MATHEMATICS

ON EIGHTFOLD PROBABILITY FUNCTIONS

BY

H. S. STEYN AND A. J. B. WIID *)

(Communicated by Prof. D. VAN DANTZIG at the meeting of September 28, 1957)

1. *Introduction*

It was shown by STEYN (1956) that for $u=t^2$ the series $F_1(a; b_1, b_2; c; t, u)$ which is well known as Appell's F_1 series, is a probability generating function generating interesting probability functions of factorial trinomial types. The same idea was used by STEYN (1957) to show that the series $F(a; b_1, b_2, b_3; c; t_1, t_2, t_3)$ generates for $t_1=t$, $t_2=u$ and $t_3=tu$ a system of probability functions of fourfold type. In the present paper attention will first be drawn to the system of trivariate probability functions which is generated by the hypergeometric series.

$$(1)\qquad \begin{cases} F \equiv F(a;\ b_1, b_2, \ldots, b_7;\ c;\ t_1, t_2, \ldots, t_7) = \\ \quad = \sum\limits_{r=0}^{\infty} \dfrac{(a+r-1)^{!r}}{(c+r-1)^{!r}} \sum\limits_{\Sigma r_i = r} \prod\limits_{i=1}^{7} \dfrac{(b_i+r_i-1)^{!r_i}\, t_i^{r_i}}{r_i!}, \end{cases}$$

where $t_1=t$, $t_2=u$, $t_3=v$, $t_4=tu$, $t_5=uv$, $t_6=vt$ and $t_7=tuv$. A few examples of eightfold types will then be given.

2. *The Moment Generating Function*

The series (1) is absolutely convergent for $|t|<1$, $|u|<1$, $|v|<1$, and converges for $t=1$, $u=1$, $v=1$ when $c-a-\sum\limits_{1}^{7} b_i>0$ (STEYN, 1955).

The trivariate probability function $f(x, y, z)$, given as the coefficient of $t^x u^y v^z$ in the expansion of $C\cdot F$ where $C^{-1}=F(a;\ b_1, \ldots, b_7;\ c;\ 1, \ldots, 1) = F(a;\ \sum\limits_{1}^{7} b_i;\ c;\ 1)$, will now be considered by using the properties of the series (1). The moment generating function $M=M(\alpha, \beta, \gamma)$ is obtained from (1) by substituting $t=e^{\alpha}$, $u=e^{\beta}$, $v=e^{\gamma}$ in $C\cdot F$. It is well-known that the moments follow from the moment generating function by differentiating and putting $\alpha=\beta=\gamma=0$. Using (1) and remembering that $t_i (i=1, \ldots, 7)$ are now functions of α, β and γ it follows that,

$$(2)\qquad \frac{\partial}{\partial\alpha} M = C\left\{\frac{\partial}{\partial t_1} F e^{\alpha} + \frac{\partial}{\partial t_4} F e^{\alpha+\beta} + \frac{\partial}{\partial t_6} F e^{\alpha+\gamma} + \frac{\partial}{\partial t_7} F e^{\alpha+\beta+\gamma}\right\}$$

*) South African Council for Scientific and Industrial Research, Pretoria.

so that

$$\mu'_{100} = \left[\frac{\partial}{\partial\alpha} M\right]_{\alpha=\beta=\gamma=0} = \frac{F(a+1; \sum_1^7 b_i+1; c+1; 1)\, a(b_1+b_4+b_6+b_7)}{F(a; \sum_1^7 b_i; c; 1)\, c}$$

$$= \frac{a(b_1+b_4+b_6+b_7)}{c-a-\sum_1^7 b_i - 1}$$

Similarly,

$$\mu'_{010} = \frac{a(b_2+b_4+b_5+b_7)}{c-a-\sum_1^7 b_i-1} \text{ and } \mu'_{001} = \frac{a(b_3+b_5+b_6+b_7)}{c-a-\sum_1^7 b_i-1}.$$

The second order moments can be obtained in the same way but will rather be derived independently when examples are considered in section 6.

3. *The Regression Equation*

It will now be shown that the regression equation of **x** on **y** and **z** will only be linear if a certain condition between the parameters exists.

Writing F_0 for $[F]_{\alpha=0} = F(a;\ b_1, b_2+b_4, b_3+b_6, b_5+b_7;\ c;\ 1, e^\beta, e^\gamma, e^{\beta+\gamma})$ and showing each time only those parameters in F_0 which are altered, it follows that

$$\left[\frac{1}{C} M\right]_{\alpha=0} = F_0$$

$$(3)\quad \begin{cases} \left[\frac{1}{C}\frac{\partial}{\partial\alpha} M\right]_{\alpha=0} = \frac{ab_1}{c} F_0(a+1, b_1+1, c+1) + \frac{ab_4}{c} F_0(a+1, b_2+b_4+1, c+1)\, e^\beta + \\ \quad + \frac{ab_6}{c} F_0(a+1, b_3+b_6+1, c+1)\, e^\gamma + \frac{ab_7}{c} F_0(a+1, b_5+b_7+1, c+1)\, e^{\beta+\gamma} \end{cases}$$

$$(4)\quad \begin{cases} \left[\frac{1}{C}\frac{\partial}{\partial\beta} M\right]_{\alpha=0} = \frac{a(b_2+b_4)}{c} F_0(a+1, b_2+b_4+1, c+1)\, e^\beta + \\ \quad + \frac{a(b_5+b_7)}{c} F_0(a+1, b_5+b_7+1, c+1)\, e^{\beta+\gamma} \end{cases}$$

$$(5)\quad \begin{cases} \left[\frac{1}{C}\frac{\partial}{\partial\gamma} M\right]_{\alpha=0} = \frac{a(b_3+b_6)}{c} F_0(a+1, b_3+b_6+1, c+1)\, e^\gamma + \\ \quad + \frac{a(b_5+b_7)}{c} F_0(a+1, b_5+b_7+1, c+1)\, e^{\beta+\gamma}. \end{cases}$$

Remembering that the first term on the r.h.s. of (3) has been obtained as in (2) after differentiation of (1) w.r.t. t_1 and that $\alpha=0$ implies $t_1=1$, and using the differential equation for the series (1) (STEYN, 1955, eq. 7) it clearly follows that,

$$\frac{ab_1}{c} F_0(a+1, b_1+1, c+1) = \frac{b_1}{c-a-b_1-1} \left\{ \frac{a(b_2+b_4)}{c} F_0(a+1, b_2+b_4+1, c+1)\, e^\beta + \right.$$
$$+ \frac{a(b_3+b_6)}{c} F_0(a+1, b_3+b_6+1, c+1)\, e^\gamma +$$
$$\left. + \frac{a(b_5+b_7)}{c} F_0(a+1, b_5+b_7+1, c+1)\, e^{\beta+\gamma} + a\, F_0 \right\}$$

so that on substituting in (3) from (4) and (5),

$$(6)\quad \begin{cases} \left[\frac{\partial}{\partial\alpha} M\right]_{\alpha=0} = \left\{\frac{b_1}{c-a-b_1-1} + \frac{b_4}{b_2+b_4}\right\} \left[\frac{\partial}{\partial\beta} M\right]_{\alpha=0} + \\ \qquad + \left\{\frac{b_1}{c-a-b_1-1} + \frac{b_6}{b_3+b_6}\right\} \cdot \left[\frac{\partial}{\partial\gamma} M\right]_{\alpha=0} + \frac{ab_1}{c-a-b_1-1} [M]_{\alpha=0} \end{cases}$$

when the coefficient of $e^{\beta+\gamma} F_0(a+1, b_5+b_7+1, c+1)$ is zero i.e.

$$(7)\qquad \frac{a(b_5+b_7)}{c} \left\{\frac{b_4}{b_2+b_4} + \frac{b_6}{b_3+b_6} - \frac{b_7}{b_5+b_7} + \frac{b_1}{c-a-b_1-1}\right\} = 0.$$

Excluding the cases where one of the parameters a or $b_1, \ldots, b_7$ is zero i.e. excluding also $b_5+b_7=0$ since this would imply that $b_7=0$, this condition for linear regression becomes

$$d = \frac{b_4}{b_2+b_4} + \frac{b_6}{b_3+b_6} - \frac{b_7}{b_5+b_7} + \frac{b_1}{c-a-b_1-1} = 0.$$

Under this condition the regression equation of **x** on **y** and **z** follows from the theorem proved by STEYN (1957) as

$$(8)\quad \tilde{x} = \left(\frac{b_1}{c-a-b_1-1} + \frac{b_4}{b_2+b_4}\right) y + \left(\frac{b_1}{c-a-b_1-1} + \frac{b_6}{b_3+b_6}\right) z + \frac{ab_1}{c-a-b_1-1}.$$

The condition (7) is also necessary for a linear regression equation of **x** on **y** and **z**, for if the regression is linear then using the theorem referred to above it must follow from the expressions (3) to (5) that

$$(9)\qquad \left[\frac{\partial}{\partial\alpha} M\right]_{\alpha=0} = k_1 \left[\frac{\partial M}{\partial\beta}\right]_{\alpha=0} + k_2 \left[\frac{\partial M}{\partial\gamma}\right]_{\alpha=0} + k_3 [M]_{\alpha=0}.$$

Assuming (9) to be an identity in β and γ, and thus in e^β and e^γ, equations in k_1, k_2 and k_3, can be written down by considering:

(i) the constant term i.e.

$$k_3 F(a, b_1; c; 1) = \frac{ab_1}{c} F(a+1, b_1+1; c+1; 1); \ F(a, b; c; 1)$$

being an ordinary hypergeometric series so that

$$k_3 = \frac{ab_1}{c-a-b_1-1};$$

(ii) the coefficient of e^β i.e.

$$k_1 \frac{a(b_2+b_4)}{c} F(a+1, b_1; c+1; 1) + \frac{ab_1}{c-a-b_1-1} \cdot \frac{a(b_2+b_4)}{c} \cdot$$
$$\cdot F(a+1, b_1; c+1; 1) =$$
$$= \frac{ab_1}{c} \frac{(a+1)(b_2+b_4)}{c+1} F(a+2, b_1+1; c+2; 1) + \frac{ab_4}{c} F(a+1, b_1; c+1; 1)$$

so that

$$k_1 = \frac{b_1}{c-a-b_1-1} + \frac{b_4}{b_2+b_4};$$

(iii) the coefficient of e^{γ} which gives

$$k_2 = \frac{b_1}{c-a-b_1-1} + \frac{b_6}{b_3+b_6}.$$

The coefficients k_1, k_2 and k_3 obtained in (i), (ii) and (iii) are the same as the coefficients in (6). Hence, if and only if the relation (6) exists can the regression be linear.

The condition (7) is therefore a necessary as well as a sufficient condition for the regression of **x** on **y** and **z** to be linear.

4. *Marginal Distributions*

Remembering that the p.g.f. of $\sum_x \sum_y f(x, y, z)$ follows from (1) by writing $t=1$, $u=1$ and noting that

$$(10) \quad \begin{cases} F(a;\ b_1, b_2, b_3, b_4, b_5, b_6, b_7;\ c;\ 1, 1, v, 1, v, v, v) \\ =F(a;\ b_1+b_2+b_4, b_3+b_6+b_5+b_7;\ c;\ 1, v) \\ =C'\ F(a;\ b_3+b_5+b_6+b_7;\ c-b_1-b_2-b_4;\ v) \end{cases}$$

where C' is a constant, it is clear that the marginal distributions are hypergeometric distributions.

5. *Limiting form of the probability function for large absolute values of the parameters*

Let the absolute values of the parameters $a, b_1, \ldots, b_7, c, c-a-\sum_1^7 b_i$ all be large but of the same order in these parameters, say $O(a)$. In section 2 it was already shown that the first order moments which are equal to the first order cumulants are all of $O(a)$. From (10) in the previous section it can easily be shown (cf. STEYN, 1951) that the variances of the marginal distributions σ_x^2, σ_y^2 and σ_z^2 are also of $O(a)$. It will now be shown that also the cumulant generating function

$$L(\alpha, \beta, \gamma) = \mu'_{100}\,\alpha + \mu'_{010}\,\beta + \mu'_{001}\,\gamma + \sigma_x^2 \frac{\alpha^2}{2!} + \ldots$$

is of $O(a)$.

The function F in (1) is (apart from a constant factor) a probability generating function and for $0 \leqslant t_i \leqslant 1$ all the terms in the expansion of F will be positive, so that,

$$1 \leqslant F \leqslant F(a;\ b_1, b_2, \ldots, b_7;\ c;\ 1, 1, \ldots, 1) = F(a;\ \sum_1^7 b_i;\ c;\ 1)$$

and for $-\infty < \alpha, \beta, \gamma \leqslant 0$,

$$1 \leqslant F(a;\ b_1, b_2, \ldots, b_7;\ c;\ e^{\alpha}, e^{\beta}, \ldots, e^{\alpha+\beta+\gamma}) \leqslant F(a;\ \sum_1^7 b_i;\ c;\ 1).$$

For c not a negative integer,

$$F(a;\ \sum_1^7 b_i;\ c;\ 1) = \frac{\Gamma(c)\,\Gamma(c-a-\sum_1^7 b_i)}{\Gamma(c-a)\,\Gamma(c-\sum_1^7 b_i)}.$$

Using the well-known asymptotic formulae for the Gamma function it now follows immediately that when the parameters are all of $O(a)$, then

$$L(\alpha, \beta, \gamma) = \log M(\alpha, \beta, \gamma) =$$
$$\log F(a;\ b_1, b_2, \ldots, b_7;\ c;\ e^{\alpha}, e^{\beta}, \ldots, e^{\alpha+\beta+\gamma}) - \log F(a;\ \sum_1^7 b_i;\ c;\ 1)$$

will be of $O(a)$.

For c a negative integer, it is assumed that a is also a negative integer such that the series terminates i.e. $c = -n, a = -m$ where $n > m > 0$. Clearly then, since $b > 0$,

$$F(a;\ \sum_1^7 b_i;\ c;\ 1) < \left(1 + \frac{b+m-1}{n-m-1}\right)^m,$$

so that,

$$\log F(a;\ \sum_1^7 b_i;\ c;\ 1) \text{ is again } O(a).$$

Hence all cumulants will be of $O(a)$. In standard units, therefore, the cumulant generating function will change to

$$L(\alpha, \beta, \gamma) = \tfrac{1}{2}\{\alpha^2 + \beta^2 + \gamma^2 + 2\varrho_{xy}\alpha\beta + 2\varrho_{yz}\beta\gamma + 2\varrho_{zx}\gamma\alpha\} + O(a^{-\frac{1}{2}}),$$

so that for large absolute values of the parameters the probability function will be approximately given by the trivariate normal probability function.

6. *Examples*

(i) *The Eightfold Hypergeometric or Eightfold Factorial-Binomial*

Consider the case of a finite population of N individuals which are initially divided in the proportions $p_{111}, p_{110}, p_{101}, p_{011}, p_{100}, p_{010}, p_{001}, p_{000}$, possessing respectively the characteristics EFG, $EF\bar{G}$, $E\bar{F}G$, $\bar{E}FG$, $E\bar{F}\bar{G}$, $\bar{E}F\bar{G}$, $\bar{E}\bar{F}G$, $\bar{E}\bar{F}\bar{G}$, (where $\bar{E}$ means not E). From this population a sample of n individuals is drawn without replacement and x, y and z are the total number of successes of E, F and G respectively.

$$p = p_{111} + p_{110} + p_{101} + p_{100}, \quad p' = p_{111} + p_{110} + p_{011} + p_{010},$$
$$p'' = p_{111} + p_{101} + p_{011} + p_{001}$$

are the total probabilities for E, F and G respectively.

It is easy to see that the probability generating function is in this case given by

$$\frac{1}{N!^n} \sum_{\Sigma r_{ijk}=n} \frac{n!\,(Np_{000})!^{r_{000}}\,(Np_{100})!^{r_{100}} \ldots (Np_{111})!^{r_{111}}\, t^r u^{r'} v^{r''}}{r_{000}!\, r_{100}! \ldots r_{111}!}$$

(where r, r' and r'' are defined similar to the p, p' and p'' above)

$$(11)\quad \begin{cases} = \dfrac{(Np_{000})!^n}{N!^n} F(-n;\ -Np_{100}, -Np_{010}, -Np_{001}, -Np_{110}, -Np_{011}, \\ \quad -Np_{101}, -Np_{111};\ Np_{000} - n + 1;\ t, u, v, tu, uv, tv, tuv). \end{cases}$$

From (6) above follows that the regression equation of **x** on **y** and **z** is linear when

$$\frac{p_{111}}{p_{111}+p_{011}}-\frac{p_{110}}{p_{110}+p_{010}}-\frac{p_{101}}{p_{101}+p_{001}}+\frac{p_{100}}{p_{100}+p_{000}}=0.$$

The marginal distribution $\sum_x \sum_y f(x, y, z)$ is clearly the well-known hypergeometric probability function and, by (10), is generated by,

$$\text{Constant} \cdot F(-n;\ -Np'';\ Nq''-n+1;\ v).$$

The moments follow from the moment generating function by differentiation but also directly from the factorial moment generating function obtained from (11) by substituting $t=1+\alpha$, $u=1+\beta$ and $v=1+\gamma$ thereby obtaining the series

$$\frac{1}{N!^n}\sum_{\Sigma r_{ijk}=n}\frac{n!\,(Np_{000})!^{r_{000}}\,(Np_{100})!^{r_{100}}\ldots(Np_{111})!^{r_{111}}}{r_{000}!\,r_{100}!\ldots r_{111}!}\cdot(1+\alpha)^r\,(1+\beta)^{r'}\,(1+\gamma)^{r''},$$

whence it follows, using brackets to indicate factorial moments obtained from the first and second degree coefficients in this expansion, that $\mu'_{(100)}$ (or $\bar{x}$, the mean of **x**) $=np$. Similarly $\bar{y}=np'$, $\bar{z}=np''$. Also $\mu'_{(200)}$, the coefficient of α^2 reduces to

$$\frac{n^{!2}\,(Np)^{!2}}{N^{!2}},$$

so that

$$\sigma_{\mathbf{x}}^2=\mu'_{(200)}+\mu'_{(100)}-\mu'^2_{(100)}$$
$$=\frac{(N-n)}{N-1}\,npq,$$
$$\sigma_{\mathbf{y}}^2=\frac{(N-n)}{N-1}\,np'q' \text{ and } \sigma_{\mathbf{z}}^2=\frac{(N-n)}{N-1}\,np''q''.$$

From $\mu'_{(110)}$ the coefficient of $\alpha\beta$, it follows after some simplification that the product moment about the mean,

$$\mu_{110}=\mu'_{(110)}-\mu'_{(100)}\,\mu'_{(010)}$$
$$=\frac{(N-n)}{N-1}\,n\,(p_{111}+p_{110}-pp'), \text{ giving}$$
$$\varrho_{xy}=\frac{\mu_{110}}{\sqrt{\mu_{200}\,\mu_{020}}}=\frac{p_{111}+p_{110}-pp'}{\sqrt{pq\,p'q'}}.$$

Similarly

$$\varrho_{yz}=\frac{p_{111}+p_{011}-p'p''}{\sqrt{p'q'p''q''}} \text{ and } \varrho_{xz}=\frac{p_{111}+p_{101}-pp''}{\sqrt{pq\,p''q''}}.$$

Under condition (7) and similar condition for y and z the regression equation can be written down from (8) but also in the standard form

$$\frac{\tilde{x}-\bar{x}}{\sigma_x}=\frac{\varrho_{xy}-\varrho_{xz}\,\varrho_{yz}}{1-\varrho_{yz}^2}\left(\frac{y-\bar{y}}{\sigma_y}\right)+\frac{\varrho_{xz}-\varrho_{xy}\,\varrho_{yz}}{1-\varrho_{yz}^2}\left(\frac{z-\bar{z}}{\sigma_z}\right).$$

(ii) *The Eightfold Negative Factorial-Binomial*

If in the population of example (i) the sampling is stopped after obtaining $(m+1)$ failures (i.e. $\bar{E}\bar{F}\bar{G}$), including the last trial, the probability generating function is given by

$$(12)\quad \begin{cases} (Np_{000})^{!(m+1)} \sum_{s=0}^{N-Np_{000}} \sum_{\Sigma r_{ijk}=s} \frac{(m+s)!\,(Np_{100})^{!r_{100}}(Np_{010})^{!r_{010}}\dots(Np_{111})^{!r_{111}}}{m!\,r_{100}!\,r_{010}!\dots r_{111}!} \cdot t^r u^{r'} v^{r''} = \\ \text{(where } r, r', r'' \text{ are defined as previously)} \\ = \frac{(Np_{000})^{!(m+1)}}{N^{!(m+1)}} F(m+1;\ -Np_{100}, -Np_{010}, -Np_{001}, -Np_{110}, -Np_{011}, - \\ -Np_{101}, -Np_{111};\ -N+m+1;\ t, u, v, tu, uv, tv, tuv). \end{cases}$$

In this example the regression equation of **x** on **y** and **z** will be linear (and can be written down from (8)), if,

$$\frac{Np_{111}}{Np_{111}+Np_{011}} + \frac{Np_{100}}{Np_{100}-N-1} - \frac{Np_{110}}{Np_{110}+Np_{010}} - \frac{Np_{101}}{Np_{101}+Np_{001}} = 0.$$

Similar conditions hold for the other regression equations.

The marginal distribution $\sum_x \sum_y f(x, y, z)$ are again by (10) generated by

$$\text{Const. } F(m+1;\ -Np'';\ -Np''-Np_{000}+m+1;\ v)$$

and is therefore an univariable negative factorial binomial distribution (STEYN, 1955) arising from a finite population containing Np'' individuals which produce successes and Np_{000} individuals which produce the $m+1$ failures. Hence the name eightfold negative factorial binomial for this example.

Again, from the factorial moment generating function obtained from (12) by substituting $t=1+\alpha$, $u=1+\beta$, $v=1+\gamma$ it follows that,

$$\bar{x} = \mu'_{(100)} = \frac{(m+1)Np}{Np_{000}+1},\ \bar{y} = \frac{(m+1)\,Np'}{Np_{000}+1},\ \bar{z} = \frac{(m+1)\,Np''}{Np_{000}+1}.$$

Also, $\mu'_{(200)}$ the coefficient of α^2 follows after some simplification as

$$\mu'_{(200)} = \frac{(m+2)^{!2}\,(Np)^{!2}}{(Np_{000}+2)^{!2}}$$

so that

$$\sigma_x^2 = \mu'_{(200)} + \mu'_{(100)} - (\mu'_{100})^2$$
$$= \frac{(m+1)\,Np\,(Np+Np_{000}+1)\,(Np_{000}-m)}{(Np_{000}+2)^{!2}\,(Np_{000}+1)}$$

and similarly for σ_y^2 and σ_z^2 by changing p to p' and p'' respectively. The coefficient of $\alpha\beta$, $\mu'_{(110)}$, reduces to

$$\mu'_{(110)} = \frac{(m+1)}{(Np_{000}+2)^{!2}} \{(m+2)\,N^2pp' - (Np_{111}+Np_{110}) + \\ + (Np_{111}+Np_{110})\,(Np_{000}+2)\},$$

so that the product moments about the mean follow as

$$\mu_{110} = \mu'_{(110)} - \mu'_{(100)}\,\mu'_{(010)}$$
$$= \frac{(m+1)\,(Np_{000}-m)}{(Np_{000}+2)^{!2}\,(Np_{000}+1)}\{N^2pp' + (Np_{111}+Np_{110})\,(Np_{000}+1)\}.$$

By cyclic interchanging of the indices similar expressions for μ_{011} and μ_{101} are obtained. The correlation coefficients are given by

$$\varrho_{xy} = \frac{NpNp' + (Np_{111}+Np_{110})\,(Np_{000}+1)}{\sqrt{NpNp'\,(Np+Np_{000}+1)\,(Np'+Np_{000}+1)}}$$

and similar expressions for ϱ_{yz} and ϱ_{xz}.

(iii) *Eightfold distributions of Eggenberger and Polya types*

It is easily seen that, if after each trial giving an individual with a certain characteristic, $\Delta(>0)$, individuals possessing that characteristic are added to the population, then in the above two examples of the positive and negative factorial eightfolds the only change will be that factors of the form $\frac{Np-r}{N-s}$ will have to be replaced by

$$\frac{Np+r\Delta}{N+s\Delta} = \frac{\frac{Np}{\Delta}+r}{\frac{N}{\Delta}+s}$$

thereby replacing $-N$ by N/Δ in the expressions obtained in example (i) and (ii) above.

(iv) *Limiting forms*

When in the examples (i) and (ii) above the size of the population is much larger than the size of the sample, the distributions derived in these examples will tend to those obtained from a constant population, i.e. to distributions obtained by sampling with replacement.

(a) The Eightfold Binomial

The limiting form of the probability generating function (11) when sampling is done from an infinite population (or with replacement from a finite population) is clearly given by

$$(13)\quad \begin{cases} \displaystyle\sum_{\Sigma r_{ijk}=n} \frac{n!\,p_{000}{}^{r_{000}}\,p_{100}{}^{r_{100}}\cdots p_{111}{}^{r_{111}}\,t^{r}\,u^{r'}\,v^{r''}}{r_{000}!\,r_{100}!\cdots r_{111}!} = \\ \quad = (p_{000}+p_{100}t+p_{010}u+p_{001}v+p_{110}tu+p_{011}uv+p_{101}tv+p_{111}tuv)^n, \end{cases}$$

with binomial marginal distributions ${}^nC_x\,p^x q^{n-x}$, ${}^nC_y p'^y q'^{n-y}$ and ${}^nC_z\,p''^z q''^{n-z}$ respectively.

The condition for linear regression is the same as that obtained for example (i). This condition can also be derived directly from (13) by substituting $t=e^{\alpha}$, $u=e^{\beta}$ and $v=e^{\gamma}$, and differentiating the moment generating function thus obtained once with respect to α, then w.r.t. β and w.r.t. γ and writing down the condition for a linear relationship between

$$\frac{\partial}{\partial\alpha}M,\ \frac{\partial}{\partial\beta}M \text{ and } \frac{\partial}{\partial\gamma}M \text{ when } \alpha = 0.$$

Also for this eightfold binomial,

$$\bar{x} = np,\ \bar{y} = np',\ \bar{z} = np'',\ \sigma_x^2 = npq,$$
$$\sigma_y^2 = np'q' \text{ and } \sigma_z^2 = np''q'' \text{ while } \varrho_{xy},\ \varrho_{yz} \text{ and } \varrho_{xz}$$

remain the same as for the eightfold factorial binomial.

(b) The Eightfold Negative Binomial

The limiting form of the probability function (12) when sampling is done from an infinite population (or with replacement from a finite population) is clearly given by

$$p_{000}^{m+1} \sum_{s=0}^{\infty} \sum_{\Sigma r_{ijk}=s} \frac{(m+s)!\, p_{100}{}^{r_{100}}\, p_{010}{}^{r_{010}} \dots p_{111}{}^{r_{111}}\, t^r u^{r'} v^{r''}}{m!\, r_{100}!\, r_{010}! \dots r_{111}!} =$$
$$= p_{000}^{m+1}\{1-(p_{100}t+p_{010}u+p_{001}v+p_{110}tu+p_{011}uv+p_{101}tv+p_{111}tuv)\}^{-m-1}$$

with negative binomial marginal distributions.

The regression equation of **x** on **y** and **z** will be linear when

$$\frac{p_{111}}{p_{111}+p_{011}} - \frac{p_{100}}{1-p_{100}} - \frac{p_{110}}{p_{110}+p_{010}} - \frac{p_{101}}{p_{101}+p_{001}} = 0$$

and the moments are given as

$$\bar{x} = (m+1)\, p/p_{000},\ \bar{y} = (m+1)\, p'/p_{000},\ \bar{z} = (m+1)\, p''/p_{000},$$
$$\sigma_x^2 = (m+1)\, p(p+p_{000})/p_{000}^2,\ \sigma_y^2 = (m+1)\, p'(p'+p_{000})/p_{000}^2$$
$$\sigma_z^2 = (m+1)\, p''(p''+p_{000})/p_{000}^2,\ \varrho_{xy} = \frac{pp'+p_{000}\,p_{111}+p_{000}\,p_{110}}{\sqrt{pp'(p+p_{000})(p'+p_{000})}}$$

and similar expressions for ϱ_{yz}, ϱ_{xz}.

7. *Generalisations*

In this article the writers have studied general eightfold types. It is, however, very clear that also 2^k-fold probability distributions can be considered by considering a generalised hypergeometric series such as (1) in $t_1, \dots, t_{2^k-1}$, where

$$t_1 = u_1,\ t_2 = u_2,\ \dots,\ t_k = u_k,\ t_{k+1} = u_1u_2,\ \dots,\ t_{2^k-1} = u_1u_2, \dots, u_k$$

as the generating function. In the extension of the various examples it will mean that the associated probabilities are $p_{ij\dots l}$ where each of the k suffixes is either 0 or 1. The marginal distributions will then be factorial binomial types and negative factorial-binomial types. Further h^k-fold probability distributions can be considered by using generating functions of multivariate hypergeometric type, which in the extensions of the special examples considered above, will give rise to multivariate distributions with factorial multinomials and negative factorial multinomials (STEYN, 1956) as marginal distributions.

The probability function of example (1) and its extensions were studied in further details in an unpublished thesis by STEYN (1947).

This paper is published with the permission of the South African Council for Scientific and Industrial Research.

BIBLIOGRAPHY

STEYN, H. S., On Discrete Multivariate Probability Functions of Hypergemetric Type. Proc. Kon. Ned. Akad. v. Wetensch., Amsterdam, **A 58**, No. 5, and Indag. Math. **17**, No. 5 (1955).

———, On the Univariable Series $F(t) \equiv F(a; b_1, b_2, \ldots, b_k; c; t, t^2, \ldots, t^k)$ and its Applications in Probability Theory, Proc. Kon. Ned. Akad. v. Wetensch., Amsterdam, **A 59**, No. 2 and Indag. Math., **18**, No. 2 (1956).

———, On Regression Properties of Discrete systems of Probability Functions, Proc. Kon. Ned. Akad. v. Wetensch., Amsterdam, **A 60**, No. 1 and Indag. Math. **19**, No. 1 (1957).

———, On Multivariate Sampling with and without Replacement, Unpublished Ph.D. thesis, University of Edinburgh, 1947.