.

又配分表からは,直接に相關係數を計算することが出來ないが,これを第77表の如き關聯表に改める時は,成功數 0, 1 と考へて相關係數が計算出來ないことはない. $[A]$ 屬性は t 階級, $[B]$ 屬性が s 階級ある時, ts 個の相關係數を計算しうる,が相關係數

はこの際あまり用ひられない.

18. 配分係數

前條で說明した關聯係數は, A_m, B_n の選擇のしかた次第で,幾通りも計算しうるのである. 若し, $[A]$ と $[B]$ とが全く獨立であるならば,即ち,無關聯であるならば,すべての關聯係數 Q が 0 でなければならぬ.

$$(A_mB_n) - \frac{(A_m)(B_n)}{N} = \varDelta_{mn}$$

と置けば, A_m と B_n とが無關聯である爲には, $\varDelta_{mn}$ が 0 である. すべての區劃の $\varDelta$ が 0 でなければならぬ. でなければある程度の關聯がある. 今 $\dfrac{(A_m)(B_n)}{N}$ を以て $\varDelta_{mn}{}^2$ を除して $\dfrac{N\varDelta_{mn}{}^2}{(A_m)(B_n)}$ を作り,次に其の總和 χ^2 を求む. 即ち,

$$\chi^2 = \sum\left(\frac{N\varDelta_{mn}{}^2}{(A_m)(B_n)}\right)$$

又は,

$$\chi^2 = \sum\left\{\frac{N\delta_{mn}{}^2}{(A_m)_p(B_n)_p}\right\}$$

を作つて見る. χ^2 は常に正である. 又 A と B とが完全に無關聯ならば, χ^2 は 0 である. そこで,

$$C = \sqrt{\frac{\chi^2}{N+\chi^2}} \quad \cdots\cdots\cdots\cdots\cdots\cdots (\text{X. } 17)$$

を考ふるに A と B とが無關聯なる時には, $C = 0$, 又 χ^2 が 0 より漸次大となれば, C は漸次大となり, $\chi^2 = \infty$ なる時 $C = 1$ となる. 平方根の正なる値をとれば, C は

$$0 \leqq C < 1$$

の値を取る. C を**配分係數**,又は **Pearson の配分係數**と云ふ(Contingency coefficient, Pearson's mean-square contingency coefficient.)

しかし,かくして計算せられたる C は,他の分類法から計算せる係數と比較が出來ない不利がある．同一の計測群からは他の分類法を行つて計算しても,同一(又は近似の)係數が出て來なければならぬが,此の C は其の性質を持つて居ない．

又,C に正負を考へ得る．例7の如く,毛髪の淡色は眼の青色と正に關聯し,毛髪の黒色又は褐色の眼の色の褐色と正關聯して居る．かくの如く全體として大體正に關聯して居る時は C は正の値をとる．若し,表全體として負の關聯にある時は,C は負の値を取る．然る時は $-1 < C < +1$

となる．然し,C の正負を定め得ないことが多いから,C に正負を考へるのは初學者には危險で,絶對値を取るのが安全である．C の最小値は0であるが,最大値を詳しく考へて見る．

$$\chi^2 = \sum \frac{N\Delta_{mn}{}^2}{(A_m)(B_n)},$$

而して, $$(A_mB_n)_0 = \frac{(A_m)(B_n)}{N}$$

と置けば,

$$\Delta_{mn} = (A_mB_n) - (A_mB_n)_0$$

であるから,

$$\Delta_{mn}{}^2 = (A_mB_n)^2 + (A_mB_n)_0{}^2 - 2(A_mB_n)(A_mB_n)_0,$$

$$\therefore \quad \chi^2 = \sum \frac{(A_mB_n)^2 + (A_mB_n)_0{}^2 - 2(A_mB_n)(A_mB_n)_0}{(A_mB_n)_0},$$

$$= \sum \left\{ \frac{(A_mB_n)^2}{(A_mB_n)_0} + (A_mB_n)_0 - 2(A_mB_n) \right\},$$

而して $\sum (A_mB_n) = N$, 又 $\sum (A_mB_n)_0 = N$ であるから,

$$\chi^2 = \sum \frac{(A_mB_n)^2}{(A_mB_n)_0} - N,$$

故に $\sum \frac{(A_mB_n)^2}{(A_mB_n)_0} = s$ を以て表すならば

$$\chi^2 = s - N$$

故に $$C = \sqrt{\frac{s-N}{s}} \quad \cdots\cdots\cdots\cdots\cdots\cdots (X.\ 18)$$

さて $t \times t$ 個の階級の配分表に於て,完全なる關聯がある時には,總ての m に關して,$(A_mB_m) = (A_m) = (B_m)$ となり,總ての度數は表の對角線上に集り,他の區劃の度數は0となる．此の時は

$$s = \sum \frac{(A_mB_m)^2}{(A_mB_m)_0} = \sum \frac{(A_m)^2N}{(A_m)^2} = \sum N = Nt,$$

故にこの時は,

$$C = \sqrt{\frac{s-N}{s}} = \sqrt{\frac{t-1}{t}},$$

即ち完全に關聯がある時,C は $\sqrt{\frac{t-1}{t}}$ となる．若し,A も B も共に三階級に分類せられて居る時は,$C = \sqrt{\frac{2}{3}}$ となり,これが C の取り得る最大値である．種々の t について C の取り得る最大値は第78表に示す通りである．

第 78 表

種々の t に對し C の取りうる最大價

t	C の取り得る最大値
2	$C = 0{\cdot}707$
3	0·816
4	0·866
5	0·894
6	0·913
7	0·926
8	0·935
9	0·943
10	0·949

例 10. 例8(第76表)に於ける配分表より配分係數 C を求めよ．

解. 先づ $$(A_mB_n)_0 = \frac{(A_m)(B_n)}{N}$$

を求め,これを各級度數の下に括弧の中に記入する(第79表)．これより

$$\frac{(A_mB_n)^2}{(A_mB_n)_0}, \qquad s = \sum \frac{(A_mB_n)^2}{(A_mB_n)_0}$$

を計算して, Cを求む.

第 79 表

毛髪の色 / 眼の色	淡色	褐色	黒色	赤色	計
青	1,768 (1,169)	807 (1,088)	189 (506)	47 (48·0)	2,811
綠及び灰白	946 (1,303)	1,387 (1,212)	746 (563)	53 (53·4)	3,132
褐	115 (357)	438 (332)	288 (154)	16 (14·6)	857
計	2,829	2,632	1,223	116	6,800

$$
\begin{aligned}
1768^2/1169 &= 2674\\
946^2/1303 &= 687\\
115^2/357 &= 37\\
807^2/1088 &= 599\\
1387^2/1212 &= 1587\\
438^2/332 &= 578\\
189^2/506 &= 71\\
746^2/563 &= 989\\
288^2/154 &= 539\\
47^2/48 &= 46\\
53^2/53{\cdot}4 &= 53\\
16^2/14{\cdot}6 &= 18
\end{aligned}
$$

計 $s = 7{,}878$

$$N = 6{,}800$$
$$s - N = 1{,}078$$
$$C = \sqrt{\frac{s-N}{s}} = \sqrt{0{\cdot}1368} = 0{\cdot}370$$
$$C = 0{\cdot}370$$

演習問題

1. 次は多數の男性の學童について,一定の缺陷を觀察せる度數である. A は發育不全, B は神經徴候, C は榮養不良とす.

$(ABC) = 149$ $(\alpha BC) = 204$
$(AB\gamma) = 738$ $(\alpha B\gamma) = 1{,}762$
$(A\beta C) = 225$ $(\alpha\beta C) = 171$
$(A\beta\gamma) = 1{,}196$ $(\alpha\beta\gamma) = 21{,}842$

これより積極度數を計算せよ.

2. 次の數は同一の研究に於ける女兒の積極度數である.

$N = 23{,}713$ $(AB) = 587$
$(A) = 1{,}618$ $(AC) = 428$
$(B) = 2{,}015$ $(BC) = 335$
$(C) = 770$ $(ABC) = 156$

これより最終度數を計算せよ.

3. 父の眼の色の暗色なることと,子の眼の暗色なることとの間に關聯ありやを,次の表より檢査せよ. 又其の關聯係數を計算せよ.

A は父の眼の暗色 B は子の眼の暗色とす

$(AB) = 50$ $(A\beta) = 79$
$(\alpha B) = 89$ $(\alpha\beta) = 782$

4. 夫と妻の眼の色を調べ, A を夫の眼の淡色, B を妻の眼の淡色と定むれば

$(AB) = 309$ $(A\beta) = 214$
$(\alpha B) = 132$ $(\alpha\beta) = 119$

であつた. A と B との間に關聯ありや. 又關聯係數を計算せよ.

参考文獻

本章の参考書は Yule の外に次のがある.

1. Jevons, W. Stanley :- On a general system of numerically definite reasoning, pure logic and other minor works. London, 1890.

2. Yule, G. U. :- On the theory of consistence of logical class-frequencies and its geometrical representation. *Phil. Trans. Roy. Soc. Series A. Vol. 197, 1901.*

Association に關するもの :—

3. Pearson. Karl. :- On the correlation of characters not quantitatively measurable. *Phil. Trans. Roy. Soc. Series A. Vol. 195, 1900.*

4. Pearson, Karl and Heron, David :- On the theories of association. *Biometr. Vol. 9., 1913.*

5. Yule, G. U. :- On the association of attributes in statistics, etc. *Phil. Trans. Roy. Soc. Series A, Vol. 194, 1900.*

6. Yule, G. U. :- Notes on the theory of association of attributes in statistics. *Biometr. Vol. 2., 1903.*

Partial Association に關するもの:

5 及び 6 を見よ.

Contingency に關するもの:—

7. Lipps, G. F. :- Die Bestimmung der Abhängigkeit zwischen den Merkmalen eines Gegenstandes. *Berichte d. math.-phys. Klasse d. Kögl.-Säch. Ges. d. Wissensch. Leipzig, 1905.*

8. Pearson, K. :- On the theory of contingency and its relation to association and normal correlation. *Drapers' Co. Research Memoirs, Biometirc, Series I. London, 1904.*

9. Pearson, K. :- On a coefficient of class heterogenity or divergence. *Biometr. Vol. 5, 1906.*

10. Pearson K. :- On a new method of determining correlation between a measured character A and a character B of which only the percentage of cases wherein B exceeds a given intensity is recorded for each [illegible] of A. *Biometr. Vol. 7. 1909*

11. Pearson, K. :- On a new method of determining correlation, when one variable is given by alternative and other by multiple categories. *Biometr. Vol. 7, 1910.*

第十一章

重複試行の確率と變數の確率

1. 獨立試行

第九章に述べたる諸定理のうち,獨立事象に關する複確率の定理は重要であつて,種々の定理はこれより誘導せられる.

獨立事象を重ねて試行する時は,これを**獨立事象の重複試行**,又略して**獨立試行**と云ふ.

定理に入る前に先づ例につきて說明を試みる.

例 1. 一の骰子を投じた時, 1 の目が出るを成功とし, 1 以外の目の出るを不成功とする,從て $p=\frac{1}{6}$ $q=\frac{5}{6}$ であるとし,今二個の骰子を同時に投じて,二個の骰子共に成功する確率,一個が成功する確率,及二個共に成功せざる確率をとふ.

解. 前章の獨立事象の定理(II)によりて,

二個の骰子が共に成功する確率 $p^2=\frac{1}{36}$,

第一の骰子が成功し,第二が不成功なる確率 $pq=\frac{5}{36}$,

第二の骰子が成功し,第一が不成功なる確率 $qp=\frac{5}{36}$,

二個とも不成功なる確率 $q^2=\frac{25}{36}$,

卽ち二個の骰子を投じた時二回とも成功する確率は $p^2=\frac{1}{36}$,

一個が成功する確率は $2pq=\frac{10}{36}$,

一回も成功せざる確率は $q^2=\frac{25}{36}$ である,

この三つは

$$(p+q)^2 = p^2 + 2pq + q^2$$

の右邊の各項に一致する.

例 2. もし三個の骰子を投じた時,三個ともに成功する確率,二個が成功する確率,一個が成功する確率,及一個も成功せざる確率を問ふ.

解. 二個を投じた時の確率は例 1 で明になつて居る. 三個の場合にはこれに第三の骰子が加はつたのであるから,

(i) 三個の骰子共に成功する確率は,初め二個が成功する確率 p^2 と第三の骰子の成功の確率 p との複確率であるから, p^3 である.

(ii) 又三個の中二個が成功する確率は,初め二個が成功し第三が不成功なる確率 $p^2 \times q$ と,初めの二個の中一個が成功し且つ第三が成功する確率 $2pq \times p = 2p^2q$ との全確率である. 即ち, $3p^2q$ である.

(iii) 三個の骰子中一個が成功する確率は,初め二個の中一個が成功し第三が不成功なる複確率 $2pq \times q = 2pq^2$ と,初め二個共に不成功にして第三が成功する複確率 q^2p との全確率,即ち $3pq^2$ である.

(iv) 又一個も成功せざる確率は,勿論 q^3 である. 故に,三個成功する,二個成功する,一個の成功する,及び一個も成功せざる確率は,順次に, p^3, $3p^2q$, $3pq^2$, q^3 である.

即ち, $(p+q)^3$ の展開式,

$$(p+q)^3 = p^3 + 3p^2q + 3pq^2 + q^3$$

に於ける右邊の各項に相當する.

定理. 毎回の試行(又は觀測)に於て或事象の起る確率が p, 其の事象の起らざる確率 $q(=1-p)$ なる時は, n 回の獨立試行に於て r 回だけ其の事象の起る確率は

$$_nC_r p^r q^{n-r} = \frac{n(n-1)(n-2)\cdots\cdots(n-r+1)}{r!} p^r q^{n-r}$$

$$= \frac{n!}{r!(n-r)!} p^r q^{n-r} \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots (\text{XI. } 1)$$

である.

證明. 初めに連續して r 回成功し,他はことごとく不成功なる確率を求むるに,複確率の定理により,

$$\overbrace{p.p.p.\cdots\cdot}^{r\text{個}}\overbrace{q.q.q\cdots\cdot q}^{(n-r)\text{個}} = p^r q^{n-r}$$

となる. 然るに,成功 r 回,不成功 $(n-r)$ 回なる場合はこの外にもあり,成功の順序が異つても差支ない,即ち上式の p と q との順序を代へてもいいのである. 唯 n 個の中成功せる骰子を r 個だけ取り,不成功なる骰子を $(n-r)$ 個だけ取りさへすればいいので,其の順序に關りはない. n 個の中成功せるもの r 個だけ取る組合せの數は $_nC_r$ だけあり. 而してこれらの $_nC_r$ 個の場合は皆排反事象なるが故に,兎に角 r 個成功し他は不成功なる確率は, $p^r q^{n-r}$ を $_nC_r$ 個加へたるものに當る. 即ち, $_nC_r p^r q^{n-r}$ である. 仍て定理を證明し得た.[1]

$\frac{n(n-1)(n-2)\cdots\cdot(n-r+1)}{r!} p^r q^{n-r}$ に於て, r に n, $(n-1)$, $(n-2)$, $\cdots\cdot$ 等を代入すれば,順次, p^n, $np^{n-1}q$, $\frac{n(n-1)}{2!} p^{n-2}q^2$, $\cdots\cdots\cdots$

[1] この定理は數學的歸納法を用ふれば,一層手際よく證明出來る. 讀者試みられよ.

等となる.

此の定理は二項定理と關係をつけて記憶すべきである. 二項定理によれば,

$$(p+q)^n = p^n + np^{n-1}q + \frac{n(n-1)}{1.2}p^{n-2}q^2 + \cdots\cdots$$
$$+\frac{n(n-1)\cdots\cdots(n-r+1)}{r!}p^{n-r}q^r + \cdots\cdots + npq^{n-1} + q^n$$

又は他の書きあらはし方をすれば,

$$(p+q)^n = {}_nC_0 p^n + {}_nC_1 p^{n-1}q + {}_nC_2 p^{n-2}q^2 + \cdots\cdots$$
$$+ {}_nC_r p^{n-r}q^r + \cdots\cdots + {}_nC_{n-1} pq^{n-1} + {}_nC_n q^n$$

となる.

この展開式は二つとも同じことを他の附號を用ひて書いただけであるが,右邊の第一項は n 回の獨立試行に於て全部が成功する確率,第二項は $n-1$ 回成功し一回失敗する確率,第三項は $n-2$ 回成功し, 2 回不成功なる確率, $\cdots\cdots$ に一致する.

一般項は $n-r$ 回成功し r 回不成功なる確率である. 此の公式は本章以後屢々用ひられるが, p と q とを置き換へて,

$$\left.\begin{aligned}(q+p)^n &= q^n + npq^{n-1} + \frac{n(n-1)}{1.2}p^2q^{n-2} + \frac{n(n-1)(n-2)}{1.2.3.}p^3q^{n-3} + \cdots\\ &\cdots\cdots\frac{n(n-1)\cdots\cdots(n-r+1)}{r!}p^rq^{n-r} + \cdots\cdots np^{n-1}q + p^n\\ &= q^n + {}_nC_1 pq^{n-1} + {}_nC_2 p^2q^{n-2} + \cdots {}_nC_r p^rq^{n-r} + \cdots {}_nC_{n-1}p^{n-1}q + p^n\end{aligned}\right\}(\text{XI. }2)$$

とすると一層都合がいい.

この式の第一項は 0 回成功する確率,第二項は一回成功する確率,第三項は 2 回成功する確率,第四項は $\cdots\cdots$, 而して r 回成功し $n-r$ 回不成功なる確率は第 $r+1$ 項(一般項)であつて,

$${}_nC_r p^r q^{n-r} = \frac{n(n-1)(n-2)\cdots\cdots(n-r+1)}{r!}p^rq^{n-r}$$
$$= \frac{n!}{r!(n-r)!}p^rq^{n-r}$$

であらはすことが出來る. (XI. 2) の右邊は成功不成功の全ての確率の和であるから,當然右邊は 1 でなければならぬ. 左邊は $(p+q)^n=1$ である. 又, n 個の試行中少くとも r 個其の事象の起る確率は,全確率の定理を用ひて,

$$p^n + {}_nC_{n-1}p^{n-1}q + \cdots\cdots {}_nC_r p^rq^{n-r},$$

又は $$1-\{q^n + {}_nC_1 q^{n-1}p + \cdots\cdots + {}_nC_{r-1}p^{r-1}q^{n-r+1}\}$$

で計算することが出來る.

例 3. 10個の骰子を投じ,その中三個だけが 1 の目の出る確率を求む.

10個の骰子を投じて,その 1 が出ることは互に獨立である. 又 1 が出る確率は何れも $\frac{1}{6}$, 1 が出ざる確率は $\frac{5}{6}$ であるから,之を上式にあてはめて,

$${}_{10}C_3 p^3 q^7 = \frac{10.9.8}{1.2.3}\cdot\frac{1}{6^3}\cdot\frac{5^7}{6^7} = 0{\cdot}155$$

例 4. 10個の骰子を投じ, 1 の目の出ることを成功とし,成功數 0, 1, 2, $\cdots\cdots$ 7 なる確率を計算せよ. 但し初めは $p=\frac{1}{6}$ とし,次に $p=0{\cdot}1625$ として計算せよ.

解. 第80表第二列及び第三列がそれである.

第二列は $p=\frac{1}{6}$ として成功の確率を計算したもの,第三列は $p=0{\cdot}1625$ として計算せる成功の確率を記入したものである. かく計算せられたる成功の確率が事實に一致するか否かを實

第 80 表

r (成功數)	先驗的確率 $_nC_rp^rq^{n-r}\left(p=\frac{1}{6}\right)$	先驗的確率 $_nC_rp^rq^{n-r}(p=0{\cdot}1625)$	實驗上の分布及確率 成功の度數分布	實驗上の分布及確率 經驗的確率
0	0·162	0·170	180	0·180
1	0·323	0·329	319	0·319
2	0·291	0·288	282	0·282
3	0·155	0·149	151	0·151
4	0·054	0·050	55	0·055
5	0·013	0·012	10	0·010
6	0·002	0·002	2	0·002
7	0·0002	0·0001	1	0·001
8—10	0·0000	0·0000	—	—
計	1·0002	1·0001	1,000	1·000

驗する目的で,著者は10個の骰子を投じ(1), 1 が出た骰子を成功として,成功個數を記錄し,1000投の成功度數分布表を作つた所,第四列の數となつた. この數より經驗的確率を計算したのが第五列であるが,これが可なりよく先驗的の成功確率(第二列又は第三列)に一致してゐる. 他の諸家も多くの實驗を行つて居るが,余の實驗と同じく,皆先驗的成功確率が經驗的の確率に可

(1) 獨立事象であるから10個の骰子を一度に投じてもいい. 又一個の骰子を十回投じて,之に代へても同じことである. しかし嚴格に論ずるならば,10個の骰子を一度に投じたのでは,各骰子の持つ個性が異ふから,同一の骰子を10回投じた方がいい. しかし上記の問題ではかくの如き試驗を更に1000回くりかへすのであるから,誤解を起し易い. 故に10個の骰子を1000回投ずると考へた方が良い. 且つ理論的には個性のない完全な骰子を考へるのであるから, 10個を投ずると云ふ風に說明した. 以後之に準ふ.

なりよく一致して居り; 唯僅かに一致せざるは,實驗を行つた骰子の個性が著しく偏して居るか,又は偶然の影響によるものであつて,投ずる回數の多くなればなる程,その一致は密接となつてくるのである.

獨立事象の成功の確率の分布が $(p+q)^n$ の展開の各項で計算することが出來ることは,かく實驗上でも明になつた.

又成功回數の分布は $N(p+q)^n$ の展開に於ける各項であらはすことが出來る. 此處に N は觀測總數である.

2. 二項式展開

n が正の整數なる時は, $(p+q)^n$ は, p, q の如何にかかはらず展開することが出來る. 今展開した各項を一瞥するも不要ではあるまい.

(i) $(p+q)^n$ の展開に於ける項數は $(n+1)$ であるから, n 個の獨立試行に於て,互に排反なる成功回數分布の項數は $(n+1)$ 個ある筈である. 例 4 に於ける如く,10個の骰子を投じた場合は, $n=10$ であるから項數は11である. 然るに,この例における實驗では, 8 個の項を記載したばかりで,他は確率があまり小であるから省略したのである. 若し又,この試驗を更に數萬回繰返して行ふ時は,成功數 9 なる場合や成功數10なる場合も,實驗的に出て來たのであらう.

(ii) 二項係數の總和は 2^n である.

上記展開式に於て, $p=q=\dfrac{1}{2}$ と置けば,

$$\left(\frac{1}{2}+\frac{1}{2}\right)^n=\left(\frac{1}{2}\right)^n\left\{1+n+\frac{n(n-1)}{1.2.}+\frac{n(n-1)(n-2)}{1.2.3}+\cdots\cdots\right.$$

$$+\cdots\frac{n(n-1)\cdots(n-r+1)}{r!}+\cdots+1\}$$

$$2^n=1+n+\frac{n(n-1)}{1.2}+\frac{n(n-1)(n-2)}{1.2.3}+\cdots\cdots$$

$$+\cdots+\frac{n(n-1)\cdots(n-r+1)}{r!}+\cdots+1$$

又は,　$2^n=1+{}_nC_1+{}_nC_2+\cdots+{}_nC_r+\cdots+{}_nC_{n-1}+1$

卽ち二項式展開に於ける係數の級數が得られる. 其の總和は 2^n である.

(iii) この係數は相稱であつて,第 r 項の係數と第 $(n-r)$ 項の係數とは相等しい. 卽ち,

$${}_nC_1={}_nC_{n-1}\quad {}_nC_2={}_nC_{n-2}\cdots\cdots {}_nC_r={}_nC_{n-r}\qquad \text{である.}$$

證明. 第 r 項の係數は ${}_nC_{r-1}$ 第 $n-r$ 項の係數は ${}_nC_{n-r+1}$ である. 卽ち,

$${}_nC_{r-1}=\frac{n!}{(r-1)!(n-r+1)!}=\frac{n!}{(n-r+1)!(r-1)!}={}_nC_{n-r+1}$$

(iv) $p=q=\frac{1}{2}$ なる時は,

$(p+q)^n$ の展開の各項は相稱である. 卽ち

$${}_nC_rp^rq^{n-r}={}_nC_{n-r}p^{n-r}q^r\quad \text{である.}$$

證明は容易である; 讀者試みられよ.

(v) $p=q=\frac{1}{2}$ とし, n を 2 より順次 3, 4, 5 $\cdots\cdots$ とすれば,その展開式は次の通りとなる.

$$1=\left(\frac{1}{2}+\frac{1}{2}\right)^2=\frac{1}{4}(1+2+1)$$

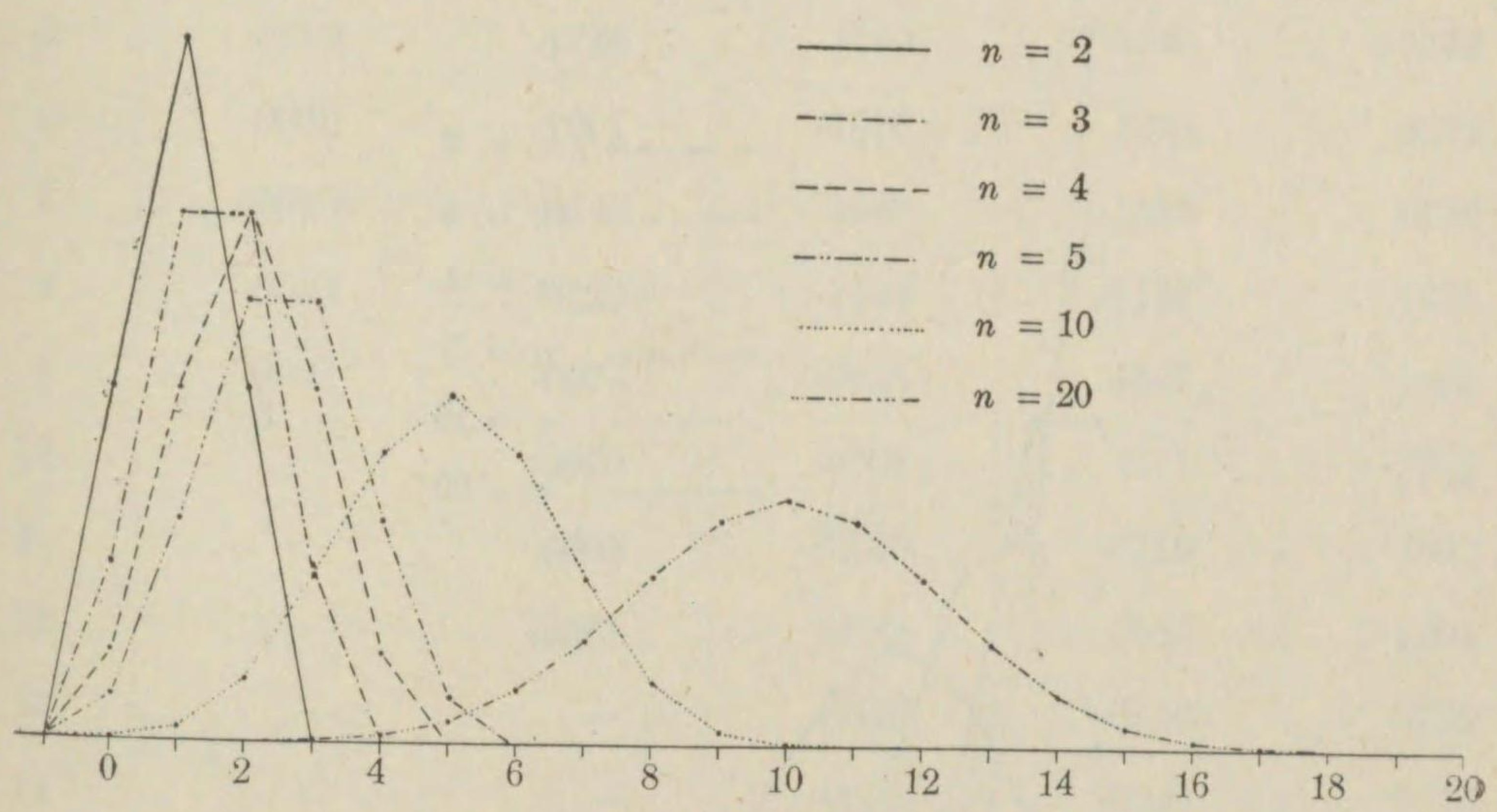
$$1=\left(\frac{1}{2}+\frac{1}{2}\right)^3=\frac{1}{8}(1+3+3+1)$$

$$1=\left(\frac{1}{2}+\frac{1}{2}\right)^4=\frac{1}{16}(1+4+6+4+1)$$

$$1=\left(\frac{1}{2}+\frac{1}{2}\right)^5=\frac{1}{32}(1+5+10+10+5+1)$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

括弧の中は二項級數の係數であつて,第四章に既に說明した數列が出來上る. 而して展開式における各項の和は卽ち右邊は勿論 1 に等しい. 次圖は n を 2, 3, 4, $\cdots\cdots$ とした時の各項の變化を圖示したものである. 橫軸は成功數,縱軸は確率を取つたものである. n の大なるに從ひ,成功數の分布範圍廣くなり,從て確率は小となる.



第 52 圖　$\left(\frac{1}{2}+\frac{1}{2}\right)^n$ の展開の各項を示すグラフ.

(vi) 次に p が q に等しくない時は,相稱でない. 卽ち

$${}_nC_rp^rq^{n-r}\neq{}_nC_{n-r}p^{n-r}q^r$$

何となれば, (iii) より　${}_nC_r={}_nC_{n-r}$

而して,　p^rq^{n-r} と $p^{n-r}q^r$ とは等しくないからである.

今 $p=\frac{1}{6}$ $q=\frac{5}{6}$ とし, n を種々變更して,各項を計算すると,次の通りとなる.

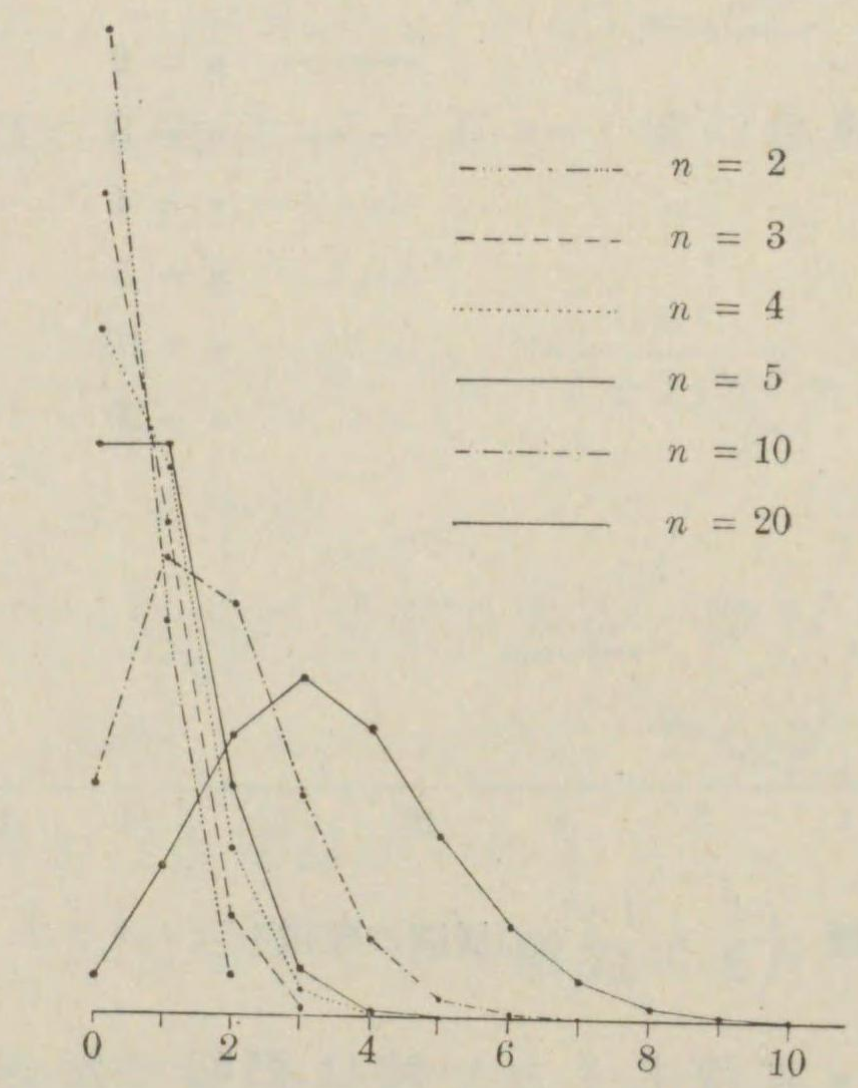
$$1=\left(\frac{5}{6}+\frac{1}{6}\right)^2=\frac{1}{36}(25+10+1),$$

$$1=\left(\frac{5}{6}+\frac{1}{6}\right)^3=\frac{1}{216}(125+75+15+1),$$

$$1=\left(\frac{5}{6}+\frac{1}{6}\right)^4=\frac{1}{1296}(625+500+150+20+1),$$

$$1=\left(\frac{5}{6}+\frac{1}{6}\right)^5=\frac{1}{7776}(3125+3125+1250+250+25+1),$$

$p=q$ の場合でも, n を漸次大にして來れば,この不相稱の程度も顯著でなくなる. 第53圖はこれを示す.



第 53 圖 $\left(\frac{5}{6}+\frac{1}{6}\right)^n$ の展開の各項を示すグラフ.

(vii) $(p+q)^{20}$ の右邊の各項を計算して,第81表とした. $p=0{\cdot}1$ より $0{\cdot}5$ まで順次にかへて見たのである. $p=0{\cdot}6$ の場合は

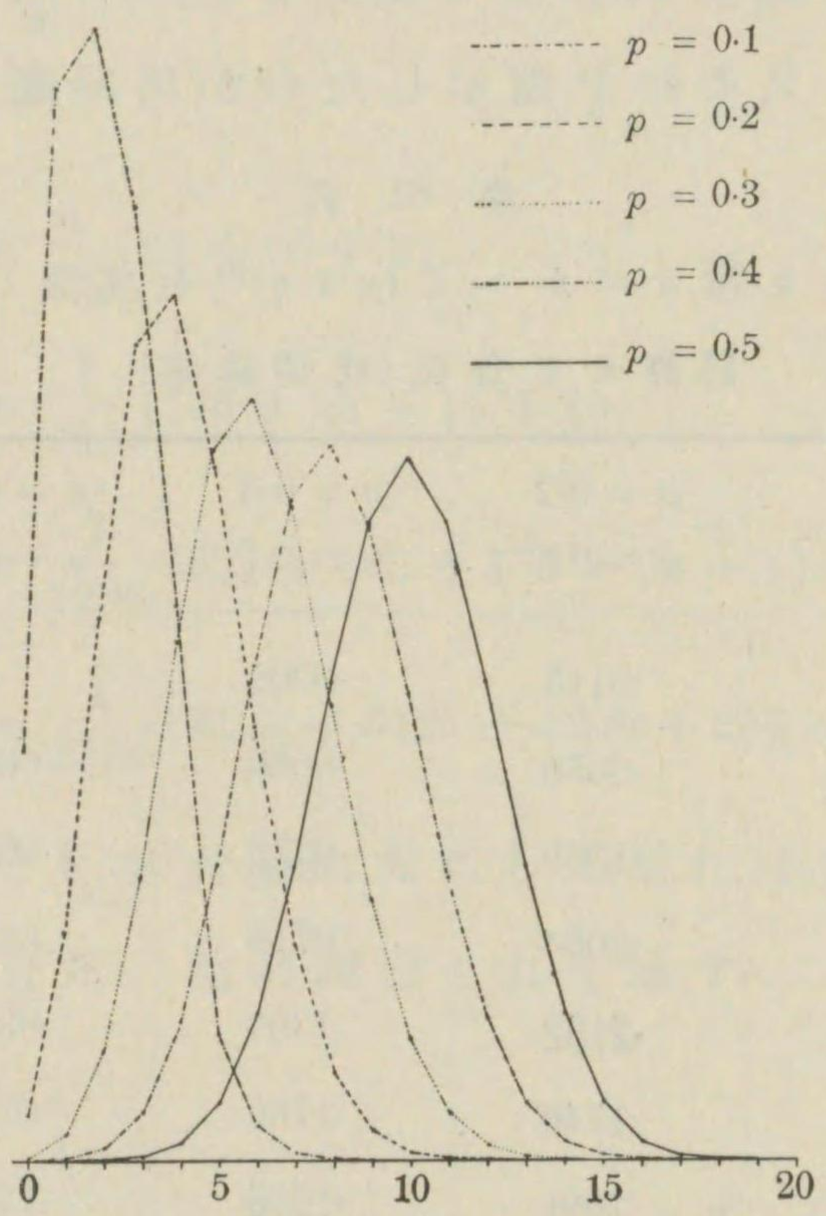
$p=0{\cdot}4$ の分布を逆にしたものであるから省略した. $p=0{\cdot}7$ 以上もその通り. 又これを圖示したのが第54圖である.

第 81 表

p を種々にかへて $(p+q)^{20}$ を展開し

計算せる各項(成功確率)

成功數	$p=0{\cdot}1$ $q=0{\cdot}9$	$p=0{\cdot}2$ $q=0{\cdot}8$	$p=0{\cdot}3$ $q=0{\cdot}7$	$p=0{\cdot}4$ $q=0{\cdot}6$	$p=0{\cdot}5$ $q=0{\cdot}5$
0	·1216	·0115	·0008	—	—
1	·2702	·0576	·0068	·0005	—
2	**·2852**	·1369	·0278	·0031	·0002
3	·1901	·2054	·0716	·0123	·0011
4	·0898	**·2182**	·1304	·0350	·0046
5	·0319	·1746	·1789	·0746	·0148
6	·0089	·1091	**·1916**	·1244	·0370
7	·0020	·0545	·1643	·1659	·0739
8	·0004	·0222	·1144	**·1797**	·1201
9	·0001	·0074	·0654	·1597	·1602
10	—	·0020	·0308	·1171	**·1762**
11	—	·0005	·0120	·0710	·1602
12	—	·0001	·0039	·0355	·1201
13	—	—	·0010	·0146	·0739
14	—	—	·0002	·0049	·0370
15	—	—	—	·0013	·0148
16	—	—	—	·0003	·0046
17	—	—	—	—	·0011
18	—	—	—	—	·0002
19	—	—	—	—	—
20	—	—	—	—	—



第 54 圖 $(p+q)^{20}$ に於て p を種々に變更し展開各項の變化を示すグラフ. 但し $q=1-p$ とす.

3. 最大確率

一般には,今後,生起(又は成功)の確率 p なる事象の, n 個の獨立試行なる語を用ふることとする. さて,成功の確率 p なる事象の n 個の獨立試行に於て,成功數幾個なることが最も確らしきか,即ち,最大の確率は成功數幾個なる時であるか,と云ふのが本條の問題である.

成功數 r なる確率は,(XI. 1) により,

$${}_nC_r q^{n-r}p^r = \frac{n(n-1)\cdots\cdots(n-r+1)}{1.2.\cdots\cdots r} q^{n-r}p^r$$

今,この項が最大項でありとすれば,

$$\frac{n(n-1)\cdots\cdots(n-r+2)}{(r-1)!} q^{n-r+1}p^{r-1} < \frac{n(n-1)\cdots\cdots(n-r+1)}{r!} q^{n-r}p^r$$

$$> \frac{n(n-1)\cdots\cdots(n-r)}{(r+1)!} q^{n-r-1}p^{r+1}$$

でなければならぬ. 此の二個の不等式を解けば,

$$q < \frac{n-r+1}{r} p \quad \cdots\cdots\cdots\cdots\cdots\cdots\text{[A]}$$

及び,

$$q > \frac{n-r}{r+1} p \quad \cdots\cdots\cdots\cdots\cdots\cdots\text{[B]}$$

となる. この式 [A] より,

$$rq < (n-r+1)p$$

となり; $q=1-p$ を代入すれば,

$$r < (n+1)p \quad \cdots\cdots\cdots\cdots\cdots\cdots\text{(a)}$$

が導かれる.

又, [B] より,

$$rq + q > (n-r)p$$

これに $q=1-p$ を代入すれば

$$r+1-p > np,$$

$$r > (n+1)p - 1, \quad \cdots\cdots\cdots\cdots\cdots\cdots\text{(b)}$$

(a) 及び (b) より,

$$(n+1)p - 1 < r < (n+1)p, \quad \cdots\cdots\cdots\text{(XI. 3. a)}$$

今, $(n+1)p = s$ とすれば,

$$s-1 < r < s, \quad \cdots\cdots\cdots\cdots\cdots\cdots\text{(XI. 3. b)}$$

即ち, s と $s-1$ との間にある自然數(正の整數) r を取り,成功數 r 個なる時最大の確率を有する. s もし整數なる時は,成功數 s なる時及び $s-1$ なるとき共に最大確率が有る.

例 5. 例 4 に於て,最大確率は何處にありやを計算せよ.

$$n=10,\quad p=\frac{1}{6},\quad q=\frac{5}{6},$$

$(n+1)p-1<r<(n+1)p,$

$\frac{5}{6}<r<\frac{11}{6}$, r は 1 でなければならぬ. 成功回數 1 なるとき,卽ち第二項が最大確率を有し,例 4 (第80表) に於ける計算また實驗にも一致する. 又, $p=0{\cdot}1625$ として計算するも同樣の結果となる. $r=1$ であるから最大確率は ${}_{10}C_1pq^9=0{\cdot}323$ $\left(p=\frac{1}{6}\right.$ としたる時$\left.\right)$ である.

例 6. 骰子を20個投じた時の最大確率は如何. 但し, $p=\frac{1}{6}$ とす.

$$s=(n+1)p=\frac{21}{6}=3{\cdot}5,$$

$$3{\cdot}5-1<r<3{\cdot}5,$$

r は整數なるが故に, $r=3$ なる時,最大確率を有すべきである. 而して, $r=3$ の時の確率は, ${}_{20}C_3p^3q^{17}$ であるから,

$${}_{20}C_3p^3q^{17}=\frac{20.\,19.\,18}{1.\,2.\,3}\cdot\frac{1}{6^3}\cdot\frac{5^{17}}{6^{17}}=0{\cdot}238,$$

$n=10$ なる時は,最大確率は $0{\cdot}323$ であつたが, $n=20$ となる時は, $0{\cdot}238$ である. かく p が一定なる時は, n が大となる時 r も亦大となるが,その r に相當する最大確率は n が大となるに從つて小となるものである.

4. 三個以上の排反事象の獨立試行

前條までは,生起すると生起せざると,唯二個の排反事象を繰返して試行したのである. 今,獨立事象, $E_1, E_2, \cdots E_m$ あり, E_1 の起る確率を p_1, E_2 の確率を $p_2, \cdots E_m$ の確率を p_m とす. この事象の n 個の試行中, E_1 が r_1 回, E_2 が r_2 回, $\cdots$, E_m が r_m 回起る確率は,

$$\frac{n!}{r_1!\,r_2!\,r_3!\cdots r_m!}p_1^{r_1}p_2^{r_2}\cdots p_m^{r_m}\quad\cdots\cdots\cdots\cdots(\mathrm{XI.}\ 4)$$

である. これは,多項式 $(p_1+p_2+p_3+\cdots p_m)^n$ の展開の一般項の公式である. 故に,このことを次の如く云ひかへることが出來る.

定理. $E_1, E_2, E_3, \cdots E_m$ の起る確率を $p_1, p_2, \cdots p_3, \cdots p_m$ とす. 今,この事象の n 個の獨立試行に於て, $E_1, E_2, E_3, \cdots E_m$ の起る確率は,すべて $(p_1+p_2+p_3+\cdots p_m)^n$ の展開式の各項で示すことが出來る. 卽ち,

$$(p_1+p_2+p_3+\cdots p_m)^n=p_1^n+p_2^n+\cdots p_m^n+np_1^{n-1}p_2+np_1^{n-1}p_3\cdots$$

この定理はあまり必要でないから證明は省略する.

例 7. n 個の骰子を投じ,1 が 2 個,2 が 3 個出る確率をとふ. 1 が 2 個,2 が 3 個なるを以て, 1, 2 以外の目は $n-5$ 個起る筈である. 故に,

$$P=\frac{n!}{2!\,3!\,(n-5)!}\left(\frac{1}{6}\right)^2\left(\frac{1}{6}\right)^3\left(\frac{4}{6}\right)^{n-5}$$

である.

5. 毎回の確率がことなる獨立試行

これまでは或事象の起る確率は常に一定であつたのであるが,本條に於ては,

第一回	の試行に於て事象の起る確率	p_1,	其餘事象の確率	q_1,
第二回	〃	p_2,	〃	q_2,
第三回	〃	p_3,	〃	q_3,
…………	…………	………,	…………	……,
第 n 回	〃	p_n,	〃	q_n

とす．然る時は，n 回の試行に於て，此事象が r 回起る確率は，

$$(p_1x+q_1)(p_2x+q_2)(p_3x+q_3)\cdots\cdots(p_nx+q_n)$$

なる式を展開して，x の降冪の順とし，

$$P_nx^n+P_{n-1}x^{n-1}+P_{n-2}x^{n-2}+\cdots\cdots P_1x+P_0$$

なる形としたる時，r 回だけ起る確率は x^r の係數 P_r に等し．この式に於て

$$p_1=p_2=p_3\cdots\cdots=p_n \text{ とし，} q_1=q_2=\cdots\cdots=q_n$$

とすれば，3 條に說きたる通りとなる．故にこの定理は獨立試行の最一般なる形とすることが出來る．しかし，前條と共に保險數學者には必要であらうが，生物統計學に緣が遠いから詳しくは說かない．

6. 變數の確率

確率の最も簡單なるものは，或事象が生起するか生起しないかの確率，或は屬性が存在するか存在しないか等の確率，即ち二個の排反事象の確率である．次には，骰子を投じた場合の如く，A, B, C, D, E, F 等の數個の面の何れが出るか，又花の色の白，赤，紫等の何れの色が出るかと云ふ樣に，n 個の排反事象の中，何れかの生起の確率であつた．しかるに，本章では成功數 0 なる確率，成功數 1 なる確率，2 なる確率又成功數 n なる確率を論じたのである．即ち，事象は 0 より n までの數を取りうるのである．かくの如き事象を，變數を以て表はしうる場合の確率を變數の確率と云ふ．次の如く定義す．

定義． 變數 X が種々の値を取りうる場合に，X の個々の値をとる確率を**變數の確率**と云ふ．

しかし，變數の確率と云ふも，決して，事象生起の確率と全く異つたものでは無い．例へば，銅貨 n 個を投ずる時は普通變數の確率として論ずるのであるが，一個の貨幣を投じた時でも，表の出ることを成功數 1 とし，裏の出ることを成功數 0 と考へれば，變數の確率として論ずることも出來る．又，骰子を投じ A, B, C, D, E, F の六つの面の何れか一つが出ると云ふ代りに，普通行はれて居る樣に，1, 2, 3, 4, 5, 6 の面の何れか一つが出るのであると考へるならば，此も變數の確率となるのである．1 なる面の出る確率と云ふも，1 なる値を取る確率と云ふも同一のことである．それのみならず，貨幣を投じ表が出れば 10 圓，裏が出れば 1 圓を受取るものと假定すれば表又は裏の出る確率は 10 圓，又は 1 圓を受取る確率となるを以て變數の確率となる．確率を論じうるものならば，何でも，變數の確率となしうる．

變數に二種ありて，一は實數値を飛び飛びに取る場合で，**不連續變數**と云ひ，例へば，0, 1, 2, ……, n と云ふ樣に，自然數のみを取る樣な場合，又，自然數ならずとも實數値を飛び飛びにとる場合である．反之，a と b との間の實數値を(有理數と無理數)ことごとく取りうる場合は，**連續變數**と云ふ．從て，變數の確率にも**不連續變數の確率**と**連續變數の確率**とがある．此の分類は變數の性質から行つたのであるが，確率の性質からすれば，**先驗的變數確率**と**經驗的變數確率**とに分類しうる．成功數の確率や其の他一般に先驗的確率を變數の確率に變更せるものは前者に屬し；花瓣の數や，一莢中の種子の數，身長頭指數等の度數分布表より近似値を計算するものは後者に屬する．

例 8. 油菊 (Chrysanthemum lavandulaefolium) の舌狀花の度數分布は,第一章第 1 表にも例示した所であるが,これより經驗的確率を計算し,且つ,圖示せよ. これは不連續變數の確率の例である.

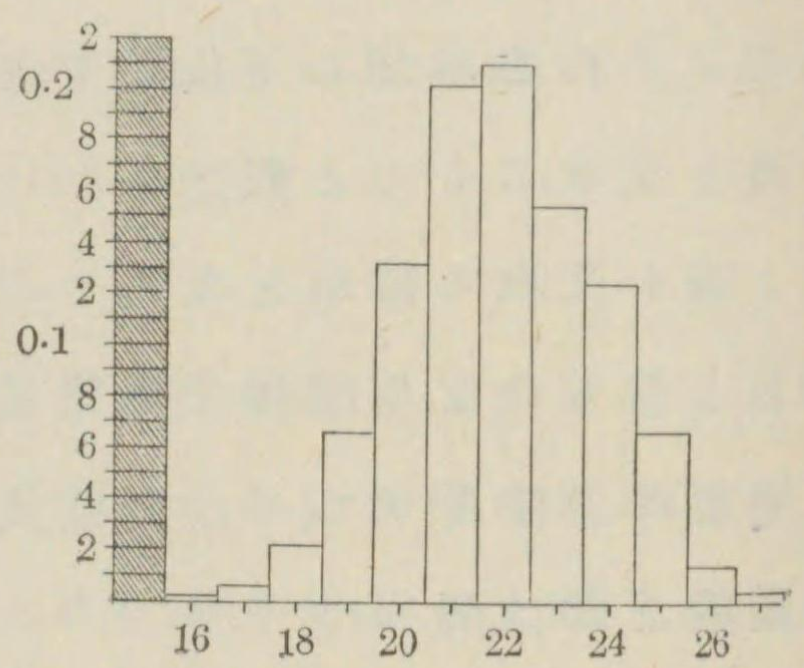
第 82 表

油菊の舌狀花の確率.

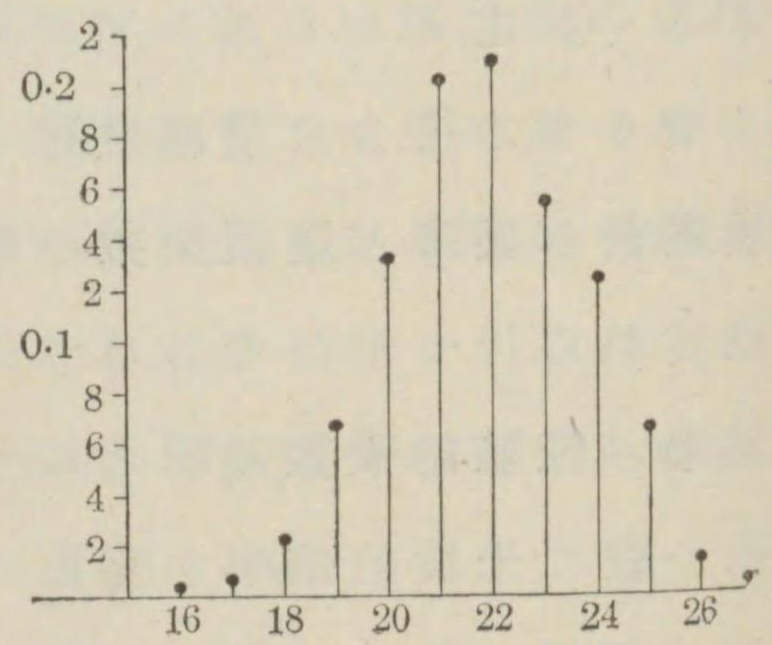
舌狀花數 X	度數 f	確率 $P(X)$	$XP(X)$
16	1	·002	·032
17	3	·006	·102
18	11	·022	·396
19	33	·066	1·254
20	66	·132	2·140
21	100	·201	4·221
22	104	·209	4·598
23	77	·154	3·542
24	62	·124	2·976
25	33	·066	1·650
26	7	·014	·364
27	2	·004	·108
計	499	1·000	21·883

圖示すれば第55圖甲或は乙となるが,乙は變數の不連續の狀態をよく表はしてゐる。第82表第4列については例11を見よ。

甲



乙



第 55 圖 油菊の舌狀花の確率分布をあらはすグラフ.

7. 期望值 (又數學的期望值)

定義 I. 不連續變數の確率に於て,變數と其の確率との積の總和を,其の變數の**期望值**又は**數學的期望值**と云ふ.

X を不連續變數,其の確率を $P(X)$ にてあらはし, $E(X)$ を期望值とすれば,

$$E(X) = \Sigma X.P(X) = X_1.P_1(X_1) + X_2.P_2(X_2) + X_3.P_3(X_3) + \cdots + X_n.P_n(X_n) \quad \cdots\cdots (\text{XI. } 5)$$

である. 期望值を又**確らしき值**と云ふことがある.

例 9. 骰子の目の數を變數とし,之を投じたる時の期望值を求む.

解. 骰子の目の數は 1, 2, 3, 4, 5, 6 である. 故に期望值 $E(X)$ は,

$$E(X) = \frac{1}{6}\times1+\frac{1}{6}\times2+\frac{1}{6}\times3+\frac{1}{6}\times4+\frac{1}{6}\times5+\frac{1}{6}\times6 = \frac{7}{2}$$

例 10. 骰子を投じ,1 の目が出た時は10圓を受取り,2 の目の時は15圓,3 の目の出た時は20圓,4 の目の出た時は25圓,5 の目の出た時は30圓,6 の目の出た時は35圓を受取ることと定められたり. 今,骰子を投じたる時,受取る金額の期望值を問ふ.

解. 1 の目は10圓と結び,2 の目は15圓と結び付き,3 以下もそれぞれの金額と結びついて居る. 故に例 9 とは異つて,10圓受取る確率,15圓受取る確率,‥‥等となり,且これ等の確率は何れも $\frac{1}{6}$ である. 故に,受取る金額の期望值 $E(X)$ は,

$$E(X) = 10\times\frac{1}{6}+15\times\frac{1}{6}+20\times\frac{1}{6}+25\times\frac{1}{6}+30\times\frac{1}{6}+35\times\frac{1}{6} = 22{\cdot}50\text{ 圓}$$

かくの如く變數及期望值が金額で表はされて居る時は,其の期望值を**期望金額**と云ふ. 期望金額は上記の如き賭を無限回

行ひたる時,受け取るべき金額の平均である.

例 11. 本章例 8 に於ける舌狀花數の確率に於て,其の期望値を求む.

解. 例 9,例 10 は不連續變數に於ける先天的確率であつたが,例 11 は不連續變數の後天的確率である. $XP(X)$ を求めて,第 82 表第四列に記入したが,期望値は第四列の總和 21·883 である.

この期望値は,第二章演習問題 1 に於ける算術的平均と同一であることに注意せよ.

定理. 不連續變數の期望値は,變數に對する觀測(又は試行)を無限に多く繰返したる時,其の變數が取る値の算術的平均の極限値に等しい(1).

觀測を無限に多く行ふことが出來ないから,普通の意味では次の如く云ひあらはすべきである.

不連續變數の期望値の近似値は,其の變數の算術的平均に等しい.

證明 I. 變數を X とし,X が取りうる總ての値を,

$$X_1,\ X_2,\ X_3,\ \cdots\cdot X_i,\ \cdots\cdot X_n$$

變數がこの値を取りたる時の觀測數を

$$N_1,\ N_2,\ \cdots\cdot N_i,\ \cdots\cdot N_n$$

とし,その確率を,

$$P(X_1),\ P(X_2),\ \cdots\cdot P(X_i),\ \cdots\cdot P(X_n)$$

とする時は,

(1) 連續變數でも此の定理が適用するが今は不連續變數につきて說く.

$$P(X_1) = \lim_{N\to\infty}\frac{N_1}{N},$$

$$\cdots\cdots\cdots\cdots\cdots$$

$$\cdots\cdots\cdots\cdots\cdots$$

$$P(X_i) = \lim_{N\to\infty}\frac{N_i}{N},$$

故に,期望値は,

$$E(X) = \sum_{i=1}^{n} X_i P(X_i) = \sum_{i=1}^{n} X_i \lim_{N\to\infty}\frac{N_i}{N}$$

N は各項共通なるが故に

$$= \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{n} X_i N_i$$

而して,$\frac{1}{N}\sum_{i=1}^{n} X_i N_i$ は變數の算術的平均である. 故に,期望値は觀測數を無限に多くしたる時の變數の算術的平均である.

N が無限大でない場合は次の如く證明する.

證明 II. N 大なる時は期望値の近似値 $E(X)$ は,

$$E(X) = X_1 P(X_1) + X_2 P(X_2) + \cdots\cdot + X_n P(X_n),$$

$$= X_1\frac{N_1}{N} + X_2\frac{N_2}{N} + \cdots\cdots\cdots\cdots + X_n\frac{N_n}{N},$$

$$= \frac{1}{N}\{X_1 N_1 + X_2 N_2 + \cdots\cdots\cdots\cdot + X_n N_n\},$$

$$= M_x$$

M_x は X の算術的平均である.

8. 獨立試行に於ける成功數の期望値 (又は成功數の算術的平均)

生起の確率 p なる事象があり,この事象の n 回の獨立試行を行ふ時,生起回數(成功回數)の確率は,公式 (XI. 2) により p 及び n

から計算することが出來る．又別に例4の如く實驗に由り,卽ちN回の試行により,經驗的にも確率の近似値を計算することが出來る(第80表參照)．故に,成功數の確率には二種,卽ち先驗的確率,及び經驗的確率がある．從て,期望値もこの先驗的確率より,或は經驗的確率よりも計算することが出來るが,後者は實際は數字の計算のみであり,第二章で說いたから,今前者より期望値を計算することを試みる．是には簡單な計算式がある．

定理．n個の獨立試行に於て,各事象の生起する(成功する)確率はpなりとす．然らば成功數の期望値$E(X)$はnpである．

$$E(X)=np \quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots (\text{XI. } 6)$$

證明：證明法は種々あるが,今理解し易い方法で行つて見る．

生起の確率をpとし,又$q=1-p$とすれば,n個の獨立試行に於ては,生起數が0, 1, 2, ‥‥nなる確率は$(q+p)^n$の展開,

$$(q+p)^n=q^n+nq^{n-1}p+\frac{n(n-1)}{1.2}q^{n-2}p^2+\cdots\cdot\frac{n!}{r!(n-r)!}q^{n-r}p^r+\cdots\cdot+p^n$$

の各項である．故に,期望値は此の各項にそれぞれ成功數0, 1, 2, ‥‥r, ‥‥nを乘じたる積の總和であるから,期望値を$E(X)$とすれば,

$$E(X)=nq^{n-1}p+n(n-1)q^{n-2}p^2+\frac{n(n-1)(n-2)}{2!}q^{n-3}p^3$$

$$+\cdots\cdot\frac{n!}{(r-1)!(n-r)!}q^{n-r}p^r+\cdots\cdot+np^n$$

$$=np\left\{q^{n-1}+(n-1)q^{n-2}p+\frac{(n-1)(n-2)}{2!}q^{n-3}p^2+\cdots\cdot+p^{n-1}\right\}$$

而して,括弧中は$(q+p)^{n-1}=1$なるを以て,

$$E(X)=np$$

卽ち本定理を證明したのである．この公式は必要なものであつて$E(X)$は又成功數の算術的平均と稱せられ屢用ひらる．

9. 函數の期望値及成功數の標準偏差

不連續函數Xの確率函數を$P(X)$,又Xの任意の函數を$\phi(X)$とすれば,

$$\Sigma\phi(X)P(X)$$

をXの函數$\phi(X)$の期望値と云ふ．

さて,Xの函數$\phi(X)$の如何なる形式が,生物統計學に必要であるかと云ふに,

$$\phi(X)=X^2$$

$$\phi(X)=X^3$$

$$\phi(X)=X^4$$

等であつて, Pearson の Moment (μ) の計算はこれから導き得るのである．本編では,就中一般的なX^2の期望値を解說する．

定理 I. 變數の平方の期望値を$E(X^2)$とすれば

$$\boldsymbol{E(X^2)=\sigma^2+M^2}$$

である．但し,σは標準偏差,Mは期望値卽ち算術的平均とす．

證明．變數の値, $X_1, X_2, \cdots\cdot X_n$,

又,其の値を取る度數を, $f_1, f_2, f_3, \cdots\cdot f_n$,

期望値をMとすれば,標準偏差の定義により,

$$\sigma^2=\frac{1}{N}\sum_{i=1}^{n}(X_i-M)^2f_i,$$

$$=\frac{1}{N}\sum_{i=1}^{n}X_i^2f_i-\frac{2M}{N}\sum_{i=1}^{n}X_if_i+\frac{1}{N}NM^2,$$

しかるに, $\dfrac{1}{N}\sum_{i=1}^{n} X_i f_i = M$ であるから,

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{n} X_i^2 f_i - M^2,$$

$$\sigma^2 + M^2 = \frac{1}{N}\sum_{i=1}^{n} X_i^2 f_i,$$

$$\sigma^2 + M^2 = \sum_{i=1}^{n}\left(X_i^2 \cdot \frac{f_i}{N}\right),$$

而して,觀測數を非常に大にした時は, $\dfrac{f_i}{N}$ は $P(X_i)$ となるを以て,この式の右邊は $E(X^2)$ となる. 即ち,

$$E(X^2) = \sigma^2 + M^2 \cdots\cdots\cdots\cdots\cdots\cdots (\mathrm{A})$$

定理 II. n 個の獨立試行に於て,各事象の成功の確率 p,不成功の確率 q,即ち $1-p$ とすれば,成功數の標準偏差 σ は先驗的に,

$$\sigma = \sqrt{npq} \cdots\cdots\cdots\cdots\cdots\cdots (\mathrm{XI.}\ 7)$$

である.

証明. 前定理により

$$\sigma^2 + M^2 = \Sigma X^2 P(X)$$

而して先驗的確率 $P(X)$ は,(XI. 2) により, $(q+p)^n$ の展開の右邊の各項,

$$(q+p)^n = q^n + nq^{n-1}p + \frac{n(n-1)}{1.2}q^{n-2}p^2$$

$$+\cdots\cdots \frac{n(n-1)\cdots\cdots(n-r+1)}{r!}q^{n-r}p^r + \cdots\cdots + p^n$$

である. 而して X^2 は $0^2, 1^2, 2^2, \cdots\cdots n^2$ の値を順に取るのであるから,これを展開せる各項に乘じて,

$$\sigma^2+M^2 = nq^{n-1}p + 2n(n-1)q^{n-2}p^2 + \frac{3n(n-1)(n-2)}{2!}q^{n-3}p^3 + \cdots\cdots$$

$$+\frac{rn(n-1)\cdots\cdots(n-r+1)}{(r-1)!}q^{n-r}p^r + \cdots\cdots + n^2p^2,$$

$$= np\left\{q^{n-1} + 2(n-1)q^{n-2}p + 3\frac{(n-1)(n-2)}{2!}q^{n-3}p^2 + \cdots\cdots\right.$$

$$\left. + r\frac{(n-1)(n-2)\cdots(n-r+1)}{(r-1)!}q^{n-r}p^{r-1} + \cdots\cdots np^{n-1}\right\},$$

$$= np\left\{q^{n-1} + (n-1)q^{n-2}p + \frac{(n-1)(n-2)}{2}q^{n-3}p^2 + \cdots\cdots\right.$$

$$+\frac{(n-1)(n-2)\cdots\cdots(n-r+1)}{(r-1)!}q^{n-r}p^{r-1} + \cdots\cdots p^{n-1}$$

$$+(n-1)q^{n-2}p + 2\frac{(n-1)(n-2)}{2!}q^{n-3}p^2 + \cdots\cdots$$

$$\left. + \frac{(r-1)(n-1)(n-2)\cdots\cdots(n-r+1)}{(r-1)!}q^{n-r}p^{r-1} + \cdots\cdots (n-1)p^{n-1}\right\},$$

$$= np\left\{(q+p)^{n-1} + p(n-1)\cdot\left(q^{n-2} + (n-2)q^{n-3}p + \cdots\cdots\right.\right.$$

$$\left.\left. + \frac{(n-1)(n-2)\cdots\cdots(n-r+1)}{(r-2)}q^{n-r}p^{r-2} + \cdots\cdots + p^{n-2}\right)\right\},$$

$$= np\left\{1 + (n-1)p(q+p)^{n-2}\right\},$$

$$= np(1+np-p),$$

しかるに, $1-p=q$ であるから,

$$\sigma^2 + M^2 = np(q+np),$$

$$= npq + (np)^2,$$

しかるに,(XI. 6) に由り, $np=M$ であるから,

$$\sigma^2 = npq,$$

$$\sigma = \sqrt{npq}.$$

例 12. 全く同一に作られたる10個の骰子を投じ,其の成功數

を記錄したりとし,算術的平均M及び標準偏差σを求む. $p=\frac{1}{6}$とせよ.

解. $p=\frac{1}{6}$, $q=\frac{5}{6}$, $n=10$,

故に, $$M=np=\frac{10}{6}=1\cdot667,$$

$$\sigma=\sqrt{npq}=1\cdot178,$$

これは定理により先驗的に計算したのであるが,果して正しきかを檢する. 第80表の第二列よりM及びσを計算し,其の一致せるを見よ. 尙ほ又,10個の骰子の1000投によりて得たる度數分布(經驗的の分布)は第四列に出て居る. この表より經驗的の平均及び標準偏差M'及びσ'を計算するに

$$M'=1\cdot625$$

$$\sigma'=1\cdot189$$

理論的に先驗的確率より計算せるM及σに比し多少の相違はあるが,槪してよく一致して居る.

10. pが小なる時の標準偏差

定理. p甚小なる時は,標準偏差は近似的に

$$\sigma=\sqrt{np}=\sqrt{M}\cdots\cdots\cdots\cdots\cdots\cdots\cdots\text{(XI. 8)}$$

である.

證明. 何となれば,前條の公式(XI. 7)により,

$$\sigma=\sqrt{npq}=\sqrt{np(1-p)}.$$

$$=\sqrt{np-np^2},$$

p甚しく小なる時は, npに比しnp^2は甚しく小なるが故に,之を無視することが出來る.

故に, $$\sigma\fallingdotseq\sqrt{np},$$

即ち $$\sigma\fallingdotseq\sqrt{M}.\cdots\cdots\cdots\cdots\cdots\cdots\cdots\text{(XI. 8)}$$

例 13. プロイセンの軍隊で,20年間に馬に蹴られて死亡したものの總數122,一の聯隊で年々其の爲に死亡する者の平均は0·61であつたと云ふ. これより死亡者數の標準偏差を求む.

解. 一事象の起る先驗的の確率pも,試行回數nも明でない. 唯,$M=0\cdot61$だけが判つて居る. 又,p甚小なることも判つて居るから,本條の公式を當て篏めて

$$\sigma=\sqrt{M}=\sqrt{0\cdot61}$$

即ち, $$\sigma=0\cdot781.$$

Bortkewitsch は200個の聯隊に於ける度數分布を與へて居る. 第83表がそれで,これによると$M=0\cdot61$又上式によらず第三章の方法で表よりσを計算すれば,

$$\sigma^2=0\cdot6079,$$

$$\sigma=0\cdot779$$

となり,公式(XI. 8)を用ひたのと同一結果となる. 即ち,p甚小なる時は,分布の形式如何にかかはらず,p, qを用ひずして算術的平均より標準偏差を計算することが出來る.

第 83 表

死亡者數	度數
0	109
1	65
2	22
3	3
4	1
計	200

演習問題

1. 骰子を投じて,4, 5, 6の目の出た時を成功とし,1, 2, 3の出た時を不成功と定む. 今,12個の骰子を同時に投じたる時成功0, 1, 2, …12,なる確率を問ふ.

解. 骰子の成功の確率は(數學的定義),

$$p = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2},$$

$$q = 1 - p = \frac{1}{2},$$

故に (XI. 2) により，

$$\left(\frac{1}{2} + \frac{1}{2}\right)^{12}$$

を展開したる各項は,成功確率の分布を表はすのである. 次表の第二列がそれである.

第三列,12個の骰子を 4096 回投じ,成功數の統計を取り度數分布としたるもので, W. F. R. Weldon の實驗である. (Encycl. Brit. 11th. ed. Vol. 22. p. 394.)

第 84 表

成功數	先驗的確率	12 の骰子を 4096 回投じたる成功の度數分布	經驗的確率
0	1/4096	—	—
1	12/4096	7	7/4096
2	66/4096	60	60/4096
3	220/4096	198	198/4096
4	495/4096	430	430/4096
5	792/4096	731	731/4096
6	924/4096	948	948/4096
7	792/4096	847	847/4096
8	495/4096	536	536/4096
9	220/4096	257	257/4096
10	66/4096	71	71/4096
11	12/4096	11	11/4096
12	1/4096	—	—
計	4096/4096 = 1	4096	1

これより經驗的確率を計算すれば,第四列となる. これが第二列によく一致せるを見よ. 但し,後章に說く樣に骰子に多少の偏重があつたのであらう,分布は僅かに一方に偏して居る. 尙,第二列と第四列との精密な比較が必要な時は, χ^2-試驗の法による.

2. $n = 20$ とし, $p = 0{\cdot}1$, $p = 0{\cdot}2$, $p = 0{\cdot}3$, $p = 0{\cdot}4$, $p = 0{\cdot}5$ なる場合に於ける $(p+q)^{20}$ を展開せる各項は,第81表に示したが,今 $p = 0{\cdot}6$, $p = 0{\cdot}7$, $p = 0{\cdot}8$, $p = 0{\cdot}9$ なる時,其の展開の各項を示せ. (計算を行はず第81表を利用し書き下せ.)

3. 12個の骰子を投じて, 6 の目の出たものを成功とし,他の目の出たるを不成功とし, 4096 回投じたる時,その最も確らしき成功の分布如何.

解. 12個の骰子を投ずるものであるから,成功 0 より 12 まであるわけである. その各の確率が計算できる. 題意は 4096 回投じた時に可能なる成功分布を計算するのであるから,各の確率に 4096 を乘じたものが,その分布である. これを空欄に記入せよ. 表の第三欄は Weldon 氏の實驗になる數である,これを計算せる度數分布と比較せよ.

第 85 表

成功數	理論的成功分布 (確らしき度數分布)	(Weldon の實驗せる結果) 成功の實驗的度數分布
0		447
1		1145
2		1181
3		796
4		380
5		115
6		24
7		7
8		1
9—12		—
計		4096

4. 三個の骰子を投じ, 5 又は 6 の目の出た骰子を成功とした. 今,これを648回投じたる時の理論的成功度數分布を計算し第86表の空欄に記入せよ.

$p = \frac{1}{3}$, $q = \frac{2}{3}$, $n = 3$ として計算す. 尙, Yule が實驗の結果次表の

數を得た.

第 86 表

成功數	理論的分布	經驗的分布
0		179
1		298
2		141
3		30
計		648

5. 梅 (Prunus mume) に高麗と名づける紅梅の一品種がある. 其の花瓣の數は 5 から 8 までであつて, 600 花の觀測の結果は次の表の通りである. これより花瓣數 X が 5, 6, 7, 8 なる値をとる確率の近似値を計算し,次でその期望値を計算せよ.

第 87 表

花瓣數 X	度數 f	確率 p	$XP(x)$
5	520	0·867	4·355
6	58	0·097	0·582
7	20	0·033	0·231
8	2	0·003	0·024
計	600	1·000	$\Sigma XP(X) = 5{\cdot}172$

解. は表によりて明である. ここに期望値 5·172 は花瓣の平均に一致することに注意せよ.

6. 下記の命題を證明せよ.

(i) 二項式の展開に於て奇數番目の係數の總和は偶數番目の係數の總和に等しい. 即ち係數を

$$C_0 \quad C_1 \quad C_2 \cdots\cdots C_n$$

とすれば

$$C_0 + C_2 + C_4 + \cdots = C_1 + C_3 + C_5 + \cdots$$

(ii) $p = q = \frac{1}{2}$ とする時は, $(p+q)^n$ の展開に於て奇數番目の項の總和は偶數番目の項の總和に等しい.

即ち, $$p^n + \frac{n(n-1)}{2!}p^{n-2}q^2 + \frac{n(n-1)(n-2)(n-3)}{4!}p^{n-4}q^4 + \cdots$$
$$= np^{n-1}q + \frac{n(n-1)(n-2)}{3!}p^{n-3}q^3 + \cdots$$

7. Mendel の法則に從ふ單性雜種を,自家生殖によりて數代繁殖せしむる時,雜種は漸次分離して,固定形質を生じ,雜種の割合は漸次減少する. 其の分離比を確率論を用ひて計算せよ.

解. 一般に生殖子(卵子及び精子)には遺傳子を一個づつそなへ,接合子(卵子及び精子の接合にて生成せる生物體)には遺傳子を二重にそなへて居ると假定す. 即ち,[A]につきて云へば,生殖子は A を備へたるものと, a を備へたるものとあり; 又,接合子には AA を具へたるものと, aa を具へたるものと, Aa を具へたるものとあり,この最後のものは雜種である. 又 AA よりは生殖子 A を生じ, aa よりは生殖子 a を生じ,雜種 Aa よりは生殖子 A と a を生ず. 而して雜種より A を生ずる確率 $\frac{1}{2}$, a を生ずる確率 $\frac{1}{2}$ なりと假定す. さて, AA 及び aa なる遺傳子を具へたる生物の交配によりて, (i) F_1 (雜種第一代)では, Aa なる雜種のみを生ず. 何となれば, AA なる生物よりは生殖子 A のみを生じ, aa なる生物よりは生殖子 a のみを生ず,其の交配では, A と a との結合より他の組合せが無き爲である.

(ii) F_2 (雜種第二代): F_2 以下では自家生殖を行ふものと假定する. F_1 の Aa なる接合子よりは,生殖子(卵子でも精子でも) A と a とを生ずるる確率何れも $\frac{1}{2}$ である.

$$♂\begin{cases}(A)_p = \frac{1}{2}\\ (a)_p = \frac{1}{2}\end{cases} \qquad ♀\begin{cases}(A)_p = \frac{1}{2}\\ (a)_p = \frac{1}{2}\end{cases}$$

然るに,自家生殖であるから,此等の生殖子より接合子 AA, Aa, aa を生ずる確率は次の式によりて表はすことが出來る.

$$\left(\frac{1}{2}A + \frac{1}{2}a\right)^2 = \frac{1}{4}AA + \frac{1}{2}Aa + \frac{1}{4}aa.$$

係數は何れも確率である．即ち AA を生ずる確率 $\frac{1}{4}$, aa を生ずる確率 $\frac{1}{4}$, Aa を生ずる確率 $\frac{1}{2}$ であることを示す．

(iii) F_3 (雜種第三代)：F_2 の接合子 AA, Aa, aa より生ずる生殖子の確率．AA よりは A のみを生じ(その確率 1), 又 aa よりは a のみを生ずるが(其の確率 1), Aa よりは A と a とを生じ其の確率は何れも $\frac{1}{2}$ である．故に自家生殖では次の通りとなる．

(甲) F_2 の AA から, F_3 に AA を生ずる確率は, F_2 の AA の確率 $\frac{1}{4}$ とこれより AA を生ずる確率 1 との複確率によりて,(從屬事象)

$$\frac{1}{4}\times 1=\frac{1}{4}$$

(乙) F_2 の Aa からは $\frac{1}{4}AA$, $\frac{2}{4}Aa$, $\frac{1}{4}aa$ を生ずること (ii) に說いた通りであるが F_2 では Aa の確率は $\frac{1}{2}$ であるから,複確率の定理により,

Aa より AA の生ずる確率, $\frac{1}{2}\times\frac{1}{4}=\frac{1}{8}$

Aa より Aa を生ずる確率, $\frac{1}{2}\times\frac{1}{2}=\frac{1}{4}$

Aa より aa を生ずる確率, $\frac{1}{2}\times\frac{1}{4}=\frac{1}{8}$

(丙) F_2 の aa よりは, F_3 に aa のみ生ずるが,これも甲と同じく複確率の定理により, $\frac{1}{4}\times 1=\frac{1}{4}$

そこで,甲乙丙全體としては,

F_3 に AA を生ずる確率は, $\frac{1}{4}+\frac{1}{8}=\frac{3}{8}$

F_3 に Aa を生ずる確率は, $\frac{1}{4}\quad=\frac{2}{8}$

F_3 に aa を生ずる確率は, $\frac{1}{4}+\frac{1}{8}=\frac{3}{8}$

(iv) F_4 も同樣に考へる時は,

F_4 に AA を生ずる確率は, $\frac{1}{4}+\frac{1}{8}+\frac{1}{16}=\frac{7}{16}$

F_4 に Aa を生ずる確率は, $\frac{1}{4}\times\frac{1}{2}=\frac{2}{16}$

F_4 に aa を生ずる確率は, $\frac{1}{4}+\frac{1}{8}+\frac{1}{16}=\frac{7}{16}$

(v) 一般に, F_n 代まで自家生殖を續ける時, F_n 代に於て AA, Aa, aa を生ずる確率は,全確率の定理を應用して,

(甲) AA を生ずる確率 $\frac{1}{2^2}+\frac{1}{2^3}+\frac{1}{2^4}+\cdots\cdots+\frac{1}{2^n}=\frac{2^{n-1}-1}{2^n}$

(乙) aa を生ずる確率は同じく, $\frac{2^{n-1}-1}{2^n}$

(丙) 雜種 Aa を生ずる確率は, $1-2\cdot\frac{2^{n-1}-1}{2^n}=\frac{2}{2^n}$

一般に雜種を生ずる確率は n 大となるに從ひ著しく小となる．

8. 二屬性につき F_1 に雜種, $AaBb$ を作りこれより自家生殖をつづくる時, F_n 代にて雜種の分離は如何？

但し, $[A]$ と $[B]$ とは全く獨立なりと假定す．

解. (i) F_1 に於ては $AaBb$ のみを生ず．

(ii) F_2 に於て $[A]$ のみを考ふるときは AA, Aa, aa を生ずる確率は $\frac{1}{4}AA$, $\frac{1}{2}Aa$, $\frac{1}{4}aa$ で表はすことが出來た．$[B]$ につきても同樣である．而して $[A]$ と $[B]$ とはあらゆる組合せを作るのであり,又假定により $[A]$ と $[B]$ とは獨立事象であるから, $AABB$, $AaBB$ 等を生ずる確率は,次の式の右邊で表はすことが出來る．

$$\left(\frac{1}{4}AA+\frac{1}{2}Aa+\frac{1}{4}aa\right)\left(\frac{1}{4}BB+\frac{1}{2}Bb+\frac{1}{4}bb\right)$$

$$=\frac{1}{16}AABB+\frac{1}{16}AAbb+\frac{1}{16}aaBB+\frac{1}{16}aabb$$

$$+\frac{2}{16}AABb+\frac{2}{16}AaBB+\frac{2}{16}Aabb+\frac{2}{16}aaBb$$

$$+\frac{4}{16}AaBb$$

四個の文字の組合せはそれぞれ遺傳子の組合せであり,その係數は其の組合せの起る確率である．獨立事象であるから斯く簡單に計算出來るのである．支配の法則のために,

AB の形質をあらはす者は, $AABB$, $AABb$, $AaBB$, $AaBb$ であつて其の確率は確率加法の定理により, $\frac{9}{16}$

Ab が表はれるものは,

$AAbb$ と $Aabb$ と　であつて,確率は加法により $\frac{3}{16}$

aB が表はれるものは,

$aaBB,\ aaBb$　で確率の和は $\frac{3}{16}$

ab が表面に表はれるもの,

$aabb$　のみで,確率は $\frac{1}{16}$

である割合から云へば, $(AB):(Ab):(aB):(ab)=9:3:3:1$

これが F_2 における兩性雜種の Mendel 比である.

(iii) F_n: 若し $[A]$ のみを考へて見れば, AA, Aa, aa を生ずる確率は

$$\frac{2^{n-1}-1}{2^n}AA,\quad \frac{2}{2^n}Aa,\quad \frac{2^{n-1}-1}{2^n}aa$$

で表はすことが出來た.

$[B]$ につきても同樣である. 且つ $[A]$ と $[B]$ とは獨立であるから, F_n 代に於て $AABB$ $AABb$ 等を生ずる確率は,次式で求めることが出來る.

$$\left(\frac{2^{n-1}-1}{2^n}AA+\frac{2}{2^n}Aa+\frac{2^{n-1}-1}{2^n}aa\right)\left(\frac{2^{n-1}-1}{2^n}BB+\frac{2}{2^n}Bb+\frac{2^{n-1}-1}{2^n}bb\right)$$

$$=\left(\frac{2^{n-1}-1}{2^n}\right)^2(AABB+AAbb+aaBB+aabb)$$

$$+\left(\frac{1}{2^{n-1}}\times\frac{2^{n-1}-1}{2^n}\right)(AABb+AaBB+Aabb+aaBb)$$

$$+\frac{1}{2^{2n-2}}AaBb$$

而して,うはべに表はれる形質のみを考ふるならば,

$$(AB)_p=(AABB)_p+(AABb)_p+(AaBB)_p+(AaBb)_p=\frac{2^{2n-2}+2^n+1}{2^{2n}}$$

$$(Ab)_p=(AAbb)_p+(Aabb)_p=\frac{2^{2n-2}-1}{2^{2n}},$$

$$(aB)_p=(aaBB)_p+(aaBb)_p=\frac{2^{2n-2}-1}{2^{2n}},$$

$$(ab)_p=(aabb)_p=\frac{2^{2n-2}-2^n+1}{2^{2n}}.$$

9. 三屬性 $[A]$, $[B]$, $[C]$ につきて Mendel 比を計算せよ.

解. 雜種第二代 (F_2) に於て, $[A]$ なる屬性のみを考ふれば, Mendel 比は問題 7 の如く $\left(\frac{1}{2}A+\frac{1}{2}a\right)^2$ の展開式の各項で表はす事が出來る.

若し, $[A]$, $[B]$, $[C]$ を考ふる時,各屬性が互に獨立であるから, 問題 8 の如く計算しうる.

$$\left(\frac{1}{4}AA+\frac{1}{2}Aa+\frac{1}{4}aa\right)\left(\frac{1}{4}BB+\frac{1}{2}Bb+\frac{1}{4}bb\right)\left(\frac{1}{4}CC+\frac{1}{2}Cc+\frac{1}{4}cc\right)$$

$$=\frac{1}{64}AABBCC+\frac{1}{64}AABBcc+\frac{1}{64}AAbbCC+\frac{1}{64}aaBBCC$$

$$+\frac{1}{64}AAbbcc+\frac{1}{64}aaBBcc+\frac{1}{64}aabbCC+\frac{1}{64}aabbcc$$

$$+\frac{2}{64}AaBBCC+\frac{2}{64}AaBBcc+\frac{2}{64}AabbCC+\frac{2}{64}Aabbcc$$

$$+\frac{2}{64}AABbCC+\frac{2}{64}AABbcc+\frac{2}{64}aaBbCC+\frac{2}{64}aaBbcc$$

$$+\frac{2}{64}AABBCc+\frac{2}{64}AAbbCc+\frac{2}{64}aaBBCc+\frac{2}{64}aabbCc$$

$$+\frac{4}{64}AaBbCC+\frac{4}{64}AaBbcc+\frac{4}{64}AaBBCc+\frac{4}{64}AabbCc$$

$$+\frac{4}{64}AABbCc+\frac{4}{64}aaBbCc$$

$$+\frac{8}{64}AaBbCc.$$

全體で,文字の組合せの數は 27 個あり. これは,遺傳子の組合せの數であるが, Mendel の支配の法則で,うはべに表はれる形質を見るならば次の通となる.

$$(ABC)_p=(AABBCC)_p+(AABBCc)_p+(AABbCC)_p+(AaBBCC)_p$$

$$+(AABbCc)_p+(AaBBCc)_p+(AaBbCC)_p+(AaBbCc)_p$$

$$=\frac{27}{64}$$

$$(ABc)_p=(AABBcc)_p+(AABbcc)_p+(AaBBcc)_p+(AaBbcc)_p=\frac{9}{64}$$

$$(AbC)_p=(AAbbCC)_p+(AAbbCc)_p+(AabbCC)_p+(AabbCc)_p=\frac{9}{64}$$

$$(aBC)_p=(aaBBCC)_p+(aaBBCc)_p+(aaBbCC)_p+(aaBbCc)_p=\frac{9}{64}$$

$$(Abc)_p=(AAbbcc)_p+(Aabbcc)_p=\frac{3}{64}$$

$$(aBc)_p = (aaBBcc)_p + (aaBbcc)_p = \frac{3}{64}$$

$$(abC)_p = (aabbCC)_p + (aabbCc)_p = \frac{3}{64}$$

$$(abc)_p = (aabbcc)_p = \frac{1}{64}$$

全部についての比は,

$$\frac{27}{64} : \frac{9}{64} : \frac{9}{64} : \frac{9}{64} : \frac{3}{64} : \frac{3}{64} : \frac{3}{64} : \frac{1}{64} = 27:9:9:9:3:3:3:1$$

(ii) F_n 代に於てこれら形質の組合せは次の式で計算することが出來る.

$$\left(\frac{2^{n-1}-1}{2^n}AA + \frac{2}{2^n}Aa + \frac{2^{n-1}-1}{2^n}aa\right)\left(\frac{2^{n-1}-1}{2^n}BB + \frac{2}{2^n}Bb + \frac{2^{n-1}-1}{2^n}bb\right)$$
$$\left(\frac{2^{n-1}-1}{2^n}CC + \frac{2}{2^n}Cc + \frac{2^{n-1}-1}{2^n}cc\right)$$
$$= \left(\frac{2^{n-1}-1}{2^n}\right)^3\{AABBCC + AABBcc + AAbbCC + aaBBCC$$
$$+ AAbbcc + aaBBcc + aabbCC + aabbcc\}$$
$$+ \frac{2}{2^n}\cdot\left(\frac{2^{n-1}-1}{2^n}\right)^2\{AaBBCC + AaBBcc + AabbCC + Aabbcc$$
$$+ BbAACC + BbAAcc + BbaaCC + Bbaacc$$
$$+ CcAABB + CcAAbb + CcaaBB + Ccaabb\}$$
$$+ \left(\frac{2}{2^n}\right)^2\left(\frac{2^{n-1}-1}{2^n}\right)\{AaBbCC + AaBbcc + AaCcBB + AaCcbb$$
$$+ BbCcAA + BbCcaa\}$$
$$+ \left(\frac{2}{2^n}\right)^3 AaBbCc.$$

第十二章

正常確率函數と Laplace 及び Bernoulli の定理

1. 獨立事象の重複試行に於て,成功數の確率は $(p+q)^n$ の展開式の各項で表示出來た. n が大でない時は,此の式は充分使用に堪へるのであるが,若し n が大となる時,展開の項數は甚多くなり,例へば $n=100$ なる時は,項數は 101 となり; これをことごとく計算することは勿論,其の一部分を計算することも容易でない. 若し,展開の各項に代はる他の函數が見出されれば,以後の研究に便利である. 本章はこの函數を見出し,且つ,その性質を研究せんとするものである.

2. 最大確率の近似値

前章に於ては,成功確率 p なる n 個の獨立試行に於て,成功數 r が如何なる價を取りたる時,確率が最大なりやを論じた. 卽ち r が $(n+1)p$ と, $(n+1)p-1$ との間にある整數値を取りたる時,その確率が最大である. 若し, $(n+1)p$ が整數である時は,最大確率をとる r が 2 個ありて, $(n+1)p$ と $(n+1)p-1$ とが共に其れである.

又 r が決定すれば,その確率 P は,

$$P = {}_nC_r p^r q^{n-r} = \frac{n!}{r!(n-r)!}p^r q^{n-r}$$

で計算することが出來た. さて n 甚だ大なる時は,この r, 卽ち $(n+1)p$ と $(n+1)p-1$ との間にある整數に代ふるに,この二値

中間の一値

$$r \fallingdotseq np \quad \cdots\cdots\cdots\cdots\cdots\cdots (\text{XII. 1. a})$$

を以てするも大なる差支がない.

従て,　$n-r \fallingdotseq n-np,$

即ち,　$n-r \fallingdotseq nq \quad \cdots\cdots\cdots\cdots\cdots\cdots (\text{XII. 1. b})$

となる. これを最大確率の式に代入し,且つ Stirling[1] の式を適用する時は,

$$P=\frac{n!}{(n-r)!\,r!}p^rq^{n-r}=\frac{n!}{(np)!\,(nq)!}p^{np}q^{nq},$$

$$=\frac{n^ne^{-n}\sqrt{2n\pi}\,p^{np}q^{nq}}{(np)^{np}e^{-np}\sqrt{2np\pi}(nq)^{nq}e^{-nq}\sqrt{2nq\pi}}$$

$$=\frac{n^ne^{-n}\sqrt{2n\pi}\,p^{np}q^{n}}{n^np^{np}q^{nq}e^{-n}.2n\pi\sqrt{pq}}$$

$$=\frac{1}{\sqrt{2n\pi pq}}$$

となる.

このPをあらはすに y_0 を以てす.

$$\left.\begin{array}{l} y_0=\dfrac{1}{\sqrt{2n\pi pq}} \\ \text{又}\quad \sqrt{npq}=\sigma \quad \text{であるから,之を置換して} \\ y_0=\dfrac{1}{\sqrt{2\pi}.\sigma}. \end{array}\right\}\cdots(\text{XII. 2})$$

この公式に於て,n 個の事象のことごとくに p は共通でなければならぬ. 又たとひ,n 無限大なりと雖も,p 又は q の何れかが甚だ小であつて,np,nq が甚しく大とならざる時は,この式は適用できない (Poisson の法則參照).

(1) 註 Stirling の公式,

$$n!=n^n.e^{-n}\sqrt{2n\pi}$$

につきては,附錄を見よ.

又曰く,(np), (nq) は共に必ずしも整數でない故に,$(np)!$, $(nq)!$ は意義のないことになる. これにはガンマ函數を知らなければならないが,此處では (np) に最も近い整數の階乘, (nq) に最も近い整數の階乘と解して差支ない.

例 1. 骰子を 360 回投じたる時, 1 の目の出るのが何回なることが最確らしきか? 又その時の確率を計算せよ.

解. 骰子に於て 1 の目の出る確率は, $p=\dfrac{1}{6}$ とする. これを 360 回投ずるに,最大確率の項 r は np である.

$$r=np=360\times\frac{1}{6}=60$$

即,1 の目の出ること,60 回なることが最確らしい. 而して,1 の目が 60 回出る確率,即ち $r=60$ なる時の確率は,(XII. 2)に由り,

$$y_0=\frac{1}{\sqrt{2\times360\times\frac{1}{6}\times\frac{5}{6}\cdot\pi}}=\frac{1}{\sqrt{100\pi}}=0{\cdot}0056,$$

$$\text{答}\left\{\begin{array}{l} r=60, \\ P_{(n,r)}=0{\cdot}056, \end{array}\right.$$

尙,n が大なる時は,n 個の獨立試行に於て,成功數 r 回なる確率は,r の如何なる値に對しても小となる. $n=360$ なる時は,最大確率でも尙 0·056 である.

3. 正常確率函數の誘導

定理. 生起の確率が p なる事象に對して,互に獨立なる n 個の試行をなすとき,其の事象が $r=np+X$ 回[1] 生起する確率, y

(1) ここに $r=np+X$ と置きたることに注意せよ. r は 0 より n までの整數値を取る變數であるが,今 np を原點として計つた變數 X に變更したのである. $X=0$ と置けば $r=np$ 即ち最大確率の項となる. X は正或は負である.

又,np が必ずしも整數でないから,X も亦必ずしも整數でないが,不連續變數である.

は近似的に,

$$y \fallingdotseq \frac{1}{\sqrt{2\pi npq}} e^{-\frac{X^2}{2npq}}, \quad \cdots\cdots\cdots\cdots \text{(XII. 3. a)}$$

で表はすことが出來る. 但し n, np, nq は共に甚だ大なるものとす[1].

証明. $P_{(n,r)} = \dfrac{n!}{r!(n-r)!} p^r q^{n-r}$

とし, n, np, nq 甚だ大なるを以て, Stirling の公式を適用して,

$$P_{(n,r)} = \sqrt{\frac{2\pi n}{2\pi r 2\pi(n-r)}} \cdot \frac{n^n e^{-n} p^r q^{n-r}}{r^r e^{-r}(n-r)^{n-r} e^{-(n-r)}}$$

$$= \sqrt{\frac{n}{2\pi r.(n-r)}} \cdot n^n \left(\frac{p}{r}\right)^r \left(\frac{q}{n-r}\right)^{n-r},$$

$$\therefore \quad P_{(n,r)} = \sqrt{\frac{n}{2\pi r(n-r)}} \left(\frac{np}{r}\right)^r \left(\frac{nq}{n-r}\right)^{n-r} \cdots\cdots\cdots\cdots[\text{A}]$$

今,

$$\left.\begin{aligned} U &= \sqrt{\frac{n}{2\pi r(n-r)}} \\ V &= \left(\frac{np}{r}\right)^r \left(\frac{nq}{n-r}\right)^{n-r} \end{aligned}\right\} \cdots\cdots\cdots\cdots[\text{B}]$$

とすれば, $\qquad P_{(n,r)} = U \cdot V$

となる. 又, $r = np + X$ とすれば,

$$n - r = n - np - X = nq - X$$

なるを以て,

$$U = \sqrt{\frac{n}{2\pi(np+X)(nq-X)}} = \sqrt{\frac{1}{2\pi npq\left(1+\dfrac{X}{np}\right)\left(1-\dfrac{X}{nq}\right)}},$$

[1] 厳格に云へば, $n \to \infty$ なる極限に於て,この公式が成立するのであるが,例2,例3に示す通り, n やや大なれば用ひうるのである.

$$= \frac{1}{\sqrt{2\pi npq}} \left(1+\frac{X}{np}\right)^{-\frac{1}{2}} \left(1-\frac{X}{nq}\right)^{-\frac{1}{2}},$$

$\left(1+\dfrac{X}{np}\right)^{-\frac{1}{2}}$ 及び $\left(1-\dfrac{X}{np}\right)^{-\frac{1}{2}}$ を展開すれば,附錄數學公式により (公. 40)

$$U = \frac{1}{\sqrt{2\pi npq}}\left(1 - \frac{X}{2np} + \frac{1.3}{2.4}\left(\frac{X}{np}\right)^2 \cdots\right)\left(1 + \frac{1}{2nq} + \frac{1.3}{2.4}\left(\frac{X}{nq}\right)^2 \cdots\cdots\right),$$

$$= \frac{1}{\sqrt{2\pi npq}}\left(1 + \frac{X}{2nq} - \frac{X}{2np} + \varepsilon\right),$$

ε は $\dfrac{X}{np}$ 及び $\dfrac{X}{nq}$ に關し二乘以上の冪の和であつて, n 甚だ大なる時はこれを省略することが出來る. 故に,近似値は

$$U \fallingdotseq \frac{1}{\sqrt{2\pi npq}}\left(1 + \frac{(p-q)X}{2npq}\right); \cdots\cdots\cdots\cdots[\text{C}]$$

又, $p = q$ なる時,若くは $p \neq q$ であつても n, 甚しく大なる時は,更に省略して,

$$U \fallingdotseq \frac{1}{\sqrt{2\pi npq}} \quad \cdots\cdots\cdots\cdots\cdots\cdots[\text{D}]$$

とすることが出來る.

次に, $\quad V = \left(\dfrac{np}{r}\right)^r \left(\dfrac{nq}{n-r}\right)^{n-r} = \left(\dfrac{r}{np}\right)^{-r} \left(\dfrac{n-r}{nq}\right)^{-(n-r)},$

即ち, $\quad V = \left(\dfrac{np+X}{np}\right)^{-(np+X)} \left(\dfrac{nq-X}{nq}\right)^{-(nq-X)}$

$$= \left(1+\frac{X}{np}\right)^{-(np+X)} \left(1-\frac{X}{nq}\right)^{-(nq-X)},$$

$$\log V = -(np+X)\log\left(1+\frac{X}{np}\right) - (nq-X)\log\left(1-\frac{X}{nq}\right)$$

而して,右邊の對數は $\dfrac{X}{np}$ の冪級數に展開することが出來る (附錄數學公式により).

$$\log V = -(np+X)\left(\frac{X}{np} - \frac{X^2}{2(np)^2} + \frac{X^3}{3(np)^3}\cdots\cdots\right)$$

$$-(nq-X)\left(-\frac{X}{nq} - \frac{X^2}{2(nq)^2} - \frac{X^3}{3(np)^3}\cdots\cdots\right),$$

$$= -X + \frac{X^2}{2np} - \frac{X^3}{3(np)^2}\cdots\cdots - \frac{X^2}{np} + \frac{X^3}{2(np)^2} - \frac{X^4}{3(np)^3}\cdots\cdots$$

$$+X + \frac{X^2}{2nq} + \frac{X^3}{3(nq)^2}\cdots\cdots - \frac{X^2}{nq} - \frac{X^3}{2(nq)^2}\cdots\cdots,$$

$$= \frac{-X^2}{2np} - \frac{X^2}{2nq} + \rho_p + \rho_q.$$

但し, ρ_p は $\frac{X}{np}$ に關し二乘以上の冪の和, ρ_q は $\frac{X}{nq}$ に關し二乘以上の冪の和であつて, np, nq が甚だ大なる時はこれを省略することが出來るから,

$$\log V \fallingdotseq \frac{-X^2}{2npq};$$

されば, $$V = e^{-\frac{X^2}{2npq}}.\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots[E]$$

故に, $[D]$ 及び $[E]$ より

$$P_{(n,r)} = U\cdot V = \frac{1}{\sqrt{2\pi npq}}e^{-\frac{X^2}{2npq}}.$$

$P_{(n,r)}$ を表はすに y を以てする時は,

$$y = \frac{1}{\sqrt{2\pi npq}}e^{-\frac{X^2}{2npq}},\cdots\cdots\cdots\cdots\cdots(\text{XII. 3. a})$$

或は,詳しく

$$y = \frac{1}{\sqrt{2\pi npq}}e^{-\frac{X^2}{2npq}}\left(1 + \frac{(p-q)X}{2npq}\right).\quad\cdots\cdots(\text{XII. 4})$$

(3. a) は最屢用ひられ,正常確率函數と名づけらる. (3. a) 式に代ふるに種々の形があるが,何れも同一のものを稍ことなりたる形式で表示したにすぎない. 例へば,

$\sqrt{npq} = \sigma$ であるから,

$$y = \frac{1}{\sqrt{2\pi}.\sigma}e^{-\frac{X^2}{2\sigma^2}}.\cdots\cdots\cdots\cdots\cdots(\text{XII. 3. b})$$

$\frac{1}{\sqrt{2}\,\sigma} = h$ とすれば

$$y = \frac{h}{\sqrt{\pi}}e^{-h^2x^2},\cdots\cdots\cdots\cdots\cdots\cdots(\text{XII. 3. c})$$

又 $\frac{X}{\sigma} = \xi$ と置く時は,

$$y = \frac{1}{\sqrt{2\pi}\,\sigma}e^{-\frac{\xi^2}{2}}\quad\cdots\cdots\cdots\cdots\cdots(\text{XII. 3. d})$$

等何れも用ひられることがある.

(3) の四式又は (4) は $(p+q)^n$ の展開の各項に代へることが出來るが,この公式の誘導に當り Stirling の公式を用ひたり,又其の他二三ヶ所で $\frac{1}{n^2}$ を含む項を切り棄てたりした. 故に,以上の公式は $n\to\infty$ なる極限に於てのみ正しいのであるが,然らずとも,n が大でありさへすれば兎に角使用に堪へる. ことに $p=q$ なる時は,n が50以下でも可なりよく一致することは,例 2 に示す通りである. 又 $p\neq q$ で, n 甚しく大ならざる時は, (4) の式を使用すべきであるが,例 3 に示す通り $n=48$ でも尚未だ多少の誤差がある. 若し(XII. 3)式を用ふれば誤差が一層甚しい故に, p と q との差が斯く大なる時は,(XII. 3)又は(XII. 4)を使用せんが爲には, n が尚一層大となるを必要とする.

例 2. (XII. 3)が $(p+q)^n$ の各項に代用することが出來るか否かを檢べて見よう. 今 $n=48$ とし, $p=\frac{1}{2}$, $q=\frac{1}{2}$ として $(p+q)^n$ を展開し,其の各項と, $y = \frac{1}{\sqrt{2\pi}\,\sigma}e^{-\frac{X^2}{2\sigma^2}}$ より計算せる各項とを比較する(第89表).

$$\sigma = \sqrt{npq} = \sqrt{12} = 3\cdot464,$$

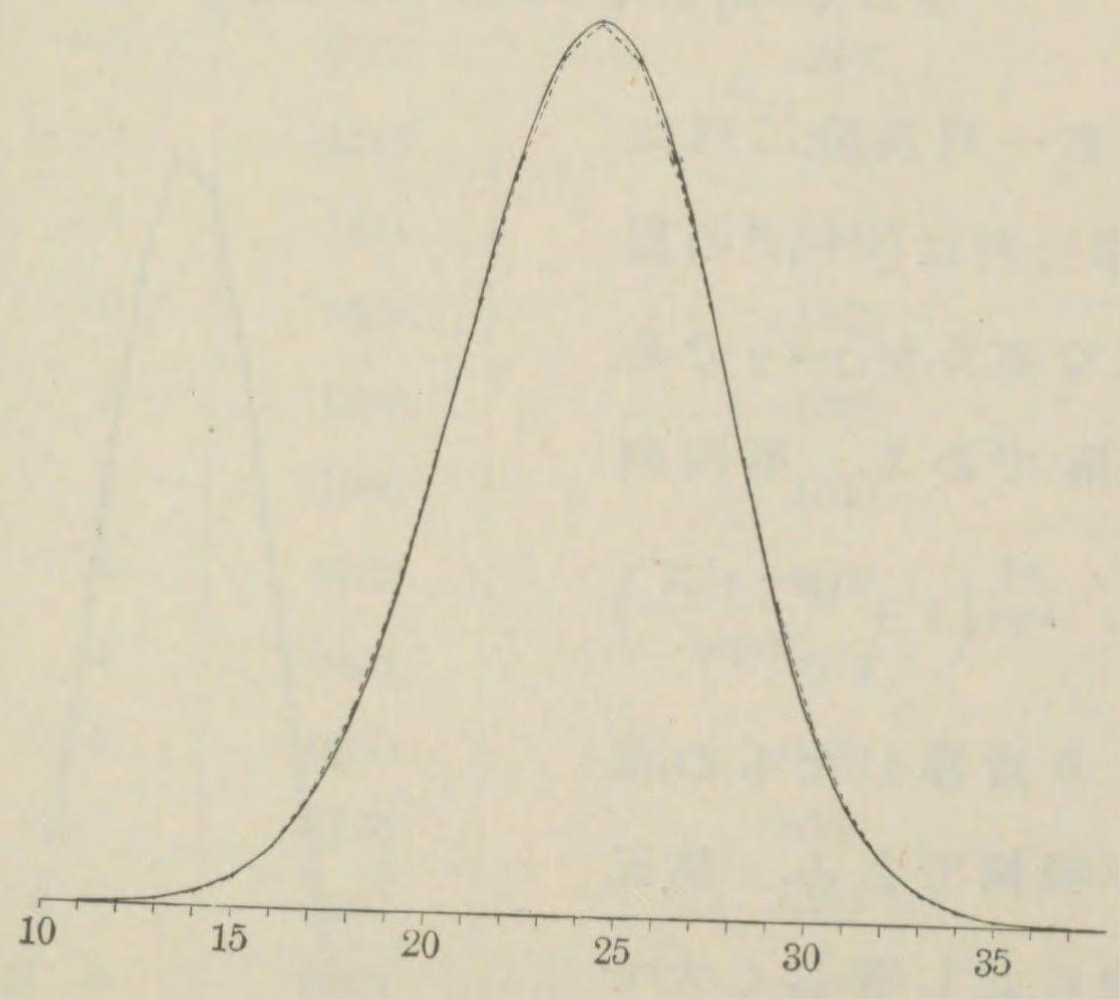
第 89 表

$\left(\frac{1}{2}+\frac{1}{2}\right)^{48}$ の展開の諸項と函數 $\frac{1}{\sqrt{2\pi npq}}e^{-\frac{X^2}{2npq}}$ より計算せる數値との比較を示す表.

r	$X=r-np$	${}_nC_rp^rq^{n-r}$	$\frac{1}{\sqrt{2\pi npq}}e^{-\frac{X^2}{2npq}}$	Δ
0—10	$-\infty$—-14	0·00003	0·00005	+0·00002
11	−13	·00008	·00010	+ ·00002
12	−12	·00025	·00029	+ ·00004
13	−11	·00069	·00075	+ ·00006
14	−10	·00171	·00179	+ ·00008
15	− 9	·00388	·00396	+ ·00008
16	− 8	·00801	·00800	− ·00001
17	− 7	·01508	·01495	− ·00013
18	− 6	·02597	·02570	− ·00027
19	− 5	·04101	·04060	− ·00041
20	− 4	·05946	·05913	− ·00033
21	− 3	·07928	·07915	− ·00013
22	− 2	·09729	·09749	+ ·00020
23	− 1	·10998	·11046	+ ·00048
24	0	·11456	·11516	+ ·00060
25	+ 1	·10998	·11046	+ ·00048
26	+ 2	·09729	·09749	+ ·00020
27	+ 3	·07928	·07915	− ·00013
28	+ 4	·05946	·05913	− ·00033
29	+ 5	·04101	·04060	− ·00041
30	+ 6	·02597	·02570	− ·00027
31	+ 7	·01508	·01495	− ·00013
32	+ 8	·00801	·00800	− ·00001
33	+ 9	·00388	·00396	+ ·00008
34	+10	·00171	·00179	+ ·00008
35	+11	·00069	·00075	+ ·00006
36	+12	·00025	·00029	+ ·00004
37	+13	·00008	·00010	+ ·00002
38—X	+14—$+\infty$	·00003	·00005	+ ·00002
計		1·00000	1·00000	

又,最大確率は $r'=np=24$ の項にあり,第89表は以下の計算を示す. 第一列は r を記入し,第二列には $X=r-np$ より計算せる X を記入したのである. 表の第三列は ${}_nC_rp^rq^{n-r}$ より各項の近似値を計算し(附表VI参照),第四列は $\frac{1}{\sqrt{2\pi npq}}e^{-\frac{X^2}{2npq}}$ を計算したものである(附表II参照). この二列を比較すれば,分布の中央即ち np に近き所は誤差稍大であるが,兩翼に近づけば誤差 $\varDelta$ は小となる. 第五列が誤差である.

第56圖はこの二つの數列を圖示比較したもので,實線は $\frac{1}{\sqrt{2\pi npq}}e^{-\frac{X^2}{2npq}}$ を表はし,虛線で連絡せる點は ${}_nC_rp^rq^{n-r}$ の各項に相當するものである.



第 56 圖 第 89 表をグラフとせるもの.

この二つがよく一致し n 大なる時は, $(p+q)^n$ の展開の各項に代ふるに, $\frac{1}{\sqrt{2\pi npq}}e^{-\frac{X^2}{2npq}}$ を以てすることが出來るを示してゐる. 尚 r が24の時最大確率を有し, r が10以下なる時,及び38

以上なる時は,共に,確率甚しく小となるが為に計算するに及ばないのである.

例 3. $n=48,\ p=\dfrac{1}{6},\ q=\dfrac{5}{6}$ とし,(XII. 4) 式にあてはめて確率函数を計算し,これを ${}_nC_r p^r q^{n-r}$ の各項と比較せよ.

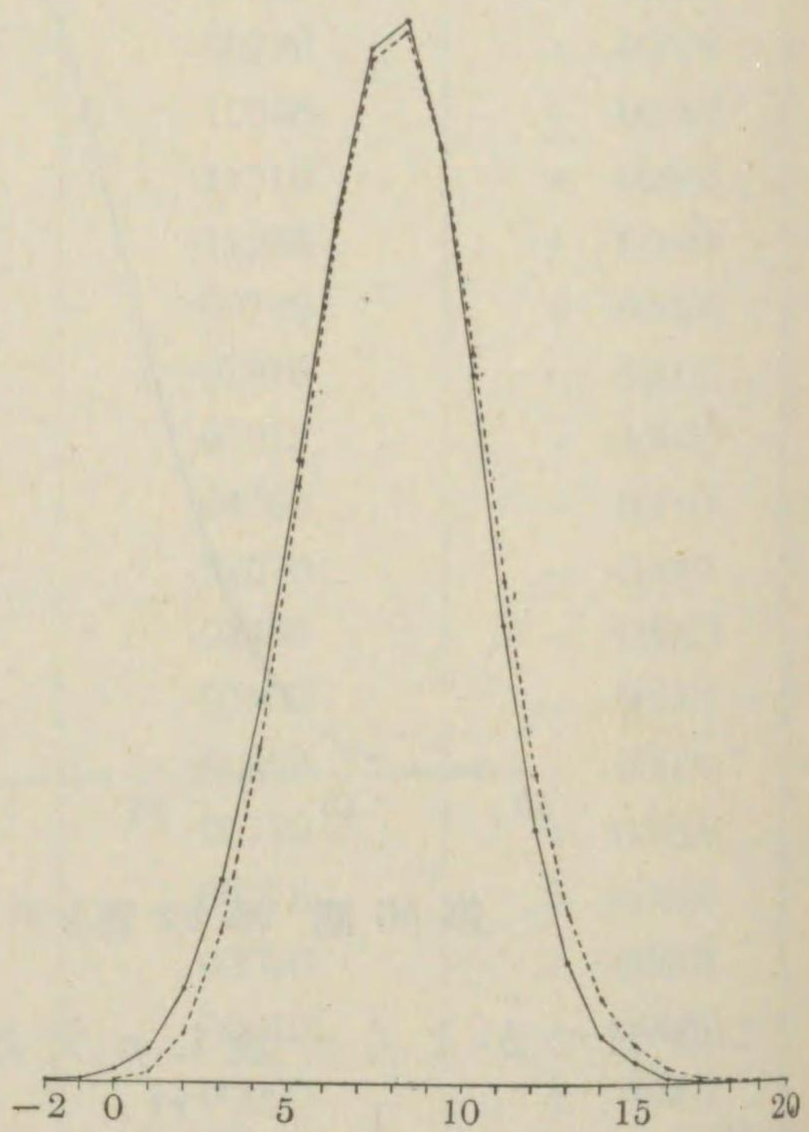
解. 最大確率の項 r' は,

$$r'=np=48\times\frac{1}{6}=8,$$

即ち,成功数 8 なる場合が最大確率を有し,その確率は正しくは 0·1529 である. また,(XII. 2) 式によつて計算すれば 0·1545 となる.

$$\sigma=\sqrt{npq}=\sqrt{6{\cdot}6667}=2{\cdot}582$$

さて,表の第一列及第二列は例 2 の通り,第三列は $(p+q)^n$ の展開式の各項であるが,$p\neq q$ であるから不相称である. 第四列は $\dfrac{1}{\sqrt{2\pi npq}}e^{-\frac{X^2}{2npq}}\left(1+\dfrac{(p-q)X}{2npq}\right)$ の展開式より計算したもの,従てこれも不相称である. 第五列 $\varDelta$ は前表に比し著しく大である. しかし其れでも,可なりよく二数列が一致する. 第 57 圖は第三列及び第四列を圖示比較したものである.



第 57 圖　第 90 表をグラフとせるもの

第 90 表

$\left(\dfrac{5}{6}+\dfrac{1}{6}\right)^{48}$ 展開の諸項と函数 $\dfrac{1}{\sqrt{2\pi npq}}e^{-\frac{X^2}{2npq}}\left(1+\dfrac{(p-q)X^2}{2npq}\right)$ との比較.

r	X $(r-np)$	${}_nC_r p^r q^{n-r}$	$\frac{1}{\sqrt{2\pi npq}}e^{-\frac{X^2}{2npq}}\left(1+\frac{(p-q)X}{2npq}\right)$	Δ
	$-\infty$—-10		·0002	+0·0002
	−9		·0005	+ ·0005
0	−8	·0002	·0020	+ ·0018
1	−7	·0015	·0050	+ ·0035
2	−6	·0071	·0138	+ ·0067
3	−5	·0219	·0296	+ ·0077
4	−4	·0489	·0558	+ ·0067
5	−3	·0867	·0905	+ ·0038
6	−2	·1242	·1258	+ ·0016
7	−1	·1491	·1505	+ ·0014
8	0	·1529	·1545	+ ·0016
9	1	·1359	·1362	+ ·0003
10	2	·1060	·1030	− ·0030
11	3	·0732	·0669	− ·0063
12	4	·0452	·0373	− ·0079
13	5	·0250	·0178	− ·0072
14	6	·0125	·0069	− ·0056
15	7	·0057	·0028	− ·0029
16	8	·0023	·0006	− ·0017
17	9	·0009	·0002	− ·0007
18	10	·0003	·0000	− ·0003
19	11	·0001	·0000	− ·0001
		1·0000	1·0000	

注意1. 確率を表はすには以上の二式を用ひるのであるが,若し度數分布表を作らんとならば,(3)又は(4)式の右邊にN(觀測總數)を乘ずればいいのである. 例へば,

$$Y' = \frac{N}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}} \quad \cdots\cdots\cdots(4.\ e)$$

の如くする. Y'は度數である.

注意2. $(p+q)^n$の右邊の各項は$n+1$個あるのみで,各項は不連續である. しかるに,(3)式及び(4)式は,共に連續の函數であることに,讀者は不審せらるるであらう. 函數形は連續だから,若し變數Xが連續ならば,(3)式及び(4)式のYは連續に相違ない. しかし$X=r-np$に於て,npは常數であるから,Xはrと同じく不連續變數である. 故に,(3)式及(4)式のXはとびとびに,$n+1$個の價をとるのみであるから,從てYも亦$n+1$個の値を取るのみで,不連續である. 函數形は連續だが實際Yは連續でない. 又Xは必ずしも整數でない.

4. 函數 $Y=\frac{1}{\sqrt{2\pi}.\sigma}e^{-\frac{X^2}{2\sigma^2}}$ の性質

(i) Yは常に正である. 其の右邊は決して負となることがないからである. 又,元來確率は負となるべきものでないことは明である.

(ii) 正常確率函數, $Y=\frac{1}{\sqrt{2\pi}.\sigma q}e^{-\frac{X^2}{2\sigma^2}}$ は偶函數である. 何となれば,上式Xに代ふるに$-X$を以てするも,Yの値に變りはない. 即ち,曲線はY軸に對して對稱である(第25圖其の他).

(iii) Yは$X=0$なる時,極大値をとる.

(iv) $X=0$よりXの絕對値を增大するに從ひて,Yは漸次小となり,$|X|=\infty$に於て,Yは0に收斂す.

(v) 曲線, $Y=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{X^2}{2\sigma^2}}$ は$X=\pm\sigma$に於て變曲點を有す(問題2).

(vi) これを圖示すれば第25圖の如き曲線となる.

5. Laplace の定理

前條の公式,

$$Y = \frac{1}{\sqrt{2\pi npq}} e^{-\frac{x^2}{2npq}}$$

はn, p, qが判明せる時,生起數rなる確率(又は生起數の偏差Xなる確率)を求むる方程式であつた. 然るに,n甚だ大なる時は,Yは甚だ小となり,Yを一つづつ計算することは,事實上無意義となることが多い. 例へば,銀貨を10,000回投じて,表が5,000回出る確率は甚だ小であつて,實際表が5,000回出ることは,殆ど豫期することが出來ない. かくの如き時は,表が5,000回乃至5,010回出る確率,或は5,010—5,020回出る確率と云ふ風に,變數を階級に分類し,變數が階級內の値を取る確率を求めた方が意義がある. Laplaceの定理はこれに關したものである.

定理. n回の獨立試行中,事象の生起の回數がs回乃至t回の範圍にある確率は,$n\to\infty$なる極限に於て,

$$\left.\begin{aligned} &\frac{1}{\sqrt{2\pi}}\int_{\alpha_1}^{\alpha_2} e^{-\frac{\xi^2}{2}}d\xi, \\ &\text{に等しい. 但し,} \\ &\sigma=\sqrt{npq}, \qquad \xi=\frac{X}{\sigma}, \\ &\alpha_1=\frac{X_s}{\sigma}=\frac{s-np}{\sigma}, \\ &\alpha_2=\frac{X_t}{\sigma}=\frac{t-np}{\sigma}, \end{aligned}\right\} \quad \cdots\cdots(\text{XII. }5)$$

とす. 之を**Laplace の定理**と云ふ.

n無限大ならずとも,n大なる時は,此の公式を用ひうる.

証明. s と t との間にある生起數(卽ち變數)は m 個ありとし,

$$s=r_1,\ r_2,\ r_3,\ r_4,\ \cdots\ r_{m-1},\ r_m=t,$$

なる値をとる. これより,

$$\xi=\frac{X}{\sigma}=\frac{r-np}{\sqrt{npq}}$$

に於ける ξ を計算することをう. 其の値も m 個ありて,

$$\alpha_1=\xi_1,\ \xi_2,\ \xi_3,\ \cdots\ \xi_{m-1},\ \xi_m=\alpha_2;$$

なる値をとる. n 甚だ大なる時(或は無限に大なる時),生起回數 r_1 なる確率, $f(n,r_1)$ は前條の公式により,

$$f(n,r_1)=\frac{e^{-\frac{\xi_1^2}{2}}}{\sqrt{2\pi npq}},$$

又 r_2 なる確率は,

$$f(n,r_2)=\frac{e^{-\frac{\xi_2^2}{2}}}{\sqrt{2\pi npq}},$$

$$\cdots\cdots\cdots\cdots\cdots\cdots$$

$$\cdots\cdots\cdots\cdots\cdots\cdots$$

故に,事象の生起回數が s と t との間にある確率は,これらの和

$$\sum_{i=1}^{m}\frac{e^{-\frac{\xi_i^2}{2}}}{\sqrt{2\pi npq}}=\sum_{i=1}^{m}\frac{e^{-\frac{\xi_i^2}{2}}}{\sqrt{2\pi}\,\sigma},\ \cdots\cdots\cdots\cdots\cdots\cdots[\mathrm{A}]$$

今　$$\xi_i=\frac{r-np}{\sigma},$$

に於て r が一個だけ大になりたる時,又

$$\xi_{i+1}=\frac{r+1-np}{\sigma}=\xi_i+\varDelta\xi$$ になつたとする. この二式より, $r=np+\xi_i\sigma$

$$r+1=np+(\xi_i+\varDelta\xi)\sigma,$$

邊々相減じて,

$$1=\sigma\varDelta\xi.$$

n 甚大なる時は σ 大となり, $\varDelta\xi$ も甚だ小となる. この式を[A]式に代入する時は,

$$\sum_{i=1}^{m}\frac{e^{-\frac{\xi_i^2}{2}}}{\sqrt{2\pi}}\varDelta\xi,\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots[\mathrm{B}]$$

となる. [B]式は, $n\to\infty$ 從て $m\to\infty$ なる時に,

$$\frac{1}{\sqrt{2\pi}}\int_{\alpha_1}^{\alpha_2}e^{-\frac{\xi^2}{2}}d\xi$$

となる. 卽ち Laplace の定理を證明し得た.

この積分の値は α_1 と α_2 との函數であるから,これを $\phi(\alpha_1,\alpha_2)$ にてあらはすならば,

$$\phi(\alpha_1,\alpha_2)=\frac{1}{\sqrt{2\pi}}\int_{\alpha_1}^{\alpha_2}e^{-\frac{\xi^2}{2}}d\xi.\quad\cdots\cdots(\mathrm{XII.}\ 5)$$

(5)式の積分値の計算は困難である.

$$\frac{1}{\sqrt{2\pi}}\int_{0}^{+\infty}e^{-\frac{\xi^2}{2}}d\xi=\frac{1}{2},$$

又は,

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}e^{-\frac{\xi^2}{2}}d\xi=1$$

の如く,特種なる場合は,比較的簡單に出來るのであるが,一般には容易でない. しかし,古來多くの學者により,積分計算が與へられて居る. 附表 III. は Sheppard の計算になるもので,

$$\phi(\alpha)=\frac{1}{\sqrt{2\pi}}\int_{0}^{\alpha}e^{-\frac{\xi^2}{2}}d\xi$$

が與へられて居る. この $\phi(\alpha)$ は α のみの函數であるから, α が明になれば $\phi(\alpha)$ が表によりて計算が出來るのである. 又, $\phi(\alpha_1\alpha_2)$ は次式により $\phi(\alpha_1)$ と $\phi(\alpha_2)$ とより計算することが出

來る. 即ち,

$$\frac{1}{\sqrt{2\pi}}\int_{\alpha_1}^{\alpha_2} e^{-\frac{\xi^2}{2}}d\xi = \frac{1}{\sqrt{2\pi}}\int_{0}^{\alpha_2} e^{-\frac{\xi^2}{2}}d\xi - \frac{1}{\sqrt{2\pi}}\int_{0}^{\alpha_1} e^{-\frac{\xi^2}{2}}d\xi.$$

即ち,　$\phi(\alpha_1\alpha_2) = \phi(\alpha_2) - \phi(\alpha_1)$.

式の右邊は表 III から計算することが出來る. 尙ほこの積分の性質及び附表 III の使用に關しては,詳しく第十三章を參照せよ.

注意. Laplace の定理は又, s 回及 t 回なる語に代ふるに, $np+\alpha_1\sqrt{npq}$ と $np+\alpha_2\sqrt{npq}$ を以てすることがある. 然る時は定理は次の如くなる.

n 回の獨立試行中事象の起る回數 r が

$$np+\alpha_1\sqrt{npq} < r < np+\alpha_2\sqrt{npq}$$

なる關係を滿足する確率は, $n\to\infty$ なる極限に於て,

$$\frac{1}{\sqrt{2\pi}}\int_{\alpha_1}^{\alpha_2} e^{-\frac{\xi^2}{2}}d\xi$$

である.

例 4. 銀貨を 10,000 回投じたる時,表の出る回數が 4,900 回と 5,100 回との間にある確率を問ふ.

解. $p=\frac{1}{2}$,　$n=10{,}000$,　$np=5{,}000$,

$$\sigma=\sqrt{npq}=\sqrt{2{,}500}=50,$$

$$\alpha_1=\frac{s-np}{\sigma}=\frac{4900-5000}{50}=-2,$$

$$\alpha_2=\frac{t-np}{\sigma}=\frac{5100-5000}{50}=+2,$$

故に

$$\phi(\alpha_1\alpha_2)=\frac{1}{\sqrt{2\pi}}\int_{-2}^{+2} e^{-\frac{\xi^2}{2}}d\xi.$$

而して, $e^{-\frac{\xi^2}{2}}$ は偶函數であるから(4 條 2),

$$\phi(\alpha_1\alpha_2)=\frac{2}{\sqrt{2\pi}}\int_{0}^{+2} e^{-\frac{\xi^2}{2}}d\xi = 2\phi(2),$$

しかるに,この式の右邊は, $\phi(2)$ は附表 II で $\alpha=2$ に對して直ちに求めうるから,

$$\phi(\alpha_1\alpha_2) = 2\times 0{\cdot}47725 = 0{\cdot}9545.$$

即ち完全なる圓板を, 10,000 回投じたるとき,表が出る回數は, 5,000 より 100 回以內の差異を有するもの, 9,500 回あり. これ以外にあるもの, 10,000 回中僅に 500 である(附表 III の用法參照).

例 5. 表の出る確率 $\frac{1}{2}$ なる圓板を, 10,000 個 (10,000 回) 投じたる時,表が 5,000—5,010, 5,010—5,020, ‥‥等なる確率を求む(第十三章例 3 參照).

解. 第 91 表は計算の順序を示すものである. 第 5 列はその計算結果である.

(i) 先づ,第一に階級の幅及數を決定しなければならぬ. これは任意に決定してもいいが,階級數 20 乃至 30 位になる樣に,階級の幅を決定するのが便利である. 階級數少なきに失する時は用をなさないが,多きに過ぎる時は煩雜なばかりで,概觀が出來ないと云ふ缺點がある. (詳しくは,次章を見よ.) $np=5000$ を界として,これより前と後とは對稱になるから,今は表の如く, 5,000—5,010 を一階級とし, 5150 までを 15 階級に分類した. さうすれば計 30 の階級が出來上るが,分布は對稱であるから, 5000 を境とし何れか一方を計算すれば,他は計算するに及ばないと云ふ便宜がある[1].

[1] $r=5000$ を階級の境界の値とするか,又は階級の中央となる樣に階級を選ぶが良い. でなければ斯く計算の省略は出來ない.

尙一つ注意すべきことは,階級を(5000—5009), (5010—5019)と云ふ風に分類してはならぬ事である. その理由は $Y=\frac{1}{\sqrt{2\pi}\,\sigma}e^{-\frac{x^2}{2\sigma^2}}$ は連續函數であるから,階級間に隙間を作つて置けば,結果は合計 10000 とならず,甚しき不都合が出來るからである. しかし,10,000 個の銅貨を投ずる試驗を N 回繰返して,この計算の眞否を試みる時,若し階級境界の値が出た時は,出た度數の半數を上の階級に,半數を下の階級に繰入れて,度數分布表を作製すべきである.

(ii) 階級が決定出來たらば,各階級の境界の內絕對值の大なる者と np 即ち 5000 との差を求めよ. 之を X として第二列に記入した. X は階級の中央の値でないことに注意せよ.

(iii) 次に $\alpha=\frac{X}{\sqrt{npq}}$ を計算せよ. 第三列がそれである.

(iv) 次に附表 III によつて, $\phi(\alpha)$ を求むるのである(附表使用法參照). 附表は小數點を除く爲に, $\phi(\alpha)$ の各數を10000倍してある. 即ち,出て來る數は $10{,}000\phi(\alpha)$ で, 10,000 觀察に於ける分布である.

かく　$\phi(\alpha_1)$　$\phi(\alpha_2)$　$\cdots\cdots$　を求め第4列に記入する.

(v) 第一階級, 5,000—5,010 の階級の度數は $\phi(\alpha_1)$ 即 793 であるが,第二階級(5010—5020)では, $1554-793=761$ である. 第三の階級以下第二にならふ. 何となれば α_1, α_2 を階級の境界とすれば,

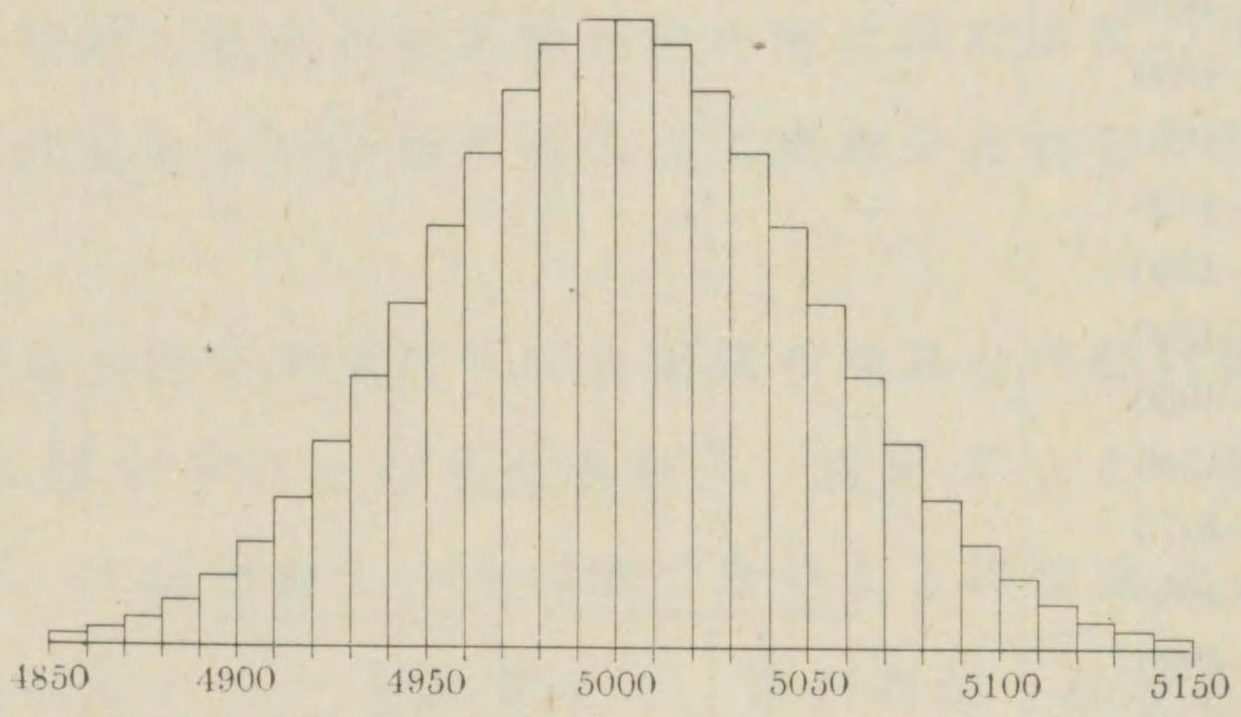
$$\phi(\alpha_1,\alpha_2)=\phi(\alpha_2)-\phi(\alpha_1)$$

であり, $\phi(\alpha_2)$ は各階級の行に記され, $\phi(\alpha_1)$ は其の直前の階級に記入されてあるからである. かく出來上りたる $\phi(\alpha_1\alpha_2)$ は 10,000 分比の分布であつて,第五列に記入せらる. 確率を表示するに

第 91 表

r の階級	X	$\frac{X}{\sigma}=\alpha$	$\phi(\alpha)$	10000分比分布
0—4850				13
4850—4860				13
4860—4870				21
4870—4880				35
4880—4890				57
4890—4900				89
4900—4910				131
4910—4920	—	—	—	189
4920—4930				260
4930—4940				343
4940—4950				436
4950—4960				532
4960—4970				624
4970—4980				703
4980—4990				761
4990—5000				793
5000—5010	+ 10	0·2	793	793
5010—5020	+ 20	0·4	1554	761
5020—5030	+ 30	0·6	2257	703
5030—5040	+ 40	0·8	2881	624
5040—5050	+ 50	1·0	3413	532
5050—5060	+ 60	1·2	3849	436
5060—5070	+ 70	1·4	4192	343
5070—5080	+ 80	1·6	4452	260
5080—5090	+ 90	1·8	4641	189
5090—5100	+100	2·0	4772	131
5100—5110	+110	2·2	4861	89
5110—5120	+120	2·4	4918	57
5120—5130	+130	2·6	4953	35
5130—5140	+140	2·8	4974	21
5140—5150	+150	3·0	4987	13
5150—+∞	+ ∞	∞	5000	13
計	—	—	—	10,000

は,第五列の數を 10,000 を以て割らねばならぬ. かく出來た部分は表の後半である. 生起數 0 より 5,000 までの度數(又は確率)の分布表はこれと對稱であるから,計算せずとも直ちに記入することが出來る. この確率の分布表を Histogram に圖示すれば次の如くになる.



第 58 圖

系 1. n 個の獨立試行に於て,事象の生起する回數と試行回數との比, $\frac{r}{n}$ が

$$p+\alpha_1\sqrt{\frac{pq}{n}}<\frac{r}{n}<p+\alpha_2\sqrt{\frac{pq}{n}}$$

を滿足する確率 P は, n 甚だ大なる時は近似的に,

$$P=\frac{1}{\sqrt{2\pi}}\int_{\alpha_1}^{\alpha_2}e^{-\frac{\xi^2}{2}}d\xi$$

である.

證明. Laplace の定理では,生起回數 r が

$s=np+\alpha_1\sqrt{npq}$ と $t=np+\alpha_2\sqrt{npq}$ との間にある確率を與へた. 今この不等式

$$np+\alpha_1\sqrt{npq}<r<np+\alpha_2\sqrt{npq}$$

は不等式, $p+\alpha_1\sqrt{\frac{pq}{n}}<\frac{r}{n}<p+\alpha_2\sqrt{\frac{pq}{n}}$ と同一である. 故に, Laplace の定理は上の通り云ひあらはすことが出來る. 系 1 は Laplace の定理にすぎないのである.

系 2. n 回の獨立試行に於て,生起の度數を r とすれば $\frac{r}{n}$ が $p-\alpha\sqrt{\frac{pq}{n}}$ と $p+\alpha\sqrt{\frac{pq}{n}}$ との間にある確率は, $n\to\infty$ なる極限に於ては,

$$\frac{1}{\sqrt{2\pi}}\int_{-\alpha}^{+\alpha}e^{-\frac{\xi^2}{2}}d\xi,\quad 即ち,\quad \sqrt{\frac{2}{\pi}}\int_0^{+\alpha}e^{-\frac{\xi^2}{2}}d\xi$$

である.

證明. 系 1 に於て $\alpha_1=0$, $\alpha_2=\alpha$ と置けば,生起率 $\frac{r}{n}$ が p と $p+\alpha\sqrt{\frac{pq}{n}}$ との間にある確率, P_1 が計算せらる.

即ち, $$P_1=\frac{1}{\sqrt{2\pi}}\int_0^{\alpha}e^{-\frac{\xi^2}{2}}d\xi,$$

又, $\alpha_2=0$, $\alpha_1=-\alpha$ と置けば生起率が $p-\alpha\sqrt{\frac{pq}{n}}$ と p との間にある確率, P_2 は

$$P_2=\frac{1}{\sqrt{2\pi}}\int_{-\alpha}^{0}e^{-\frac{\xi^2}{2}}\quad となる.$$

$e^{-\frac{\xi^2}{2}}$ は偶函數であるからこの式の右邊は

$$P_2=\frac{1}{\sqrt{2\pi}}\int_0^{\alpha}e^{-\frac{\xi^2}{2}}d\xi\quad に等し.$$

即ち,生起率が $p-\alpha\sqrt{\frac{pq}{n}}$ と $p+\alpha\sqrt{\frac{pq}{n}}$ との間にある確率は, P_1 と P_2 との和,即ち,

$$\frac{1}{\sqrt{2\pi}}\int_{-\alpha}^{+\alpha}e^{-\frac{\xi^2}{2}}d\xi=\frac{2}{\sqrt{2\pi}}\int_0^{+\alpha}e^{-\frac{\xi^2}{2}}d\xi$$