基礎課程

数理統計

河田敬義
丸山文行
共著

裳華房

基 礎 課 程

数 理 統 計

東京大学教授　総理府技官
理学博士　　理学士
河田敬義　丸山文行

共 著

東京 裳 華 房 発行

著作者	河田敬義　丸山文行
発行者	吉野元章
印刷者	塩原政雄

基礎課程 数理統計

定価 380円

発行元	東京都千代田区四番町8番地ノ1 振替口座東京107番 電話東京301－9166～8番	合名会社 裳華房
印刷所	東京都千代田区神田神保町1の33	政弘印刷株式会社

社団法人 自然科学書協会会員

牧製本

昭和26年3月15日 第1版印刷　昭和26年3月20日 第1版発行
昭和26年5月20日 第2版発行　昭和27年11月25日 第3版増補発行
昭和30年1月25日 第4版発行　昭和31年12月1日 修正第5版発行
昭和34年4月15日 第6版発行
昭和36年4月20日 第7版発行
昭和37年4月20日 第8版発行
昭和38年8月20日 第9版発行

序

本書は，新制大学の初年級の教科書または参考書として役立つことを目標として，数理統計の原理を余り多くの予備知識なしで理解することができるように，しかもなるべく簡潔に説明したものである．

近時，推則統計学の発展とともに，数理統計の基礎的知識が強く要求せられている．本邦においても，近代統計に関する書物が少からず刊行されて，これらによって新しい統計の意義と，実際上の方法についての理解が，急速に深められつつあることはまことに喜ばしいことである．その一面，その数学的基礎について親切な解説のあるものは乏しいようである．

本書は，新制大学における基礎課程としての役割にかんがみて，特に次の点を心掛けた．

(I) 実際の統計と，数学理論との関係を明らかにして，確率論の諸命題の表わす実際上の意味，母集団と標本との関係，統計的推則の立つ仮定等について，できるだけ明確に解説することを試みた．

(II) 近代統計は，かなり程度の高い数学を必要とするために，多くの書物は単に実際上の方法を指示するに止まって，その数学的証明を欠いている．本書では，本篇において，これら諸命題の数学的仮定を明らかに述べ，附録において，大部分の定理の証明を与えることにした[1]．

(III) 本書の特色の一つは，多くの図表を用いたことである[2]．これらの大部分は全く新たに画かれたものであって，しかも十分に精密で実際の用に供し得るものである．数値表を用いる代りに，本書では図によって大体の値を求める方法を採用した．

(IV) 例題として，社会科学や自然科学の各方面から実際の問題を取上げて，具体的に解説した．

さらに進んで数理統計に関する知識を求められる方には

H. Cramér, Mathematical methods of statistics (1946)

(Princeton University Press)

をおすすめしたい.

なお本書の分担は，本文を河田敬義が，図表と例題とを丸山文行が受持った. 本書の企劃と構成は，彌永昌吉教授に負うものである. また例題のために，白石一誠氏，木村等氏，田口玄一氏から資料を頂いた. その他竹之内脩氏はじめ統計数理研究所の方々の御援助を受けた. ここに深く感謝の意を表わしたい.

1951 年 2 月

著　者

第3版に際して，巻末に「要約と問題」を附け加え，また第5版にては「附表 1～5」を附け加えた. 本文の理解に役立つことと思う.

---

1) 即ち本篇においては順列，組合せ，二項定理，積分の概念，初等函数の簡単な性質等のみを既知として仮定する. 難かしい定理の証明はすべて附録にゆずることにした.

証明を略したのは IV, §4, (e); V, §3, (a*), (b*), (c*), §5, (d*) だけである.

なお本篇中 *) 印を附けた節は省略してもよい. それらは IV, §1, 3; §2, 3; §3, 2～3; §4, 5; V, §1, 6; §3, 2; §4, 2; §5, 2～3; §6, 3～4 である.

附録においては，積分の変数変換，複素数，行列式，連立1次方程式，n 次元解析幾何学等を用いる.

2) 確率密度の図は，原則として，縦の 0.1 に対応する長さの 1.6 倍が，横で 1 に対応するようになっている.

3) 統計用語は，使う人によってまちまちで，定まっていない. 従って本書で用いた訳語も暫定的なものである.

目　次

I まえがき

§1 記述統計

II 確　率

§1 確　率

§2 基本定理

§3 確率変数

§4 大数の法則

III 母集団と標本

§1 母集団と標本

§2 大　標　本

§3 統計的推則

IV 分　　布

§1 離散分布

§2 連続分布

§3 多次元連続分布

§4 標本分布

V 検定と推定

§1 母数に関する検定

§2 分散分析法

§3 母数の推定

§4 最小2乗法

§5 あてはまりの良さの検定

§6 検定についての補足

附録 確率変数と分布

§1 確率変数

§2 特性函数

§3 中心極限定理

§4 種々の1次元分布

§5 n 次 元 分 布

検定及び推定に用いる図表

I まえがき

§1 記述統計

1. 近代統計学の目的は，或る種の偶然的事象に関して，統計的法則を確立し，または種々の統計的推測を行うことにある．実際に統計的法則，或いは統計的推測の意味するところは，確率の理論を用いて表わされるのであって，後に次第に説明することにする．

しかしながら統計学発達の歴史においても，また各個人が統計的方法にたずさわる過程においても，まず大量を記述する段階より始まる．即ち或る観察や測定等の結果として，多数の数値が得られた場合に，それぞれの目的に応じて，その数値全体に関する様子を知るために，この数値の集合を種々分析，整理して記述するのである．この種の統計を**記述統計**という．

近代統計学の本来の目標とするところからは，観察や測定等は，何等かの統計的法則を確立し，或いは統計的推測を行うために，予め企画されたところに従って実行されるべきである．従って明確な目的なしに得られた大量を単に記述することは，近代統計学の立場からは，必ずしもつねに有意義のこととは言えない．

ここでは後に必要になって来る範囲で，記述統計の極めて基礎的な一，二の方法を説明するに止める．即ち度数分布，分散(標準偏差)，共分散(相関係数)のみについて簡単に解説する．記述統計に用いられる各種の手段について詳細に説明する意図はない．

2. 度数分布

一般に種々の値を取り得る量を**変量**という．変量を X, Y 等で表わす．変

量の取り得る値は離散的な場合(例えば，0, 1, 2,・・・等)もあり，また連続的に変化し得る場合もある．例えば1地方における1日の火災件数 X とか，一定回数骰を投げて，その中6の目の出る回数 Y 等は前者に属し，また一定年齢の男子の身長 X, 体重 Y 等は，後者に属する．連続的に変化し得る変量に対しては，通常一定の幅で区間に分けて，**級**に分類する．同一の値 x を持つ，(または同一の級 c に属する)個体の個数 f を，その**度数**という．また度数 f を，その全体の個数 n で割った商 f/n を，その**相対度数**という．各々の値 x に，(または各々の級 c に)その度数 f (又は相対度数 f/n) を対応させたものを，**度数分布**，(又は**相対度数分布**)という．

例 1. 12個の骰を投げて，その中6の目の出た骰の個数 X を考える．X の取り得る値は，0, 1, 2,・・・, 12 である．今 $n=4096$ 回12個の骰を投げて，$x=0, 1,$・・・, 12 に対する度数 f を数えて，表1の如き度数分布を得た．

(相対度数の和が 1.0000 にならないのは，4捨5入の誤差のためである．1.0000 になるように修整することもあるが，本書ではこれを行わない.)

例 2. 10歳の男児の身長(単位 cm, 読みとりは 0.1 cm まで) を $n=2451$ (人)について測定し，区間の幅を3として，級 c_1-c_{14} に分けて，表2の如き度数分布を得た．

(昭和24年度文部省が行った体育実態調査の資料による.)

通常級の個数は 8-20 ぐらいに取る．

これらの結果を図示するために，図1のような**柱状図表**(ヒストグラム)や**度数折線**を用いることもある．

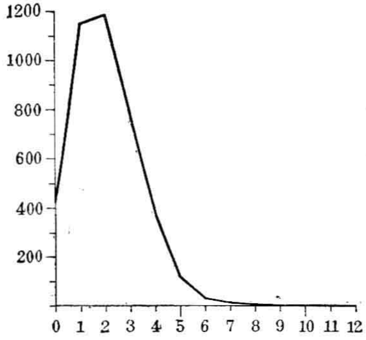
Xの値	度　数	相対度数
x	f	f/n
0	447	0.1091
1	1145	0.2795
2	1181	0.2883
3	796	0.1943
4	380	0.0928
5	115	0.0281
6	24	0.0059
7	7	0.0017
8	1	0.0002
9	0	0.0000
10	0	0.0000
11	0	0.0000
12	0	0.0000
計	4096	0.9999

表　1

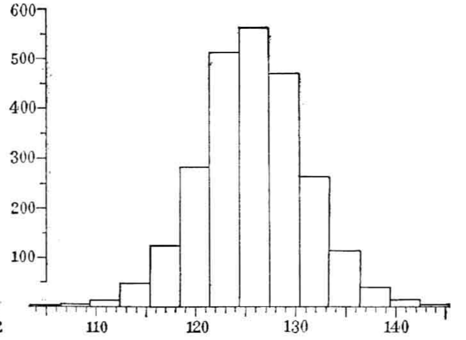
3. 平均値，分散，標準偏差

或る変量 X を測定して，n 個の値

(1)　　$\mathfrak{X}=\{x_1, x_2, \cdots, x_n\}$



例 1 の度数折線



例 2 の柱状図表

図　1

を得たとする．このとき

$$m=\frac{1}{n}(x_1+x_2+\cdot\cdot\cdot\cdot+x_n)$$

を $\mathfrak{X}$ の**平均値**という．$m=m(\mathfrak{X})$ と表わす．

今後は和 $(x_1+x_2+\cdot\cdot\cdot\cdot+x_n)$ を $\sum_{i=1}^{n} x_i$ と表わす．これに関して

$$\sum_{i=1}^{n}(x_i+y_i)=\sum_{i=1}^{n}x_i+\sum_{i=1}^{n}y_i$$

$$\sum_{i=1}^{n}ax_i=a\sum_{i=1}^{n}x_i$$

が成り立つ．これによれば $m=\frac{1}{n}\sum_{i=1}^{n}x_i$ と表わされる．

変量 X の代りに定数 $a(\neq 0)$, b によって

$$(2)\qquad X=aY+b$$

$$\left(Y=\frac{X-b}{a}\right)$$

と表わされる変量 Y を考える．(1) の n 個の値に対応して

階級	区　間	度数	相対度数
c	I	f	f/n
c_1	103.45–106.45	1	0.0004
c_2	106.45–109.45	6	0.0024
c_3	109.45–112.45	14	0.0057
c_4	112.45–115.45	49	0.0200
c_5	115.45–118.45	126	0.0514
c_6	118.45–121.45	282	0.1151
c_7	121.45–124.45	511	0.2085
c_8	124.45–127.45	562	0.2293
c_9	127.45–130.45	469	0.1914
c_{10}	130.45–133.45	262	0.1069
c_{11}	133.45–136.45	115	0.0469
c_{12}	136.45–139.45	40	0.0163
c_{13}	139.45–142.45	12	0.0049
c_{14}	142.45–145.45	2	0.0008
計		2451	1.0000

表　2

$$x_i = ay_i + b \qquad \left(y_i = \frac{x_i - b}{a}\right) \qquad (i = 1, 2, \cdots, n)$$

により定まる y_i の集合を

$$(3) \qquad \mathfrak{Y} = \{y_1, y_2, \cdots, y_n\}$$

とおく.

$$m = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}\sum_{i=1}^{n}(ay_i + b) = a\left\{\frac{1}{n}\sum_{i=1}^{n} y_i\right\} + b$$

即ち
$$m(\mathfrak{X}) = a \cdot m(\mathfrak{Y}) + b$$

となる.

$\mathfrak{X}$ の平均値 m のまわりのちらばりの度合を示す一つの値として

$$\sigma^2 = \frac{1}{n}\{(x_1 - m)^2 + (x_2 - m)^2 + \cdots + (x_n - m)^2\}$$

を，$\mathfrak{X}$ の**分散**といって，$\sigma^2 = \sigma^2(\mathfrak{X})$ と表わす．σ^2 の平方根 $\sigma(\geqq 0)$ を $\mathfrak{X}$ の**標準偏差**という．分散は

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - m)^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - 2m\left\{\frac{1}{n}\sum_{i=1}^{n} x_i\right\} + m^2$$

$$= \frac{1}{n}\sum_{i=1}^{n} x_i^2 - m^2$$

とも表わされる.

変量 X の代りに，(2) のごとく Y を取り，(1) の $\mathfrak{X}$ から，(3) の $\mathfrak{Y}$ を作る．$m = m(\mathfrak{X})$, $m' = m(\mathfrak{Y})$, $\sigma^2 = \sigma^2(\mathfrak{X})$, $\sigma'^2 = \sigma^2(\mathfrak{Y})$ とおく.

$$(4) \quad \sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - m)^2 = \frac{1}{n}\sum_{i=1}^{n}\{(ay_i + b) - (am' + b)\}^2$$

$$= a^2\left\{\frac{1}{n}\sum_{i=1}^{n}(y_i - m')^2\right\}$$

$$= a^2\sigma'^2 = a^2\left\{\frac{1}{n}\sum_{i=1}^{n} y_i^2 - m'^2\right\} = a^2\left\{\frac{1}{n}\sum_{i=1}^{n} y_i^2\right\} - (m - b)^2$$

と表わされる.

もしも $\mathfrak{X}$ の n 個の値のうち，相異なるものが $\{x_1', \cdots, x_r'\}$ で，x_i' の度数が f_i であるとする： $n = f_1 + f_2 + \cdots + f_r$.

平均値 $m=m(\mathfrak{X})$, 分散 $\sigma^2=\sigma^2(\mathfrak{X})$ は

Xの値	x_1'	x_2'	・・・	x_r'
度 数	f_1	f_2		f_r

(5) $$m=\sum_{i=1}^{r}\frac{f_i}{n}x_i'$$

(6) $$\sigma^2=\sum_{i=1}^{r}\frac{f_i}{n}(x_i'-m)^2=\sum_{j=1}^{r}\frac{f_i}{n}x_i'^2-m^2$$

また $x_i'=ay_i'+b$ $(i=1,\cdot\cdot\cdot,r)$ とおくと

$$m=a\left\{\sum_{i=1}^{r}\frac{f_i}{n}y'_i\right\}+b$$

$$\sigma^2=a^2\left\{\sum_{i=1}^{r}\frac{f_i}{n}y_i'^2\right\}-(m-b)^2$$

と表わされる.

連続的変量を階級に分けて度数分布を取った場合には，各級の区間の中央の値をその代表値とする.

例 1 の X に対して $X=Y+2$ $(a=1,\ b=2)$ とおくと

$$m=\frac{1}{n}\sum_{i=1}^{r}f_iy_i'+b=\frac{-1}{4096}+2$$

$$=2.000$$

$$\sigma^2=\frac{1}{n}\sum_{i=1}^{r}f_iy_i'^2-(m-b)^2$$

$$=\frac{6879}{4096}-0.000=1.679$$

$$\sigma=1.296$$

x'	y_i'	f_i	f_iy_i'	$f_iy_i'^2$
0	−2	447	− 894	1788
1	−1	1145	−1145	1145
2	0	1181	0	0
3	1	796	796	796
4	2	380	760	1520
5	3	115	345	1035
6	4	24	96	384
7	5	7	35	175
8	6	1	6	36
計		4096	−1	6879

表 3

例 2 の変量 X に対しては，各区間の中央値を代表値 x_i' として，$a=3$, $b=125.95$に取り $X=3Y+125.95$ とおけば

$$m=\frac{a}{n}\sum f_iy_i'+b=\frac{3}{2451}\times(-192)+125.95$$

$$=-0.235+125.95=125.715$$

$$\sigma^2=\frac{a^2}{n}\sum f_iy_i'^2-(m-b)^2=\frac{9}{2451}\times 7736-(0.235)^2$$

$$=28.4064-0.0552=28.3512$$

$\sigma = 5.325$

$f_i y_i'^2$ は $f_i y_i'$ に y_i' を掛けて計算する.

x_i'	y_i'	f_i	$f_i y_i'$	$f_i y_i'^2$
104.95	-7	1	$-$ 7	49
107.95	-6	6	$-$ 36	216
110.95	-5	14	$-$ 70	350
113.95	-4	49	-196	784
116.95	-3	126	-378	1134
119.95	-2	282	-564	1128
122.95	-1	511	-511	511
125.95	0	562	0	0
128.95	1	469	469	469
131.95	2	262	524	1048
134.95	3	115	345	1035
137.95	4	40	160	640
140.95	5	12	60	300
143.95	6	2	12	72
		2451	-192	7736

表

4. 共分散, 相関係数

一定年齢の学生の身長 X と体重 Y を測定した場合のように, 各個体に 2 組の数値が同時に定まる場合を考える. いま変量の対 (X, Y) の n 組の測定値として

$$(\mathfrak{X}, \mathfrak{Y}) = \{(x_1, y_1), (x_2, y_2), \cdot\cdot\cdot, (x_n, y_n)\}$$

を得たとする. $\mathfrak{X} = \{x_1, x_2, \cdot\cdot\cdot x_n\}$, $\mathfrak{Y} = \{y_1, y_2, \cdot\cdot\cdot, y_n\}$ の平均値, 分散を, それぞれ $m_1 = m(\mathfrak{X}), m_2 = m(\mathfrak{Y}), \sigma_1^2 = \sigma^2(\mathfrak{X})\ \sigma_2^2 = \sigma^2(\mathfrak{Y})$ とする. そのとき

$$(7)\quad \mu = \frac{1}{n}\{(x_1 - m_1)(y_1 - m_2) + (x_2 - m_1)(y_2 - m_2) + \cdot\cdot\cdot + (x_n - m_1)(y_n - m_2)\}$$

を $(\mathfrak{X}, \mathfrak{Y})$ の**共分散**といって, $\mu = \mu(\mathfrak{X}, \mathfrak{Y})$ と表わす. 又

$$\rho = \mu/\sigma_1\sigma_2$$

を $(\mathfrak{X}, \mathfrak{Y})$ の**相関係数**といって, $\rho = \rho(\mathfrak{X}, \mathfrak{Y})$ と表わす. 公式

$$\left(\sum_{i=1}^{n} a_i b_i\right)^2 \leqq \left(\sum_{i=1}^{n} a_i^2\right)\left(\sum_{i=1}^{n} b_i^2\right)$$

において, $a_i = x_i - m_1$, $b_i = y_i - m_2$, とおいて, 両辺を n^2 で割れば

$$\mu^2 \leqq \sigma_1^2 \cdot \sigma_2^2$$

を得る. 故に

$$-1 \leqq \rho \leqq 1$$

なる関係が成り立つ. 特に定数 a, b によって, $Y = aX + b$ と表わされ,

$$y_i = ax_i + b \quad (i = 1, 2, \cdot\cdot\cdot, n)$$

なる関係があったとすれば，これを (7) に代入すれば，$\mu=a\sigma_1^2$，及び $\sigma_2^2=a^2\sigma_1^2$ となるから

$$\begin{cases} a>0 \quad \text{ならば} \quad \rho=1 \\ a<0 \quad \text{ならば} \quad \rho=-1 \end{cases}$$

である．即ち 2 変量 X, Y の間に 1 次関係 $Y=aX+b$ があれば，$(\mathfrak{X}, \mathfrak{Y})$ の相関係数 ρ は両極端の値 ±1 を取る．一般に ρ の値は $\mathfrak{X}$, $\mathfrak{Y}$ の間の相互関係の度合を示す値で，絶対値の大きい程関係の密接なことを示すものと見なされる．$\rho>0$, $\rho<0$, $\rho=0$ に対して，X, Y はそれぞれ正に相関する，負に相関する，相関が無いという．

定数 $a>0$, $c>0$ 及び b, d によって

$$X=aU+b \qquad Y=cV+d$$

となる変量 U, V を考える．

$$x_i=au_i+b \qquad y_i=cv_i+d \qquad (i=1, 2, \cdots, n)$$

によって

$$(\mathfrak{U}, \mathfrak{V})=\{(u_1, v_1), (u_2, v_2), \cdots, (u_n, v_n)\}$$

を作る．$\mathfrak{U}=\{u_1, \cdots, u_n\}$, $\mathfrak{V}=\{v_1, \cdots, v_n\}$ に対して，$m_1'=m(\mathfrak{U})$, $m_2'=m(\mathfrak{V})$, $\sigma_1'^2=\sigma^2(\mathfrak{U})$, $\sigma_2'^2=\sigma^2(\mathfrak{V})$, $\mu'=\mu(\mathfrak{U}, \mathfrak{V})$, $\rho'=\rho(\mathfrak{U}, \mathfrak{V})$ とおくと

$$\mu=\frac{1}{n}\sum_{i=1}^{n}(x_i-m_1)(y_i-m_2)=\frac{ac}{n}\sum_{i=1}^{n}(u_i-m_1')(v_i-m_2')=ac\mu'$$

となる．一方 (4) によって，$\sigma_1=a\sigma_1'$, $\sigma_2=c\sigma_2'$ であるから，

$$\rho(\mathfrak{X}, \mathfrak{Y})=\rho=\rho'=\rho(\mathfrak{U}, \mathfrak{V})$$

となる．

今後二重添数を持った和 $(x_{11}+x_{12}+\cdots+x_{1s}+x_{21}+\cdots+x_{2s}+\cdots+x_{r1}+x_{r2}+\cdots+x_{rs})$ のことを $\sum_{i=1}^{r}\sum_{j=1}^{s}x_{ij}$ と表わす，これに関して

$$\sum_{i=1}^{r}\sum_{j=1}^{s}x_{ij}=\sum_{j=1}^{s}\sum_{i=1}^{r}x_{ij}=\sum_{i=1}^{r}\left(\sum_{j=1}^{s}x_{ij}\right)=\sum_{j=1}^{s}\left(\sum_{i=1}^{r}x_{ij}\right)$$

が成り立つ．同様に

$$\sum_{i=1}^{r}\sum_{j=1}^{s}(x_{ij}+y_{ij})=\sum_{i=1}^{r}\sum_{j=1}^{s}x_{ij}+\sum_{i=1}^{r}\sum_{j=1}^{s}y_{ij}$$

$$\sum_{i=1}^{r}\sum_{j=1}^{s}ax_{ij}=a\cdot\sum_{i=1}^{r}\sum_{j=1}^{s}x_{ij}$$

である．

$\mathfrak{X}$ の n 個の値の中相異なるものを $\{x_1',\cdot\cdot\cdot, x_r'\}$, $\mathfrak{Y}$ の中相異なるものを $\{y_1',\cdot\cdot\cdot, y_s'\}$ とする．$(\mathfrak{X}, \mathfrak{Y})$ の中

$$(x_i',\ y_j')\qquad(i=1,\cdot\cdot\cdot,\ r;\ j=1,\cdot\cdot\cdot,\ s)$$

の度数を $f_{ij}\geqq 0$ とする：

$$\sum_{i=1}^{r}\sum_{j=1}^{s}f_{ij}=n$$

また $\mathfrak{X}$ の中で x_i' の度数を g_i, $\mathfrak{Y}$ の中で, y_j' の度数を h_j とすると

$$g_i=\sum_{j=1}^{s}f_{ij}\qquad h_j=\sum_{i=1}^{r}f_{ij}\qquad(i=1,\cdot\cdot\cdot,r;\ j=1,\cdot\cdot\cdot,s)$$

である．従って

$$m_1=\frac{1}{n}\sum_{i=1}^{r}g_ix_i'=\frac{1}{n}\sum_{i=1}^{r}\sum_{j=1}^{s}f_{ij}x_i'$$

$$m_2=\frac{1}{n}\sum_{j=1}^{s}h_jy_j'=\frac{1}{n}\sum_{i=1}^{r}\sum_{j=1}^{s}f_{ij}y_j'$$

となり．$(\mathfrak{X}, \mathfrak{Y})$ の共分散 μ は

$$(8)\qquad \begin{aligned}\mu&=\frac{1}{n}\sum_{i=1}^{r}\sum_{j=1}^{s}f_{ij}(x_i'-m_1)(y_j'-m_2)\\&=\frac{1}{n}\sum_{i=1}^{r}\sum_{j=1}^{s}f_{ij}x_i'y_j'-\frac{m_2}{n}\sum_{i=1}^{r}\sum_{j=1}^{s}f_{ij}x_i'\\&\qquad\qquad-\frac{m_1}{n}\sum_{i=1}^{r}\sum_{j=1}^{s}f_{ij}y_j'+m_1m_2\\&=\frac{1}{n}\sum_{i=1}^{r}\sum_{j=1}^{s}f_{ij}x_i'y_j'-m_1m_2\end{aligned}$$

と計算される．そこで