Philosophical Magazine Series 6

ISSN: 1941-5982 (Print) 1941-5990 (Online) Journal homepage: http://www.tandfonline.com/loi/tphm17

XLVI. The application of solid hypergeometrical series to frequency distributions in space

L. Isserlis B.A.

To cite this article: L. Isserlis B.A. (1914) XLVI. The application of solid hypergeometrical series to frequency distributions in space , Philosophical Magazine Series 6, 28:165, 379-403, DOI: 10.1080/14786440908635223

To link to this article: http://dx.doi.org/10.1080/14786440908635223

XLVI. *The Application of Solid Hypergeometrical Series to Frequency Distributions in Space.* *By* L. ISSERLIS, *B.A.* *

[Plate VII.]

§1. THE connexion between the frequency distribution of one variable character, and the hypergeometrical series is well known. It has been discussed exhaustively by Professor Karl Pearson in a series of memoirs in the Phil. Trans. from 1895 on, where a series of frequency curves are studied which may be described as parallels to the hypergeometrical series.

In a paper entitled "On certain properties of the hypergeometrical series, and on the fitting of such series to observation polygons in the theory of chance," Phil. Mag. 1899, pp. 236–246, Pearson showed how to determine from a given frequency distribution the constants of the corresponding hypergeometrical series. The method of moments was employed, and the fitting of the series was equivalent to the statement of a problem of chance with a theoretical distribution of events similar to the actual distribution.

Although much progress has been made with the study of the normal surface, no general theory of frequency surfaces analogous to Pearson's Skew frequency curves exists at present, and this paper is an attempt to make a first step towards such a theory by solving the problem of fitting a double hypergeometrical series to a frequency distribution with two variable characters.

§ 2. The corresponding chance problem may be stated as follows :—

A bag contains n balls of which pn are white and qn are black ; r balls are drawn and not replaced ; a second draw of r' balls is made. This is repeated N times. If N is a large number, the theoretical frequency of s black balls in the first draw and s' in the second is

$$\frac{\mathrm{N}\binom{r}{s}\binom{r'}{s'}\left(qn\right)_{s+s'}\left(pn\right)_{r+r'-s-s'}}{(n)_{r+r'}}.$$

We may denote this by $\mathrm{N}z(s, s')$,
where

$$\binom{h}{k}\equiv\frac{h(h-1)(h-2)\ldots(h-k+1)}{k!}$$

and

$$(h)_k\equiv h(h-1)(h-2)\ \ldots\ (h-k+1).$$

* Communicated by Prof. Karl Pearson, F.R.S.

The "fitting" of a double hypergeometrical series to a set of statistics, say the frequency of a certain age of husband and a certain age of wife, in a large number of marriages will involve among the other things the determination of r, r', n, and q in the corresponding chance problem. The value of q and the ratios of r and r' to n will be numbers characteristic of the particular type of frequency distribution.

§ 3. Let

$$[h]_k \equiv h(h+1)(h+2)\ldots(h+k-1),$$

then

$$z(s, s') = \frac{(pn)_{r+r'}}{(n)_{r+r'}} \frac{\binom{r}{s}\binom{r'}{s'}\binom{qn}{}_{s+s'}}{[pn-r-r'+1]_{s+s'}}.$$

If we put

$$-r=\alpha, \quad -r'=\alpha', \quad -qn=\beta, \quad pn-r-r'+1=\gamma,$$

then

$$z(s, s') = \frac{(pn)_{r+r'}}{(n)_{r+r'}} \frac{[\alpha]_s[\alpha']_{s'}[\beta]_{s+s'}}{s!\, s'!\, [\gamma]_{s+s'}}.$$

Let $F(\alpha, \alpha', \beta, \gamma, x, y)$ denote the double hypergeometrical series

$$\sum_s \sum_{s'} \frac{[\alpha]_s[\alpha']_{s'}[\beta]_{s+s'}}{s!\, s'!\, [\gamma]_{s+s'}} x^s y^{s'}.$$

So that

$$\Sigma\Sigma z(s, s') = \frac{(pn)_{r+r'}}{(n)_{r+r'}} F(\alpha, \alpha', \beta, \gamma, 1, 1),$$

or say $$= \Delta F_1(\alpha, \alpha', \beta, \gamma).$$

§ 4. We will begin by finding two partial differential equations satisfied by $F(\alpha, \alpha', \beta, x, y)$ as a preliminary to the calculation of the moments of the series.

The coefficient of $\dfrac{y^{s'}[\alpha']_{s'}[\beta]_{s'}}{s'!\, [\gamma]_{s'}}$ in $F(\alpha, \alpha', \beta, \gamma, x, y)$

is $$\sum_s \frac{[\alpha]_s[\beta+s']_s x^s}{s!\, [\gamma+s']_s} = F(\alpha, \beta+s', \gamma+s', x)$$
$$= X_{s'} \text{ say}.$$

Similarly write $Y_s = F(\alpha', \beta+s, \gamma+s, y)$.

Then $F(\alpha, \alpha', \beta, \gamma, xy)$

$$=X_0+X_1\frac{\alpha'.\beta}{1.\gamma}y+X_2\frac{\alpha'(\alpha'+1)}{1.2}\frac{\beta(\beta+1)}{\gamma(\gamma+1)}y^2+\ldots.$$

$$=Y_0+Y_1\frac{\alpha.\beta}{1.\gamma}x+Y_2\frac{\alpha(\alpha+1)\beta(\beta+1)}{1.2.\gamma(\gamma+1)}x^2+\ldots.$$

or $F=\sum_{s'} X_{s'}y^{s'}=\sum_{s} Y_{s}x^{s}$.

Now $F(\alpha, \beta, \gamma, x)$ satisfies the equation

$$x(1-x)\frac{d^2z}{dx^2}+(\gamma-(\alpha+\beta+1)x)\frac{dz}{dx}-\alpha\beta z=0,$$

$$\therefore x(1-x)\frac{\partial^2 X_{s'}}{\partial x^2}+((\gamma+s')-(\alpha+\beta+s'+1)x)\frac{\partial X_{s'}}{\partial x}-\alpha(\beta+s')X_{s'}=0$$

Multiply this by $y^{s'}$ and sum with respect to s'; writing z for $F(\alpha, \alpha', \beta, \gamma, x, y)$ we obtain

$$x(1-x)\frac{\partial^2 z}{\partial x^2}+(\gamma-(\alpha+\beta+1)x)\frac{\partial z}{\partial x}-\alpha\beta z+(1-x)\Sigma s'\frac{\partial}{\partial x}(X_{s'}y^{s'})$$
$$-\alpha\Sigma s'X_{s'}y^{s'}=0.$$

But $y\frac{\partial z}{\partial y}=\Sigma s'X_{s'}y^{s'}$,

$$\therefore x(1-x)\frac{\partial^2 z}{\partial x^2}+(\gamma-(\alpha+\beta+1)x)\frac{\partial z}{\partial x}-\alpha\beta z+(1-x)\frac{\partial}{\partial x}\left(y\frac{\partial z}{\partial y}\right)-\alpha y\frac{\partial z}{\partial y}=0;$$

or finally z satisfies the two partial differential equations

$$x(1-x)\frac{\partial^2 z}{\partial x^2}+\{\gamma-(\alpha+\beta+1)x\}\frac{\partial z}{\partial x}+y(1-x)\frac{\partial^2 z}{\partial x\partial y}-\alpha y\frac{\partial z}{\partial y}-\alpha\beta z=0 \quad . \quad (1)$$

$$y(1-y)\frac{\partial^2 z}{\partial y^2}+\{\gamma-(\alpha'+\beta+1)\}\frac{\partial z}{\partial y}+x(1-y)\frac{\partial^2 z}{\partial x\partial y}-\alpha' x\frac{\partial z}{\partial x}-\alpha'\beta z=0 \quad . \quad (2)$$

§ 5. *The moments of* $F_1(\alpha, \alpha_1, \beta, \gamma)$. We imagine an ordinate of magnitude $z(s, s')$ erected at the point $x=cs$, $y=c's'$, where c, c' are constants to be determined in the fitting. Let V be the volume of the polyhedron of which the tops of the ordinates are vertices, *i. e.* a polyhedron made up of elementary prisms on base cc', and of height $z(s, s')$, and let $p_{tt'}$ denote the (t, t')th product moment about the centroid vertical, *the elements of volume of the polyhedron being concentrated along the ordinates.* Let $p'_{tt'}$ be the corresponding moment about the planes

$$x=-c, \quad y=-c'.$$

Thus $V = cc'F_1 = cc'\Sigma\Sigma A_{s,s'}$ say,

$$p'_{tt'}V = cc'\Sigma\Sigma A_{s,s'}(s+1)^t(s'+1)^{t'}c^t c'^{t'}.$$

Let $F(\alpha, \alpha', \beta, \gamma, x, y) = \Sigma\Sigma A_{ss'}x^s y^{s'}$

$$= \chi_{00},$$

and let $\chi_{t,t'} = \theta^t\phi^{t'}\chi_{00}$,

where $$\theta(u) = \frac{\partial}{\partial x}(xu) \text{ and } \phi(u) = \frac{\partial}{\partial y}(yu),$$

so that

$$\theta^t\phi^{t'}(x^s y^{s'}) = (s+1)^t(s'+1)^{t'}x^s y^{s'} \quad . \quad . \quad . \quad . \quad (3)$$

We have then

$$F_1 = (\chi_{00})_{x=y=1} = \frac{V}{cc'}$$

and

$$p'_{tt'}V = cc' \cdot c^t c'^{t'}(\chi_{t,t'})_{x=y=1},$$

$$\therefore \quad p'_{tt'} = c^t c'^{t'}\left(\frac{\chi_{t,t'}}{\chi_{00}}\right)_{x=y=1} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

χ_{00} satisfies the differential equations (1) and (2).

$$\chi_{10} = \theta z = x\frac{\partial z}{\partial x} + z, \quad \therefore \; x\frac{\partial z}{\partial x} = \chi_{10} - \chi_{00}.$$

Similarly $$y\frac{\partial z}{\partial y} = \chi_{01} - \chi_{00},$$

$$\chi_{20} = \frac{\partial}{\partial x}\left(x^2\frac{\partial z}{\partial x} + xz\right) = 3x\frac{\partial z}{\partial x} + x^2\frac{\partial^2 z}{\partial x^2} + z,$$

$$\chi_{11} = \frac{\partial}{\partial y}\left(yx\frac{\partial z}{\partial x} + yz\right) = x\frac{\partial z}{\partial x} + z + xy\frac{\partial^2 z}{\partial x\partial y} + y\frac{\partial z}{\partial y},$$

$$\chi_{02} = \frac{\partial}{\partial y}\left(y^2\frac{\partial z}{\partial y} + yz\right) = 3y\frac{\partial z}{\partial y} + y^2\frac{\partial^2 z}{\partial y^2} + z.$$

Solving these equations, we find

$$x^2\frac{\partial^2 z}{\partial x^2} = \chi_{20} - 3\chi_{10} + 2\chi_{00},$$

$$y^2\frac{\partial^2 z}{\partial y^2} = \chi_{02} - 3\chi_{01} + 2\chi_{00},$$

$$xy\frac{\partial^2 z}{\partial x\partial y} = \chi_{02} - 3\chi_{01} + 2\chi_{00}.$$

Multiply equations (1) and (2) by x, y respectively after writing $m_1 = \alpha + \beta$, $m_1' = \alpha' + \beta$, $m_2 = \alpha\beta$, $m_2' = \alpha'\beta$, and noting that $\gamma - \alpha - \alpha' - \beta - 1 = n$.

The equations (1), (2) lead to the following :

$$(1-x)x^2\frac{\partial^2 z}{\partial x^2}+(1-x)xy\frac{\partial^2 z}{\partial x\partial y}+(\gamma-\overline{m_1+1}x)\frac{\partial z}{\partial x}-\alpha xy\frac{\partial z}{\partial y}-m_2xz=0,$$

$$(1-y)y^2\frac{\partial^2 z}{\partial y^2}+(1-y)xy\frac{\partial^2 z}{\partial x\partial y}+(\gamma-\overline{m_1'+1}y)\frac{\partial z}{\partial y}-\alpha' xy\frac{\partial z}{\partial x}-m_2'yz=0;$$

and substituting for the differential coefficients their values as above, we get the difference equations

$$(1-x)(\chi_{20}-3\chi_{10}+2\chi_{00})+(1-x)[\chi_{11}-\chi_{01}-\chi_{10}+\chi_{00}]$$
$$+(\gamma-\overline{m_1+1}x)(\chi_{10}-\chi_{00})-\alpha x(\chi_{01}-\chi_{00})-m_2x\chi_{00}=0\ . \quad (5)$$

$$(1-y)(\chi_{02}-3\chi_{01}+2\chi_{00})+(1-y)(\chi_{11}-\chi_{10}-\chi_{01}+\chi_{00})$$
$$+(\gamma-\overline{m_1'+1}y)(\chi_{01}+\chi_{00})-a'y(\chi_{10}-\chi_{00})-m_2'y\chi_{00}=0\ . (6)$$

Put $x=y=1$,

$$\therefore\quad (\gamma-m_1-1)(\chi_{10}-\chi_{00})_1-\alpha(\chi_{01}-\chi_{00})_1-m_2(\chi_{00})_1=0,$$
$$(\gamma-m_1'-1)(\chi_{01}-\chi_{00})_1-\alpha'(\chi_{10}-\chi_{00})_1-m_2'(\chi_{00})_1=0,$$

or

$$(n-r')(\chi_{10})_1+r(\chi_{01})_1=(n-r'+r+m_2)(\chi_{00})_1$$
$$(n-r)(\chi_{01})_1+r'(\chi_{10})_1=(n-r+r'+m_2')(\chi_{00})_1.$$

Hence

$$\left(\frac{\chi_{10}}{\chi_{00}}\right)_{x=y=1}=\frac{n^2-nr'-nr+(n-r)m_2-rm_2'}{n(n-r-r')}$$

or

$$p'_{10}=c\left(1+\frac{nm_2-r(m_2+m_2')}{n(n-r-r')}\right)=c(1+rq).\quad .\quad (7)$$

Similarly

$$p'_{01}=c'(1+r'q)\quad .\ .\ .\ .\ .\ .\ .\quad (8)$$

These values agree with those found by Pearson (Phil. Mag. 1899, p. 238, eqn. 7). This must be true in general, *i.e.* the "marginal" moments p'_{10}, p'_{20}, ... p'_{t0} must have the same values as Pearson's ν_1, ν_2, ... ν_t found for the single hypergeometrical series.

For let

$$S_s=A_{s0}+A_{s1}+A_{s2}\ldots+A_{ss'}$$
$$=\frac{[\alpha]_s[\beta]_s}{s!\,[\gamma]_s}+\ldots\ldots+\frac{[\alpha]_s[\alpha']_{s'}[\beta]_{s+s'}}{s!\,s'!\,[\gamma]_{s+s'}}+\ldots$$
$$=\frac{[\alpha]_s[\beta]_s}{s!\,[\gamma]_s}\left\{1+\ldots+\frac{[\alpha']_{s'}[\beta+s]_{s'}}{s!\,[\gamma+s]_{s'}}+\ldots\right\}$$
$$=\frac{[\alpha]_s[\beta_s]}{s!\,[\gamma]_s}\,\frac{\Pi(\gamma+s-1)\Pi(\gamma-\alpha'-\beta-1)}{\Pi(\gamma-\beta-1)\Pi(\gamma-\alpha'-1+s)}$$

which reduces to

$$S_s = \frac{(n-r)!}{(n-r-r')!} \frac{[\alpha]_s[\beta]_s}{s!\,[\gamma]_s} \frac{1}{[\gamma+s]_{r'}}$$

$$= \frac{(n-r)!}{(n-r-r')!} \frac{1}{[\gamma]_{r'}} \cdot \frac{[\alpha]_s[\beta]_s}{s!\,[\gamma+r']_s}.$$

$$\therefore \qquad p'_{t0}\,V = cc'[S_0c^t + S_1(2c)^t + \ldots]$$

or
$$p'_{t0} = c^t\left(\frac{S_0 + S_1(2)^t + S_2(3)^t + \ldots}{S_0 + S_1 + S_2 + \ldots.}\right)$$

$$= c^t \frac{\left[1 + \dfrac{\alpha\beta(2)^t}{1(\gamma+r')} + \dfrac{\alpha(\alpha+1)\beta(\beta+1)(3)^t}{1\,.\,2\,.\,(\gamma+r')(\gamma+r'+1)} + \ldots\right]}{1 + \dfrac{\alpha\beta}{1\,.\,(\gamma+r')} + \ldots.}.$$

Now our α = Pearson's α,

our β = Pearson's β,

and our $\gamma + r'$ = Pearson's γ;

$\therefore$ our p'_{t0} = Pearson's ν_t.

§ 6. To find the higher moments and product moments we write equations (5) (6) in the form

$$(1-x)[\chi_{20} + \chi_{11} + \chi_{10}(m_1 - 3) - \chi_{01}(r+1) + \chi_{00}(m_2 - m_1 + r + 2)]$$
$$+ (n-r')\chi_{10} + r\chi_{01} + \chi_{00}(-r - m_2 + r' - n) = 0 \quad . \quad . \quad (9)$$

$$(1-y)[\chi_{02} + \chi_{11} + \chi_{01}(m_1' - 3) - \chi_{10}(r'+1) + \chi_{00}(m_2' - m_1' + r' + 2)]$$
$$+ (n-r)\chi_{01} + r'\chi_{10} + \chi_{00}(-r' - m_2' + r - n) = 0, \quad . \quad . \quad (10)$$

or
$$(1-x)P + Q = 0, \quad . \quad . \quad . \quad . \quad . \quad . \quad (11)$$
$$(1-y)R + S = 0. \quad . \quad . \quad . \quad . \quad . \quad . \quad (12)$$

Now
$$\theta(xu) = x(1+\theta)u,$$
$$\theta^k(xu) = x(1+\theta)^k u,$$

and $\therefore$ $\theta^k(1-x)u = [\theta^k - (1+\theta^k]u + (1-x)(1+\theta)^k u.$

Similarly $\phi^k(1-y)u = [\phi^k - (1+\phi)^k]u + (1-y)(1+\phi)^k u.$

Hence

$$\theta^h\phi^k[(1-x)u] = [(\theta^h - (1+\theta)^h]\phi^k u + (1-x)(1+\theta)^h\phi^k u$$
$$\theta^h\phi^k[(1-y)u] = [\phi^k - (1+\phi)^k]\theta^h u + (1-y)(1+\phi)^k\theta^h u.$$

$$\therefore \qquad [\theta^h\phi^k(1-x)u]_{x=y=1} = \{[\theta^h - (1+\theta)^h]\phi^k u\}_1 \quad . \quad (13)$$
$$[\theta^h\phi^k(1-y)u]_{x=y=1} = \{\phi^k - (1+\phi)^k)\theta^h u\}_1. \quad . \quad (14)$$

Also $\theta\chi_{tt'} = \chi_{t+1,t'}, \quad \phi\chi_{tt'} = \chi_{t,t'+1}.$

Operating on each of equations 11, 12 with θ and ϕ and putting $x=y=1$ in the result, we obtain, since by (13), (14)

$$[\theta(\overline{1-x}\mathrm{P}+\mathrm{Q})]_1=[-\mathrm{P}+\theta\mathrm{Q}]_1$$
$$[\phi(\overline{1-x}\mathrm{P}+\mathrm{Q}]_1=(\phi\mathrm{Q})_1$$
$$[\theta(\overline{1-y}\mathrm{R}+\mathrm{S})]_1=(\theta\mathrm{S})_1$$
$$[\phi(\overline{1-y}\mathrm{R}+\mathrm{S})]_1=[-\mathrm{R}+\phi\mathrm{S}]_1,$$

the following equations :—

$$\left(\begin{array}{ll} -\chi_{20}-\chi_{11}-\chi_{10}(m_1-3)+\chi_{01}(r+1)-\chi_{00}(m_2-m_1+r+2) & \\ +(n-r')\chi_{20}+r\chi_{11}+\chi_{10}(-r+m_2+r'-n)=0, \quad . \quad . \quad . \quad . \quad . & . \quad (15) \\ (n-r')\chi_{11}+r\chi_{02}+\chi_{01}(-r-m_2+r'-n)=0, \quad . \quad . \quad . \quad . \quad . & . \quad (16) \\ (n-r)\chi_{11}+r'\chi_{20}+\chi_{10}(-r'-m_2'+r-n)=0, \quad . \quad . \quad . \quad . \quad . & . \quad (17) \\ -\chi_{02}-\chi_{11}-\chi_{01}(m_1'-3)+\chi_{10}(r'+1)-\chi_{00}(m_2'-m_1'+r'+2) & \\ +(n-r)\chi_{02}+r'\chi_{11}+\chi_{01}(-r'-m_2'+r-n)=0 \quad . \quad . \quad . \quad . \quad . & . \quad (18) \end{array}\right)_{x=y=1}$$

Solving these equations, we find after some reductions,

$$\left(\frac{\chi_{11}}{\chi_{00}}\right)_1=\frac{1}{n-1}[(1+rq)(1+r'q)n-q(rr'+r+r')-1], \quad . \quad (19)$$

$$\left(\frac{\chi_{20}}{\chi_{00}}\right)_1=\frac{1}{n-1}[n(1+3rq+r(r-1)q^2)-1-r(r+2)q], \quad . \quad (20)$$

and these are the values of $\frac{p'_{11}}{cc'}$ and $\frac{p'_{20}}{c^2}$.

We can transfer these to the mean, since $p_{20}=p'_{20}-p'^2_{10}$ and $p_{11}=p'_{11}-p'_{01}p'_{10}$.

The results are

$$p_{20}=\frac{c^2rq(1-q)(n-r)}{(n-1)}, \quad . \quad . \quad . \quad . \quad . \quad (21)$$

$$p_{02}=\frac{c'^2r'q(1-q)(n-r')}{(n-1)}, \quad . \quad . \quad . \quad . \quad (22)$$

$$p_{11}=-\frac{cc'rr'q(1-q)}{(n-1)}. \quad . \quad . \quad . \quad . \quad . \quad (23)$$

As shown in § 5 we may use Pearson's ν_3 for our p_{30}, whence

$$p_{30}=\frac{c^3qr(1-q)(1-2q)(n-r)(n-2r)}{(n-1)(n-2)}, \quad . \quad . \quad (24)$$

$$p_{03}=\frac{c'^3qr'(1-q)(1-2q)(n-r')(n-2r')}{(n-1)(n-2)}. \quad . \quad (25)$$

To find p_{21} and p_{12} we again apply (13) and (14), and obtain

$$[\phi\theta(\overline{1-x}\,\mathrm{P}+\mathrm{Q})]_1=[-\phi\mathrm{P}+\theta\phi\mathrm{Q}]_1, \quad . \quad . \quad (26)$$

$$[\theta\phi(\overline{1-y}\,\mathrm{R}+\mathrm{S})]_1=[-\theta\mathrm{R}+\theta\phi\mathrm{S}]_1, \quad . \quad . \quad (27)$$

leading to the equations

$$x=y=1\left\{\begin{array}{ll}-\chi_{21}-\chi_{12}-\chi_{11}(m_1-3)+\chi_{02}(r+1)-\chi_{01}(m_2-m_1+r+2) & \\ \quad+(n-r')\chi_{21}+r\chi_{12}+\chi_{11}(-r-m_2+r'-n)=0. & . \quad . \quad (28)\\ -\chi_{12}-\chi_{21}-\chi_{11}(m_1'-3)+\chi_{20}(r'+1)-\chi_{10}(m_2'-m_1'+r'+2) & \\ \quad+(n-r)\chi_{12}+r'\chi_{21}+\chi_{11}(-r'-m_2'+r-n)=0. & . \quad . \quad (29)\end{array}\right.$$

These equations are equivalent to the following for the "raw" moments

$$\frac{p'_{21}}{c^2c'}(n-r'-1)+\frac{p'_{12}}{cc'^2}(r-1)+\frac{p'_{11}}{cc'}(r'-n+3+qn-rqn)+\frac{p'_{02}}{c'^2}(r+1)$$
$$-\frac{p'_{01}}{c'}(r'qn+qn+2r'+2)=0, \quad . \quad (30)$$

$$\frac{p'_{21}}{c^2c'}(r'-1)+\frac{p'_{12}}{ec'^2}(n-r-1)+\frac{p'_{11}}{cc'}(r-n+3+qn-r'qn)+\frac{p'_{20}}{c^2}(r'+1)$$
$$-\frac{p'_{10}}{c}(rqn+qn+2r+2)=0. \quad . \quad . \quad (31)$$

Now if $d=(1+rq)c$, $d'=(1+r'q)c'$ denote the position of the mean, the raw moments can be expressed in terms of the moments about the mean by means of the equations

$$\left.\begin{array}{l}p'_{21}=p_{21}+2dp_{11}+d'p_{20}+d^2d\\ p'_{12}=p_{12}+2d'p_{11}+dp_{02}+dd'^2\\ p'_{11}=p_{11}+dd'\\ p'_{02}=p_{02}+d'^2\\ p'_{20}=p_{20}+d^2\end{array}\right\} \quad . \quad . \quad . \quad (32)$$

Making these substitutions in (30), (31) we get after rearranging

$$\frac{p_{21}}{c^2c'}(n-r'-1)+\frac{p_{12}}{cc'^2}(r-1)+\frac{p_{20}}{c^2c'}d'(n-r'-1)+p_{02}\left(\frac{d}{cc'^2}(r-1)+\frac{r+1}{c'^2}\right)$$
$$+p_{11}\left(\frac{2d}{c^2c'}(n-r'-1)+\frac{2d'}{cc'^2}(r-1)+\frac{r'-n+3+qn-rqn}{cc'}\right)$$
$$+\frac{d^2d'}{c^2c'}(n-r'-1)+\frac{dd'^2}{cc'^2}(r-1)+\frac{dd'}{cc'}(r'-n+3+qn-rqn)+\frac{d'^2}{c'^2}(r+1)$$
$$-\frac{d'}{c'}(rqn+qn+2r+2)=0, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (33)$$

and a similar equation with suffixes and accents inter-changed.

Putting $d=(1+rq)c$, $d'=(1+r'q)c'$, we obtain after some reduction

$$\frac{p_{21}}{c^2c'}(n-r'-1)+\frac{p_{12}}{cc'^2}(r-1)+\frac{p_{20}}{c^2}(1+r'q)(n-r'-1)+\frac{p_{02}}{c'^2}r(2+r-1q)$$

$$+\frac{p_{11}}{cc'}[q(nr-2r-2r'+n)+(n-1)-r'+2r]$$
$$-(1+r'q)qr(1-q)(n-r-r')=0, \quad . \quad . \quad (34)$$

$$\frac{p_{12}}{cc'^2}(n-r-1)+\frac{p_{21}}{c^2c'}(r'-1)+\frac{p_{02}}{c'^2}(1+rq)(n-r-1)+\frac{p_{20}}{c^2}r'(2+\overline{r'-1}q)$$

$$+\frac{p_{11}}{cc'}[q(nr'-2r'-2r+n)+(n-1)-r+2r']$$
$$-(1+rq)qr'(1-q)(n-r-r')=0. \quad . \quad . \quad (35)$$

We now substitute the previously obtained values for p_{20}, p_{02}, p_{11}, and after some more reduction obtain

$$\frac{p_{21}}{c^2c'}(n-r'-1)+\frac{p_{12}}{cc'^2}(r-1)=-\frac{rr'q(1-q)(1-2q)(n-r-r')}{n-1} \quad . \quad (36)$$

$$\frac{p_{21}}{c^2c'}(r'-1)+\frac{p_{12}}{cc'^2}(n-r-1)=-\frac{rr'q(1-q)(1-2q)(n-r-r')}{n-1} \quad . \quad (37)$$

Solving these equations we find

$$p_{21}=-\frac{c^2c'rr'q(1-q)(1-2q)(n-2r)}{(n-1)(n-2)} \quad . \quad . \quad (38)$$

$$p_{12}=-\frac{cc'^2rr'q(1-q)(1-2q)(n-2r')}{(n-1)(n-2)} \quad . \quad . \quad (39)$$

Four additional marginal moments will be required, these can be deduced from Pearson *l. c.*:

$$p_{30}=\frac{c^3rq(1-q)(1-2q)(n-r)(n-2r)}{(n-1)(n-2)} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (40)$$

$$p_{03}=\frac{c'^3r'q(1-q)(1-2q)(n-r')(n-2r')}{(n-1)(n-2)} \quad . \quad . \quad . \quad . \quad . \quad . \quad (41)$$

$$p_{40}=\frac{c^4q(1-q)(n-r)}{(n-1)(n-2)(n-3)}\left\{\begin{array}{l} n^2(1+3q\overline{1-q}\,\overline{r-2}) \\ +n(3q\overline{1-q}r\overline{6-r}+1-6r \\ +6r^2(1-3q\overline{1-q}) \quad\quad) \end{array}\right\} (42)$$

$$p_{04}=\frac{c'^4q(1-q)(n-r')}{(n-1)(n-2)(n-3)}\left\{\begin{array}{l} n^2(1+3q\overline{1-q}\,\overline{r'-2}) \\ +n(3q\overline{1-q}r'\overline{6-r'}+1-6r') \\ +6r'^2(1-3q\overline{1-q}) \end{array}\right\} (43)$$

If the values for the moments be examined it will be seen that a frequency distribution is not hypergeometrical in type unless certain conditions are fulfilled. For example

$$p_{21}\,p_{20}\,p_{03} \equiv p_{12}\,p_{02}\,p_{30} \quad . \quad . \quad . \quad . \quad (44)$$

$$p_{02}\,p_{20}\,p_{21}\,p_{12} \equiv p_{11}{}^2\,p_{03}\,p_{30} \quad . \quad . \quad . \quad . \quad (45)$$

Other identities will appear later. This explains why we have to use more moments than there are unknowns to be determined.

§7. *The solution.* Let ρ denote the correlation so that

$$\rho^2 = \frac{p_{11}{}^2}{p_{20}\,p_{02}} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (46)$$

and let $$\lambda^2 = \frac{p_{20}\,p_{12}{}^2}{p_{20}\,p_{21}{}^2} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (47)$$

From (21), (22), (23) we get

$$\frac{1}{\rho^2} = \frac{p_{20}\,p_{02}}{p_{11}{}^2} = \frac{(n-r)(n-r')}{rr'} \quad . \quad . \quad . \quad . \quad (48)$$

and using (38), (39)

$$\lambda^2 = \frac{r}{r'}\,\frac{n-r}{n-r'}\left(\frac{n-2r'}{n-2r}\right)^2 \quad . \quad . \quad . \quad . \quad . \quad . \quad (49)$$

Put $\frac{n}{r}-1=\xi$, $\frac{n}{r'}-1=\eta$, then

$$\xi\eta = \frac{1}{\rho^2} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (50)$$

and $$\frac{\xi(\eta-1)^2}{\eta(\xi-1)^2} = \lambda^2, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (51)$$

whence $$\xi = \frac{1+\rho\lambda}{\rho(\rho+\lambda)}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (52)$$

λ being taken of the same sign as ρ.

Equations (50) and (52) give ξ and η.

Next from (21), (22), (40), and (41) we have

$$\frac{p_{30}{}^2}{p_{20}{}^3} = \frac{(1-2q)^2}{q(1-q)}\,\frac{n-1}{(n-2)^2}\,\frac{(n-2r)^2}{r(n-r)} \quad . \quad . \quad . \quad (53)$$

$$\frac{p_{03}{}^2}{p_{02}{}^3} = \frac{(1-2q)^2}{q(1-q)}\,\frac{n-1}{(n-2)^2}\,\frac{(n-2r')^2}{r'(n-r')}, \quad . \quad . \quad (54)$$

while from (38) and (39)

$$\frac{p_{21}^2}{p_{20}^2 p_{02}} = \frac{(1-2q)^2}{q(1-q)} \frac{n-1}{(n-2)^2} \frac{(n-2r)^2 r'}{(n-r)^2(n-r')} \quad . \quad (55)$$

$$\frac{p_{12}^2}{p_{02}^2 p_{20}} = \frac{(1-2q)^2}{q(1-q)} \frac{n-1}{(n-2)^2} \frac{(n-2r')^2 r}{(n-r')^2(n-r)} . \quad (56)$$

These last four equations involve three more identities as we have four values for the product

$$\frac{(1-2q)^2 (n-1)}{q(1-q)(n-2)^2}.$$

The best value of this quantity which plays an important part in the numerical fitting is found from (55) and (56) which involve the body of the table, and we shall take

$$\frac{(1-2q)^2 (n-1)}{q(1-q)(n-2)^2} = \frac{p_{12} p_{21} \xi^{3/2} \eta^{3/2}}{\sigma^3 \sigma'^3 (\xi-1)(\eta-1)}, \quad . \quad . \quad (57)$$

which is obtained by multiplying equations (55) and (56) and writing as is usual σ and σ' for $\sqrt{p_{20}}$ and $\sqrt{p_{02}}$.

If we remember that $\rho = p_{11}/\sigma\sigma'$ by (48), we can write (57) in the form

$$\frac{(1-2q)^2 (n-1)}{q(1-q)(n-2)^2} = \frac{p_{21} p_{12}}{p_{11}^3 (\xi-1)(\eta-1)} = \theta, \text{ say} \quad . \quad (58)$$

These identities necessitate the use of higher moments for the determination of n.

For this purpose we use (42), (43).

Following the usual notation we write

$$\beta_1 = \frac{p_{30}^2}{p_{20}^3}, \ \beta_2 = \frac{p_{40}}{p_{20}^2},$$

$$\beta_1' = \frac{p_{03}^2}{p_{02}^3}, \ \beta_2' = \frac{p_{04}}{p_{02}^2},$$

and as in Pearson's paper write

$$\left.\begin{aligned} \epsilon &= n(n-r)(1-q) \\ \epsilon' &= n(n-r')(1-q) \\ z_2 &= rqn\epsilon \\ z_2' &= r'qn\epsilon' \\ z_1 &= \epsilon + rqn \\ z_1' &= \epsilon' + r'qn \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad (59)$$

so that $\qquad \dfrac{\epsilon'}{\epsilon} = \dfrac{n-r'}{n-r}$ and $\dfrac{z_2'}{z_2} = \dfrac{r'}{r}\dfrac{n-r'}{n-r}.$

Hence (21), (22), (40), (41), (42), (43) give

$$z_2\left[\beta_2\frac{(n-2)(n-3)}{n-1}-(3n-6)\right]=n^4+n^3+6(z_1^2-n^2z_1)\ . \quad (60)$$

$$z_2\beta_1\frac{(n-2)^2}{n-1}=n^4+4(z_1^2-n^2z_1)\ . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (61)$$

$$z_2'\left[\beta_2'\frac{(n-2)(n-3)}{n-1}-(3n-6)\right]=n^4+n^3+6(z'^2-n^2z_1')\ . \quad . \quad . \quad (62)$$

$$z_2'\left[\beta_1'\frac{(n-2)^2}{n-1}\right]=n^4+4(z_1'^2-n^2z_1')\ . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (63)$$

Eliminating z_1, z_1' we obtain

$$z_2\left[3\beta_1\frac{(n-2)^2}{n-1}-2\beta_2\frac{(n-2)(n-3)}{n-1}+6n-12\right]=n^4-2n^3$$

$$z_2'\left[3\beta_1'\frac{(n-2)^2}{n-1}-2\beta_2'\frac{(n-2)(n-3)}{n-1}+6n-12\right]=n^4-2n^3.$$

But $\dfrac{z_2'}{z_2}=\dfrac{r'}{r}\dfrac{n-r'}{n-r}$ is known from (50), (52);

in fact $\dfrac{z_2'}{z_2}=\left(\dfrac{r'}{n}\right)^2\left(\dfrac{n}{r}\right)^2\dfrac{n-r'}{r'}\ \dfrac{r}{n-r}=\dfrac{(1+\xi)^2\eta}{(1+\eta)^2\xi}=\gamma$ say . (64)

$$\therefore\quad \frac{3\beta_1(n-2)-2\beta_2(n-3)+6n-6}{3\beta_1'(n-2)-2\beta_2'(n-3)+6n-6}=\gamma\ . \quad . \quad (65)$$

$$\text{or } n=6\left\{\frac{\gamma(\beta_1'-\beta_2'+1)-(\beta_1-\beta_2+1)}{\gamma(3\beta_1'-2\beta_2'+6)-(3\beta_1-2\beta_2+6)}\right\}\ . \quad . \quad (66)$$

n being known, q is given by (58) since

$$\frac{(1-2q)^2}{q(1-q)}=\theta\frac{(n-2)^2}{n-1}.$$

The values of r and r' are given by

$$r=\frac{n}{\xi+1},\ r'=\frac{n}{\eta+1}.$$

Equations (21) and (22) will give c and c' and (7), (8) will give the position of the mean.

Let N be the total number of observations, then to obtain the various ordinates of the hypergeometrical frequency polyhedron we must take the various terms of

$$\frac{N}{(n)_{r+r'}}\sum_s\sum_t\binom{r}{s}\binom{r'}{t}(qn)_{s+t}(pn)_{r+r'-s-t},$$

which is equivalent to the double series $\Delta F_1(\alpha, \alpha', \beta, \gamma)$ of §3.

§8. This solution fails for symmetrical distributions.

If $r=r'$, it follows that $\xi=\eta$ and from (64) that $\gamma=1$, so that the value of n given by (66) becomes

$$n=6\left\{\frac{(\beta_1'-\beta_2'+1)-(\beta_1-\beta_2+1)}{(3\beta_1'-2\beta_2'+6)-(3\beta_1-2\beta_2+6)}\right\} \quad . \quad . \quad (67)$$

Now for nearly symmetrical distributions β_1 would be nearly $=\beta_1'$ and $\beta_2=\beta_2'$, so that it would be the ratio of two very small numbers and liable to an exceedingly high error; in fact n is quite indeterminate in the exactly symmetrical case.

In this case, however, the solution can be completed by the use of marginal moments alone, and the formula given by Pearson for n (eqn. 32 *l. c.*) can be adopted. In fact let $\beta_3=p_{50}/p_{30}p_{20}$, $\beta_3'=p_{05}/p_{02}p_{03}$, we must have approximately

$$\frac{4\beta_3'-10\beta_2'+6\beta_1'+2}{\beta_3'-4\beta_2'+3\beta_1'+2}=\frac{4\beta_3-10\beta_2+6\beta_1+2}{\beta_3-4\beta_2+3\beta_1+2}=n.$$

We may therefore use

$$n=\sqrt{\frac{(4\beta_3'-10\beta_2'+6\beta_1'+2)(4\beta_3-10\beta_2+6\beta_1+2)}{(\beta_3'-4\beta_2'+3\beta_1'+2)(\beta_3-4\beta_2+3\beta_1+2)}}\ . \quad (68)$$

The other constants can be determined as before, or if we are at an early stage of the calculations convinced of the symmetry of the distribution, the heavy work of calculating the product moments may be omitted and the whole solution carried out by Pearson's method.

It may be observed that the solution by marginal moments only can be used even in unsymmetrical cases if (68) is approximately true. On the other hand, the higher moments being subject to very high variations in individual samples, a more accurate fit is to be expected from the use of p_{21}, p_{12}, p_{11} than from β_3, β_3' which depend on p_{50}, p_{05} and which replace these moments in the "marginal" solution.

By whichever of the above methods n is determined, its probable error, since it depends on the higher moments (the fourth or fifth), is much greater than the probable errors of $\frac{r}{n}$, $\frac{r'}{n}$ in the determination of which no moment higher than the third is employed. The quantities q, c, c' are derived from n and are subject to similar variations.

§9. Another solution involving no moments higher than the third can be obtained for distributions of discrete variables where we know *a priori* that approximately $c=c'=1$.

Let us assume that $cc'=1$. Multiplying equations (21), (22) we have

$$\frac{(n-1)^2 p_{20} p_{02}}{(n-r)(n-r')rr' q^2(1-q)^2} = 1,$$

or

$$q(1-q) = \frac{(n-1)\sigma\sigma'}{\sqrt{(n-r)(n-r')rr'}} \quad . \quad . \quad . \quad (69)$$

Now from (58) we have easily

$$q(1-q) = \frac{1}{4+\theta\dfrac{(n-2)^2}{n-1}} \quad . \quad . \quad . \quad . \quad . \quad (70)$$

Substituting $\frac{n}{\xi+1}$, $\frac{n}{\eta+1}$ for r, r' in (69) and equating the two values of $q(1-q)$ we find

$$[4(n-1)+\theta(n-2)^2]\sigma\sigma' = n^2\sqrt{\xi\eta} \,/\, (\xi+1)(\eta+1).$$

Or, since $\xi\eta = \frac{1}{\rho^2}$, n is given by

$$4(n-1)+\theta(n-2)^2 = \frac{n^2}{\rho\sigma\sigma'}\,\frac{1}{(\xi+1)(\eta+1)} \quad . \quad . \quad (71)$$

Now $\rho = p_{11}/\sigma\sigma'$.

Hence if

$$1/\phi = p_{11}(\xi+1)(\eta+1), \quad . \quad . \quad . \quad . \quad (72)$$

$$n^2(\theta-\phi)+n(4-4\theta)+4\theta-4=0 \quad . \quad . \quad . \quad (73)$$

The rest of the solution follows as in §§7 and 8. It is to be remarked that this solution applies to symmetrical distributions as well. For convenience in numerical applications a table is appended, giving the various constants in the order of calculation.

§10. *Numerical.*—In numerical applications, we begin by calculating the moments about some convenient origin and transferring them to the mean by the formulæ :—

$$p_{02} = p'_{02} - p'^2_{01},$$

$$p_{03} = p'_{03} - 3p'_{02}v'_{01} + 2p'^3_{01},$$

$$p_{04} = p'_{04} - 4p'_{03}p'_{01} + 6p'_{02}p'^2_{01} - 3p'^4_{01},$$

$$p_{05} = p'_{05} - 5p'_{04}p'_{01} + 10p'_{03}p'^2_{01} - 10p'_{02}p'^3_{01} + 4p'^4_{01},$$

$$p_{11} = p'_{11} - p'_{01}p'_{10},$$

$$p_{21} = p'_{21} - p'_{20}p'_{01} - 2p'_{11}p'_{10} + 2p'^2_{10}p'_{01},$$

and similar formulæ obtained by interchanging the suffixes.

For all three forms of solutions we calculate

$$\sigma = \sqrt{p_{20}}, \quad \sigma' = \sqrt{p_{02}},$$
$$\rho = p_{11}/\sigma\sigma',$$
$$\lambda = \sigma p_{12}/\sigma' p_{21} \quad \text{(of same sign as } \rho),$$
$$\xi = (1+\rho\lambda)/\rho(\rho+\lambda), \quad \eta = \frac{1}{\rho^2\xi},$$
$$\theta = p_{21}p_{12}/p_{11}{}^3(\xi-1)(\eta-1).$$

n is determined by the methods of § 7, 8, and 9 by the group of equations A, B, C, respectively,

$$\left.\begin{aligned} &\beta_1 = p_{30}{}^2/p_{20}{}^3, \quad \beta_1' = p_{03}{}^2/p_{02}{}^3, \\ &\beta_2 = p_{40}/p_{20}{}^2, \quad \beta_2' = p_{04}/p_{02}{}^2, \\ &\gamma = (1+\xi)^2\eta/(1+\eta)^2\xi, \\ &\frac{n}{6} = \frac{\gamma(\beta_1'-\beta_2'+1)-(\beta_1-\beta_2+1)}{\gamma(3\beta_1'-2\beta_2'+6)-(3\beta_1-2\beta_2+6)} \end{aligned}\right\} \text{(A)}$$

$$\left.\begin{aligned} &\beta_3 = p_{50}/p_{20}p_{30}, \quad \beta_3' = p_{05}/p_{02}p_{03}, \\ &n^2 = \frac{(4\beta_3-10\beta_2+6\beta_1+2)(4\beta_3'-10\beta_2'+6\beta_1'+2)}{(\beta_3-4\beta_2+3\beta_1+2)(\beta_3'-4\beta_2'+3\beta_1'+2)} \end{aligned}\right\} \text{(B)}$$

$$\left.\begin{aligned} &1/\phi = p_{11}(\xi+1)(\eta+1), \\ &n^2(\theta-\phi)+n(4-4\theta)+4\theta-4 = 0 \end{aligned}\right\} \text{(C)}$$

$$1/q(1-q) = 4+\theta(n-2)^2/(n-1)$$
$$r = n/(\xi+1), \quad r' = n/(\eta+1).$$
$$c = \frac{\sigma\sqrt{n-1}}{\sqrt{r(n-r)q(1-q)}} \qquad c' = \frac{\sigma'\sqrt{n-1}}{\sqrt{r'(n-r')q(1-q)}}.$$

Equations (C) will in general give the best value of n when they are applicable, since the errors of the lower moments are less, but they may fail to give a real value of n. This is due to the fact that it is not always possible to fit an arbitrary distribution with a hypergeometric series on the assumption $cc' = 1$.

(A) and (B) on the other hand always give a real n, for right-hand side of (B) must always be positive, as the equation is only used for symmetrical distributions.

§11. *Numerical Example* 1.

As a first numerical example we will take the following table :—

	0	1	2	3	4
0				4	3
1			18	36	9
2		12	54	36	3
3	1	12	18	4	

This table is the theoretical frequency distribution of black balls in a draw of 4 balls followed by one of 3 out of a bag containing 4 white and 6 black balls, the draw being repeated 210 times.

Taking moments about (0, 0) we find

$$p'_{10} = 2{\cdot}4, \quad p'_{20} = 6{\cdot}4, \quad p'_{30} = 18{\cdot}4, \quad p'_{40} = 56{\cdot}1143,$$
$$p'_{01} = 1{\cdot}8, \quad p'_{02} = 3{\cdot}8, \quad p'_{03} = 8{\cdot}8, \quad p'_{04} = 21{\cdot}8,$$
$$p'_{11} = 4, \quad p'_{21} = 10, \quad p'_{12} = 8.$$

Transferring to the mean at 2·4, 1·8 we find

$$p_{20} = {\cdot}64, \quad p_{30} = -{\cdot}032, \quad p_{40} = 1{\cdot}1254857,$$
$$p_{02} = {\cdot}56, \quad p_{03} = -{\cdot}056, \quad p_{04} = {\cdot}8192,$$
$$p_{11} = -{\cdot}32, \quad p_{21} = {\cdot}016, \quad p_{12} = {\cdot}032.$$

Whence we obtain

$$\rho = -{\cdot}5345, \quad \beta_1 = {\cdot}00390625, \quad \beta_1' = {\cdot}01785714,$$
$$\beta_2 = 2{\cdot}747768, \quad \beta_2' = 2{\cdot}6122449.$$

The formulæ of §10 give

$$\lambda = 2{\cdot}138, \quad \xi = 1{\cdot}5, \quad \eta = 2{\cdot}3333, \quad \gamma = \tfrac{7}{8};$$

so that, if we use

$$n = 6\frac{\gamma(\beta_1' - \beta_2' + 1) - (\beta_1 - \beta_2 + 1)}{\gamma(3\beta_1' - 2\beta_2' + 6) - (3\beta_1 - 2\beta_2 + 6)},$$

we find
$$n = \frac{16{\cdot}741074}{1{\cdot}674106} = 10.$$

Hence $r=4,\ r'=3,\ \theta=\dfrac{3}{128}$,

and $\therefore\ \dfrac{(1-2q)^2}{q(1-q)}=\dfrac{1}{6}$, so that $q=\cdot 6$ or $\cdot 4$.

We take $q=\cdot 6$ for equation (41) shows that since p_{30} is negative $q>p$.

From (21) and (22) we find $c=c'=1$.

Finally

$$\Delta=\frac{\mathrm{N}}{n(n-1)(n-2)\ldots(n-r-r'+1)}=\frac{210}{10\cdot 9\ldots 5\cdot 4}=\frac{1}{2880}.$$

$\therefore$ the given table is equivalent to the series

$$\frac{1}{2880}\sum_s\sum_{s'}\binom{4}{s}\binom{3}{s'}(6)_{s+s'}(4)_{7-s-s'}.$$

Numerical Example 2.

The following table, based on the Registrar General's report for 1910, gives the distribution according to ages of the 235,252 bachelors and the same number of spinsters who intermarried in 1910 and whose ages were stated.

Taking an arbitrary origin at 22·5 years of age for each sex, we obtain the following values for the raw moments:—

$$p'_{10}=\cdot 5227798 \quad p'_{01}=\cdot 8860797 \quad p'_{11}=1\cdot 075566$$
$$p'_{20}=1\cdot 1860516 \quad p'_{02}=1\cdot 888826 \quad p'_{21}=2\cdot 7207547$$
$$p'_{30}=2\cdot 8226710 \quad p'_{03}=5\cdot 543315 \quad p'_{12}=3\cdot 1404579$$
$$p'_{40}=10\cdot 5105716 \quad p'_{04}=22\cdot 63546$$

Thus the mean age at marriage of spinsters is 25·11390, and of bachelors 26·93040. When transferring to the mean we diminish p_{20} by $\frac{1}{12}$ and p_{40} by $\frac{1}{2}p'_{20}-\frac{7}{240}$ to allow for the grouping of the material (*cf.* W. F. Sheppard, Proc. Lond. Math. Soc. vol. xxix. pp. 353–380). The moments about the mean are:—

$$p_{20}=\cdot 8294195,\quad p_{02}=1\cdot 020356,\quad p_{11}=\cdot 612341,\quad \sigma=\cdot 910725,$$
$$p_{30}=1\cdot 2482896,\quad p_{03}=1\cdot 9137522,\quad p_{21}=1\cdot 0295785,\quad \sigma'=1\cdot 010126,$$
$$p_{40}=5\cdot 9016221,\quad p_{04}=9\cdot 5141106,\quad p_{12}=1\cdot 067845.$$

The distribution is decidedly unsymmetrical. We will first attempt a fit by moments of 3rd order.

The constants are as follows:—

$$\rho=\cdot 673016,\quad \lambda=\cdot 935105,\quad \xi=1\cdot 505454,\quad \eta=1\cdot 46650,$$
$$\theta=20\cdot 308,\quad \phi=\cdot 26427;$$

whence $n^2-3\cdot 8532n+3\cdot 8532=0$,

giving an imaginary n. Thus it is impossible to fit this table with a series whose terms are spaced equidistantly with the entries of the table.

Spinsters.

Bachelors.	15–	20–	25–	30–	35–	40–	45–	50–	55–	60–	65–	70–	
15–	2119	1142	76	11	1								3349
20–	12180	66846	12850	1176	134	21		1	1				93209
25–	2653	40768	40005	6822	810	90	22	6					91176
30–	446	7771	13247	8046	1682	244	44	4		1			31485
35–	110	1549	3242	2905	1928	433	97	9		1			10274
40–	26	339	701	870	857	507	136	29	2	3	1		3471
45–	8	90	202	291	320	255	177	35	9				1387
50–	2	28	65	88	109	105	90	49	17	5			558
55–	1	8	16	21	21	33	47	24	14	8		1	194
60–	1		3	9	7	12	13	8	13	6	3	1	76
65–			2	1	2	4	7	5	6	2	10	1	40
70–		1	1		2	2	2		1	2	8	4	23
75–			1	1			1	1	1		1	4	10
	17546	118542	70411	20241	5873	1706	636	171	64	28	23	11	235252

If we complete the solution by means of equations (A) we find

$$\beta_1 = 2{\cdot}730913 \qquad \beta_1' = 3{\cdot}44769 \qquad \gamma = 1{\cdot}00513$$
$$\beta_2 = 8{\cdot}77876 \qquad \beta_2' = 9{\cdot}13828$$
$$n = 1{\cdot}4888$$
$$q = {\cdot}0745$$
$$r = {\cdot}59422$$
$$r' = {\cdot}60362$$
$$c = 3{\cdot}3259$$
$$c' = 3{\cdot}6793$$

The starting point of the series is at $c(1+rq)$ to the left and $c'(1+r'q)$ above the mean of our table, *i.e.*, at age 12·7474 for spinsters and 12·7079 for bachelors.

Numerical Example 3.

The following table gives the *actual* distributions in 25,000 deals of the trumps in the first two hands, in whist with ordinary shuffling *.

Distribution of 25,000.

Second Hand.	First Hand. 0	1	2	3	4	5	6	7	8	
0		3	22	38	76	65	26	12	4	246
1	5	43	159	380	531	358	153	54	4	1687
2	20	183	746	1458	1590	926	323	49	7	5302
3	42	360	1451	2297	2059	934	222	34	4	7403
4	54	542	1588	2048	1497	505	101	14	1	6350
5	45	378	906	926	506	139	25	3		2928
6	34	166	312	255	96	21	2			886
7	10	47	69	34	16	2				178
8	5	2	9	4						20
	215	1724	5262	7440	6371	2950	852	166	20	25,000

* Unpublished material of Professor Karl Pearson.

Arbitrary origin at (3, 3)

$p'_{10}= \cdot 24944$	$p'_{01}= \cdot 25048$
$p'_{20}=1\cdot 72352$	$p'_{02}=1\cdot 74592$
$p'_{30}=1\cdot 64960$	$p'_{03}=1\cdot 68592$
$p'_{40}=9\cdot 11360$	$p'_{04}=9\cdot 41008$
$p'_{50}=17\cdot 10464$	$p'_{05}=17\cdot 64208$

$$p'_{21}=\cdot 020404, \quad p'_{12}=\cdot 02244, \quad p'_{11}=-\cdot 49036.$$

Mean at $\bar{x}=3\cdot 24944$, $\bar{y}=3\cdot 25048$.

Transferring to mean

$p_{20}=1\cdot 66130$	$p_{02}=1\cdot 68318$	$\beta_1=\cdot 03333$	$\beta_1'=\cdot 034462$
$p_{30}=\cdot 39094$	$p_{03}=\cdot 40545$	$\beta_2=2\cdot 9347$	$\beta_2'=2\cdot 9530$
$p_{40}=8\cdot 0995$	$p_{04}=8\cdot 36635$	$\beta_3=10\cdot 009$	$\beta_3'=9\cdot 8864$
$p_{50}=6\cdot 5004$	$p_{05}=6\cdot 7460$		

$p_{21}=-\cdot 13242$

$p_{12}=-\cdot 13611$ $\qquad \rho=-\cdot 2559.$

$p_{11}=-\cdot 42788$

$$\lambda=-1\cdot 02152, \quad \xi=3\cdot 8588, \quad \eta=3\cdot 9570, \quad \gamma=\cdot 9878.$$

Using the value of n for a symmetrical distribution

$$n=\sqrt{\frac{(4\beta_3-10\beta_2+6\beta_1+2)(4\beta_3'-10\beta_2'+6\beta_1'+2)}{(\beta_3'-4\beta_2'+3\beta_1'+2)(\beta_3-4\beta_2+3\beta_1+2)}}=49\cdot 161.$$

we complete the solution by equations B and obtain

$$r=10\cdot 117, \quad r'=9\cdot 9159, \quad q=\cdot 25557, \quad c=1\cdot 0319, \quad c'=1\cdot 0465.$$

If we determine n from moments of the first 3 orders by equations (C), we obtain $\theta=\cdot 027218$, $\phi=\cdot 097035$, so that the equation

$$(\theta-\phi)n^2+(4-4\theta)n-(4-4\theta)=0$$

becomes

$$n^2-55\cdot 737n+55\cdot 737=0$$

giving

$$n=54\cdot 723$$
$$q=\cdot 24485$$
$$r=11\cdot 263$$
$$r'=11\cdot 040$$
$$c=\cdot 99436$$
$$c'=1\cdot 0090.$$

The starting point of this series is at ($\cdot 487$, $\cdot 486$).

The various terms of the double series are given by

$$z(ss')=5{\cdot}897\binom{11{\cdot}263}{s}\binom{11{\cdot}04}{s'}\frac{(13{\cdot}399)(12{\cdot}399)\ldots(14{\cdot}399-s-s')}{(20{\cdot}21)(21{\cdot}21)\ldots(19{\cdot}21+s+s')}.$$

[When $r+r'$ is not an integer the value of $Z(s, s')$ given in § 7 is written in the form

$$\frac{N(pn)_{r+r'}}{(n)^{r+}{}_{r'}}\frac{\binom{r}{s}\binom{r'}{s'}(qn)_{s+s'}}{[pn-r-r'+1]_{s+s'}}.$$

The factor $\frac{N(pn)_{r+r'}}{(n)'_{r+r'}}$ can be evaluated by Gamma Functions or by treating it as a constant which is to be determined by equating the total of the calculated values to N. Gamma functions were employed to obtain 5·897.]

In comparing the actual frequencies with the theoretical frequencies deduced from this formula, we must not forget that the starting points are different.

In the annexed table the actual frequencies are given in small type, the calculated terms of the series in large type. They cannot be compared directly as the start of the theoretical values is at (·487, ·486) and the c, c' differ slightly from unity.

Some idea of the goodness of fit can be obtained by an examination of the model of which photographs are given in Pl. VII.

The white rectangles represent the actual frequencies, and the black the calculated ones. The theoretical frequencies interpolate with the actual ones in a very continuous manner. This is particularly noticeable in diagonal sections where theoretical and actual frequency ordinates are very nearly coplanar.

§ 12. When the starting point of the double series is known *a priori*, the solution admits of much simplification. Consider the case of discrete data, like the whist table, where (with perfect shuffling) we would expect the starting point to be at (0, 0) and $c=c'=1$.

Let m, m' be the means of the actual data.

From (7) and (8) we obtain

$$c(1+rq)=c+m,$$

$$c'(1+r'q)=c'+m',$$

or $$rq=m,\quad r'q=m'. \quad \ldots\ldots\ldots \quad (74)$$

Second Hand.	0		1		2		3		4		5		6		7		8		9	First Hand.	
0	0		3		22		38		76		65		26		12		4			246	
		6		43		129		202		183		100		34		7		1			705
1	5		43		159		380		531		358		153		54		4			1687	
		44		290		752		1024		805		381		109		19		2			3426
2	20		183		746		1458		1590		926		323		49		7			5302	
		134		767		1743		2055		1387		556		132		18		1			6793
3	42		360		1451		2297		2059		934		222		34		4			7403	
		215		1070		2106		2130		1218		406		79		9		1			7234
4	54		542		1588		2048		1497		505		101		14		1			6350	
		201		866		1460		1252		596		161		25		2					4563
5	45		378		906		926		506		139		25		3					2928	
		114		422		603		430		167		36		4							1776
6	34		166		312		255		96		21		2							886	
		40		125		149		86		26		4									430
7	10		47		69		34		16		2									178	
		9		22		22		10		2											65
8	5		2		9		4													20	
		1		2		2		1													6
9																					
	215		1724		5262		7440		6371		2950		852		166		20			25,000	
		764		3607		6966		7190		4384		1644		383		55		5			24,998

25,000 deals ordinary shuffling, small figures actual frequencies.

Large figures, terms of double hypergeometric series with start at (·487, ·486).

$n = 54{\cdot}723.$
$q = {\cdot}24485.$
$r = 11{\cdot}263.$
$r' = 11{\cdot}040.$
$c = {\cdot}99436.$
$c' = 1{\cdot}0090.$
$N\Delta = 5{\cdot}897.$

By (21) $$p_{20}=\frac{m(1-q)(nq-m)}{q(n-1)},$$

and by (24)

$$p_{30}=\frac{m(1-q)(1-2q)(nq-m)(nq-2m)}{q^2(n-1)(n-2)}.$$

These equations lead to a quadratic for q,

$$q^2\left[2\frac{p_{20}}{m}-2\frac{p_{30}}{p_{20}}-2m\right]+q\left[2\frac{p_{30}}{p_{20}}-\frac{p_{20}}{m}-\frac{p_{30}}{m}-4p_{20}+\frac{p_{30}}{p_{20}}m+3m\right]$$
$$-\left[\frac{p_{30}m}{p_{20}}-2p_{20}+m\right]=0. \quad . \quad . \quad . \quad (75)$$

A similar quadratic can be obtained from the other marginal values. If q, q' are the values of q obtained from the quadratics we may adopt $\sqrt{qq'}$ for the value in the fitting; denoting this by Q, we have

$$r=\frac{m}{Q'},\quad r'=\frac{m'}{Q'},$$

and n is given by (23), since

$$p_{11}=-\frac{rr'q(1-q)}{n-1}.$$

A simplified solution of this character in which *a priori* values are assigned to certain constants has to be employed with great care. In the case of the 25000 deals at whist (ordinary shuffling) dealt with above, the two quadratics are

$$q^2-\cdot62330q+\cdot11628=0 \quad\text{and}\quad q^2-\cdot61221q+\cdot11217=0.$$

Both have imaginary roots. It is thus impossible to fit this table with a double hypergeometric series whose start is at (0, 0).

It is a matter of interest to observe that the roots are imaginary because p_{20} and p_{02} have values 1·66130 and 1·68318, differing by about 10 per cent. from the corresponding theoretical value 1·8640 deduced from the theory of chance on the hypothesis of perfect shuffling. The effect of the faulty shuffling is to diminish the "standard deviations" in the table, *i.e.* to crowd the table closer to the mean values. There is a deficiency of high frequencies and of very low frequencies, so that the actual table cannot be fitted with a theoretical double series which allows for the occurrence of "no trumps" in the hands of both partners. It will be observed that the general solution of the preceding paragraph gave a start at about (·5, ·5).

§ 13. Note to Phil. Mag., Feb. 1899, pp. 236–246.

In fitting a hypergeometrical series to a frequency polygon by the method described in Professor Pearson's paper, certain ambiguities of sign may be removed by the following considerations. The terms of the series (1) are the chances of selecting 1, 2, 3, &c. black balls out of a bag containing pn white and qn black balls in a draw of r balls. By equation (2) $m_1 = -r - qn$ must be negative, also in the real case $r < pn$ or qn $\therefore$ $-m_1 < n$ and α is $\therefore$ numerically less than β. Again, by (22) $\epsilon = n^2 + nm_1 + m_2$ $\therefore$ in general $\epsilon > m_2$. By (23) $\sqrt{\beta_1} = \sqrt{n-1}(2z_1 - n^2)/(n-2)\sqrt{z_2}$, but $\sqrt{\beta_1} = \mu_3/\sigma^3$ where σ is the positive square root of μ_2, $\therefore$ $z_1 \lesseqgtr \frac{n^2}{2}$ according as $\mu_3 \lesseqgtr 0$.

Hence (34) should be written

$$z_1 = \tfrac{1}{2}\left(n^2 \pm \sqrt{\frac{\beta_1 z_2 (n-2)^2}{n-1}}\right),$$

the $+$ sign occurring when $\mu_3 > 0$ and the $-$ sign when $\mu_3 < 0$.

(35) becomes:—m_2, ϵ are roots of $\zeta^2 - z_1\zeta + z_2 = 0$, *$\epsilon$ being the greater root*, and (37) becomes α, β are roots of $\zeta^2 - m_1\zeta + m_2 = 0$, *where α is numerically less than β.*

As an illustration we may apply the formulæ to the series

$$1 + 24 + 90 + 80 + 15,$$

which is really $F(-4, -6, 1, 1)$, *i. e.* we should obtain

$$\alpha = -4,\ \beta = -6,\ \gamma = 1,\ p = 0{\cdot}4,\ q = 0{\cdot}6,\ r = 4,\ n = 10,\ c = 1,\ d = 2{\cdot}4.$$

We find

$$\nu_1 = 2{\cdot}4,\ \mu_2 = {\cdot}64,\ \mu_3 = -{\cdot}032,\ \mu_4 = 1{\cdot}1254\tfrac{6}{7}.\ \mu_5 = -{\cdot}21961\tfrac{1}{7}.$$
$$\beta_1 = {\cdot}00390625,\ \beta_2 = 2{\cdot}747768,\ \beta_3 = 10{\cdot}72321.$$

whence (32) gives

$$n = \frac{4\beta_3 - 10\beta_2 + 6\beta_1 + 2}{\beta_3 - 4\beta_2 + 3\beta_1 + 2} = 10.$$

From (33)

$$z_2 = \frac{n^3(n-1)}{4(n-1) + 2\beta_2(n-2) - \beta_3(n-4)} = 576.$$

By (34)

$$z_1 = \tfrac{1}{2}n^2 - \sqrt{\frac{\beta_1 z_2 (n-2)^2}{n-1}},$$

the minus sign corresponding to the negative value of μ_3. Thus $z_1 = 48$.

By (35) m_2 and ϵ are the roots of

$$\zeta^2 - z_1\zeta + z_2 = 0 \quad \text{or} \quad \zeta^2 - 48\zeta + 576 = 0,$$

so that $m_2 = \epsilon = 24$.

Using (36)

$$m_1 = \frac{\epsilon - m_2 - n^2}{n} = -10,$$

and α, β are given as roots of

$$\zeta^2 - m_1\zeta + m_2 = 0 \quad \text{or} \quad \zeta^2 + 10\zeta + 24 = 0$$

by (37).

Since α is the numerically smaller root, $\alpha = -4$, $\beta = -6$.

The remaining constants are now given without any ambiguity by (38) to (41), viz.: $\gamma = n + \alpha + \beta + 1 = 1$,

$$r = -\alpha = 4,\ q = -\frac{\beta}{n} = \cdot 6,\ p = \frac{\gamma - \alpha - 1}{n} = \cdot 4,$$

$$c = n\sqrt{\frac{\mu_2(n-1)}{z_2}} = 1, \quad \text{and} \quad d = \frac{cm_2}{n} = 2\cdot 4.$$

The Series

$$F(-3, -6, 2, 1) = 7 + 63 + 105 + 35$$

leads to $n = 10$, $z_2 = 504$. The negative sign is required in the ambiguity which occurs in the determination of z_1 since $\mu_3 < 0$. ϵ and m are given by $\zeta^2 - 46\zeta + 504 = 0$ and ϵ must be taken $= 28$, $m_2 = 18$ since $\epsilon > m_2$. The equation for α, β is found to be $\zeta^2 + 9\zeta + 18 = 0$, so that $\alpha = -3$ and $\beta = -6$ in accordance with the rule $|\alpha| < |\beta|$.

The effect of this note is to make the fitting in the paper referred to determinate and unique.

The author wishes to express his gratitude for much valuable advice and kindly encouragement received from Prof. Pearson during the preparation of the present paper.

XLVII. *On the Surface-tensions of Liquids in contact with different Gases.* *By* ALLAN FERGUSON, *B.Sc.* (*Lond.*), *Assistant-Lecturer in Physics in the University College of North Wales, Bangor**.

THE question of the effect of the nature of the super-incumbent gas on the surface-tension of the liquid with which it is in contact does not appear to have been very exhaustively investigated. The only experiments with which

* Communicated by Prof. E. Taylor Jones.