STICHTING
MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM

STATISTISCHE AFDELING

Report S 266 A

(Prepublication)

On probability distributions arising from points on a graph

by

A.R. Bloemena

The Mathematical Centre at Amsterdam, founded the 11th of February 1946, is a non-profit institution aiming at the promotion of pure mathematics and its applications, and is sponsored by the Netherlands Government trough the Netherlands Organization for Pure Research (Z.W.O.) and the Central National Council for Applied Scientific Research in the Netherlands (T.N.O.), by the Municipality of Amsterdam and by several industries.

1. Introduction

Given a set of n points, numbered 1,...,n, and a n x n matrix M, with elements m_{ij}, satisfying

(1.1) $m_{ij} = m_{ji}$ $(i \neq j)$,

(1.2) $m_{ii} = 0$,

(1.3) for each i $m_{ij} \neq 0$ for at least one j, and

(1.4) $0 \leq m_{ij} < \infty$.

In the special case that all m_{ij} are integers, the set of points and the matrix M can be interpreted as a finite multigraph (cf. C. BERGE (1958), D. KOENIG (1936)), where the number of joins between point i and j is equal to m_{ij}. If $m_{ij} = 0$, this means in this case that there is no join between i and j. Assumption (1.2) states that there are no loops. Assumption (1.3) implies that no point is isolated.

From the n points two samples are taken. We shall consider two cases.

Case I "non free sampling": from the points 1,.. ,n r_1 and r_2 points are chosen at random without replacement $(r_1+r_2 \leq n)$. The r_1 points will be denoted as black (B) points, the r_2 points as white (W) ones, while finally the $n-r_1-r_2$ remaining points are the red (R) ones.

Case II "free sampling": n independent trials are performed, each trial resulting in the event B with probability p_1, in the event W with probability p_2, and in the event R with probability $1-p_1-p_2$. Point number i is alotted the colour indicated by the outcome of the i-th trial.

Consider the random variables $\underline{x}_{ij}^{(W)}$, $\underline{x}_{ij}^{(B)}$, $\underline{y}_{ij}$ (i,j=1,.. ,n) defined by

$$\underline{x}_{ii}^{(W)} = 0 \qquad \text{spr } 0,$$

$$\underline{x}_{ii}^{(B)} = 0 \qquad \text{spr } 0,$$

$$\underline{y}_{ii} = 0 \qquad \text{spr } 0,$$

and for $i \neq j$

$$\underline{x}_{ij}^{(W)} = \begin{cases} 1 & \text{if point i and j are both white} \\ 0 & \text{if not,} \end{cases}$$

$$\underline{x}_{ij}^{(B)} = \begin{cases} 1 & \text{if point i and j are both black} \\ 0 & \text{if not,} \end{cases}$$

$$\underline{y}_{ij} = \begin{cases} 1 & \text{if one of the points i and j is black and the other is white} \\ 0 & \text{if not.} \end{cases}$$

Obviously

$$\underline{x}_{ij}^{(W)} = \underline{x}_{ji}^{(W)},$$

$$\underline{x}_{ij}^{(B)} = \underline{x}_{ji}^{(B)},$$

and $\underline{y}_{ij} = \underline{y}_{ji}$.

Define

$$\underline{x}_W = \sum_{i,j} m_{ij} \underline{x}_{ij}^{(W)},$$

(1.5) $$\underline{x}_B = \sum_{i,j} m_{ij} \underline{x}_{ij}^{(B)},$$

$$\underline{y} = \sum_{ij} m_{ij} \underline{y}_{ij},$$

We also introduce a set of random variables $\underline{z}_{ij}$, taking the values 0 or 1, ($\underline{z}_{ii}$ = 0 spr 0), and a random variable

$$\underline{z} = \sum_{ij} m_{ij} \underline{z}_{ij} .$$

We attribute to $\underline{z}_{ij}$ and to $\underline{z}$ all properties that $\underline{x}_{ij}^{(B)}$, $\underline{x}_{ij}^{(W)}$, $\underline{y}_{ij}$ and $\underline{x}_W$, $\underline{x}_B$ and $\underline{y}$ have in common. Thus $\underline{z}$ is a generalization of $\underline{x}_W$, $\underline{x}_B$ and $\underline{y}$.

In the following we shall give results on the stochastic properties of $\underline{x}_W$, $\underline{x}_B$, $\underline{y}$ and $\underline{z}$. The proofs will be given in a forthcoming thesis.

2. Previous work on the subject

P.A.P. MORAN (1948) considers a "statistical map", equivalent to our graph for $m_{ij} = 0$ or 1, where the points are chosen by "free" and "non free" sampling. He gives for both cases the first and second moments of the number of black-black joins (thus for x_B) and the third and fourth moment for the case of free sampling. He proves the asymptotic normality of x_B and y (free sampling) for a rectangular twodimensional lattice, where there are joins between neighbouring points in the direction of both axis (cf. also P.A.P. MORAN (1947)).

There exists a large number of papers on the subject by P.V. KRISHNA IYER (1948-1953). He only deals with rectangular lattices, where neighbouring points are joined in the direction of both axis, but also diagonal joins are considered in a number of his papers. The results of KRISHNA IYER are mostly on the first four moments or cumulants, and statements about asymptotic normality.

A report by VAN EEDEN and BLOEMENA (1959) contains a number of exact results for rectangular lattices (non-free samples). The present report is an outgrowth of the last mentioned paper, which arose from a study of the distribution of a statistic, obtained in a psychological test.

Some older papers on the subject are by H. TODD (1940) and D.J. FINNEY (1947).

3. Some graphtheoretical notions

Consider a set S of points and a subset U of the set of all possible joins between these points. The combination (S,U) is usually called a graph. For a detailed treatment of theory of graphs, we refer to D. KOENIG (1936) and C. BERGE (1958).

In this report the word "graph" is used in two different ways. In section 1 it has been shown that in the case that m_{ij} are non negative integers, the set of points from which the samples are taken, and the matrix M can be interpreted as a graph. In the following we use the word "graph" in a different situation.

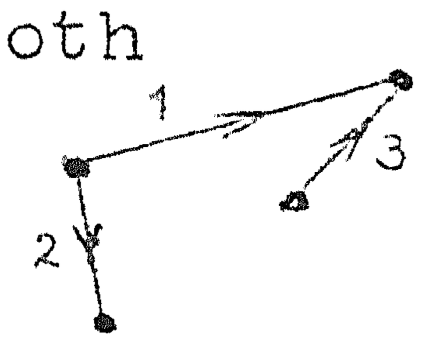
For the purposes of the following sections we use the word "graph" to denote a set of k oriented joins, labelled $J_1, \ldots, J_k$, between ℓ $(2 \leq \ell \leq 2k)$ points, such that no points are isolated (are not connected to at least one other point), and loops do not occur. Multiple joins are admitted.

A point to which join J_i is connected will be called the second point of J_i if the orientation of the join is towards the point; if not, it will be called the first point of J_i.

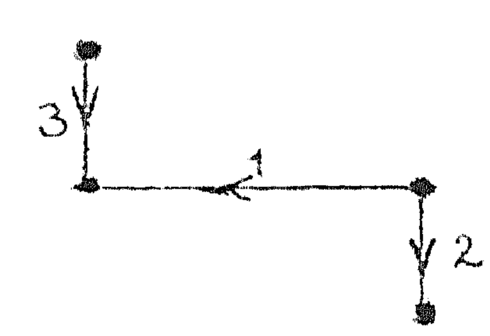
To each graph there corresponds a symmetrical 2k x 2k matrix A, consisting of k^2 2 x 2 block-matrices A_{ij} $(i,j = 1,\ldots,k)$, with elements 0 if i=j and for $i \neq j$ and $\mu, \lambda = 1,2$:

$$a_{i\mu,j\lambda} = \begin{cases} 1 \text{ if the } \mu\text{-th point of } J_i \text{ coincides with} \\ \quad \text{the } \lambda\text{-th point of } J_j. \\ 0 \text{ if not.} \end{cases}$$

All graphs having the same matrix A are considered as equivalent. E.g. both



and



have as matrix

$$\left(\begin{array}{cc|cc|cc} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ \hline 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{array}\right)$$

and are therefore equivalent.

Two graphs that are not equivalent will be called distinct.

The k x k matrix with elements

$$b_{ij} = \sum_{\mu=0}^{1} \sum_{\lambda=0}^{1} a_{i\mu,j\lambda}$$

will be called the <u>configurationmatrix</u>. Here b_{ij} $(i \neq j)$ indicates the number of common endpoints of join J_i and J_j, thus b_{ij} is either 0, 1 or 2. For i=j $b_{ij}=0$. The orientation of the joins does not influence the configuration of the graph.

Consider two graphs G_1 and G_2, each based on ℓ points and k (labelled and oriented) joins. If G_1 and G_2 are not equivalent, but a permutationmatrix P exists such that for the configuration-matrices B_1 and B_2 the relation

$$B_1 = P\, B_2\, P'$$

holds, we shall say that G_1 and G_2 have the same configuration. If two graphs G_1 and G_2 have the same configuration, this means that permuting the numbers of the joins and/or reversing the orientation of some joins of configuration G_1 can make G_1 equivalent to G_2.

A graph $G = (S,U)$ is called connected if from every point $i \in S$ one can reach any other point of S by travelling along the joins of the set U, neglecting the orientation of the joins. A graph which is not connected, can be decomposed in a number of connected components. This decomposition is unique (cf. D. KONIG, 1936, p.15). A configuration-matrix of a not connected graph (if necessary after premultiplication with a permutationmatrix P, and postmultiplication with P`) is a logical sum of the configuration-matrices of each of the connected components.

A connected graph with k joins has at most (k+1) points. It has at least two points. For ℓ satisfying

$$z \leq \ell \leq k + 1,$$

there exist finitely many, say $q_{k,\ell}$, different configurations corresponding to connected graphs, based on k joins and ℓ points. Let $C_{k,\ell}^{(\alpha)}$ be the α-th one ($\alpha = 1,\ldots,q_{k,\ell}$). The configuration of a graph having h connected components ($1 \leq h \leq [\frac{\ell}{2}]$) can now be indicated symbolically by

$$\sum_{i=1}^{h} C_{k_i,\ell_i}^{(\alpha_i)},$$

if the i-th connected component has a configuration $C_{k_i,\ell_i}^{(\alpha_i)}$. If among the h conncected components g_j have the same configuration $C_{k_j,\ell_j}^{(\alpha_j)}$, we may also write $\sum_{j=1}^{s} g_j\, C_{k_j,\ell_j}^{(\alpha_j)}$ as the symbol of the configuration of the graph.

By means of the operator N(...), operating on the symbol of a configuration, we indicate the number of distinct graphs, having this configurations. It can be proved that

$$(3.1)\quad N\Big(\sum_{j=1}^{s} g_j\, C_{k_j,\ell_j}^{(\alpha_j)}\Big) = k!\,\prod_{1}^{s} \frac{1}{g_j!}\left\{\frac{N(C_{k_j,\ell_j}^{(\alpha_j)})}{k_j!}\right\}^{g_j}$$

The calculation of $N(C_{k_j,\ell_j}^{(\alpha_j)})$ proceeds by means of recurrence relations.

4. A general expression for the moments

In order to calculate the k-th moment of $\underline{z}$, we have to consider products

$$(4.1)\qquad \underline{z}_{\nu_{1,1},\nu_{1,2}}\,\underline{z}_{\nu_{2,1},\nu_{2,2}}\cdots\underline{z}_{\nu_{k,1},\nu_{k,2}}\ ,$$

where $\nu_{1,1},\ldots,\nu_{k,2}$ are integers from the range $1,\ldots,n$. If for some $j=1,\ldots,k$

$$\nu_{j,1} = \nu_{j,2}$$

(4.1) is zero a.c. by assumption . We shall therefore only consider the case

$$(4.2)\qquad \nu_{j,1} \neq \nu_{j,2}$$

for all $j = 1,\ldots,k$.

Assume that among $\nu_{1,1},\ldots,\nu_{k,2}$, say ℓ different integers occur $(2\leq \ell \leq 2k)$. We shall denote these by $\lambda_1,\ldots,\lambda_\ell$.

To each product (4.1) there corresponds a graph: let each subscript of (4.1) corresponds to a point. If two subscripts are equal, they correspond to the same point, thus there are ℓ points in all. Let the first subscript $\nu_{j,1}$, of $\underline{z}_{\nu_{j,1},\nu_{j,2}}$ $(j=1,\ldots,k)$ correspond to the first point of a join, and the second subscript to the second point. Now we have obtained a graph with k oriented joins and ℓ points. Loops do not occur because of (4.2). Multiple joins between points correspond to powers of z. Let the graph corresponding to (4.1) have the configuration $\sum_{i=1}^{h} C_{k_i,\ell_i}^{(\alpha_i)}$, then in the following sections we show that for $\underline{z}_{ij}$ for $i \neq j$ holds:

For each $k = 1,\ldots,$ the expectation of (4.1) does not depend on the actual values of $\nu_{1,1},\ldots,\nu_{k,2}$, but only on the configuration $\sum_{i=1}^{h} C_{k_i,\ell_i}^{(\alpha_i)}$.

We can therefore use the following notation for the expectation of (4.1):

$$E(\underline{z}^{(1)}\ldots\underline{z}^{(k)}\,|\,\sum_{i=1}^{h} c_{k_i,\ell_i}^{(\alpha_i)}).$$

In calculating the k-th moment of $\underline{z}$, we also need ℓ-fold sums of the type

$$(4.3)\qquad \underbrace{\sum_{\theta_1=1}^{n}\ldots\sum_{\theta_\ell=1}^{n}}_{D^*} m_{\theta_{\mu_{1,1}},\theta_{\mu_{2,2}}}\; m_{\theta_{\mu_{2,1}},\theta_{\mu_{2,2}}}\ldots m_{\theta_{\mu_{k,1}},\theta_{\mu_{k,2}}},$$

where among the subscripts of the θ's in the general term of the sum all integers $1,\ldots,\ell$ occur. The summation is restricted by a condition D^*, yet to be specified.

If for some $j = 1,\ldots,k$

$$\mu_{j,1} = \mu_{j,2},$$

the sum is equal to zero by (1.2). We therefore consider only the case

$$(4.4)\qquad \mu_{j,1} \neq \mu_{j,2},$$

for $j = 1,\ldots,k$.

In the same way used to construct a graph corresponding to the subscripts in (4.1), we now construct a graph corresponding to the subscripts of the θ's. Let each subscript $\mu_{1,1},\ldots\mu_{k,2}$ correspond to a point. If two are equal they correspond to the same point, thus there are ℓ points in all.

Let the subscript $\mu_{j,1}$ of $m_{\theta_{\mu_{j,1}},\theta_{\mu_{j,2}}}$ $(j = 1,\ldots,k)$ correspond to the first point of a join, and the subscript $\mu_{j,2}$ to the second point. Now we have obtained a graph with k oriented joins and ℓ points. Loops do not occur because of (4.4). Let the graph have the configuration $\sum_{i=1}^{h} c_{k_i,\ell_i}^{(\alpha_i)}$, then the following notation for (4.3) will be used:

1) if there is no restriction D^*:

$$\sum\{m^{(1)}\ldots m^{(k)}\,|\,\sum_{i=1}^{h} c_{k_i,\ell_i}^{(\alpha_i)}\},$$

2) if restriction D^* is that in the summation indices $\theta_1,\dots,\theta_\ell$ are all unequal:

$$\sum{}^{*}\left\{ m^{(1)} \dots m^{(k)} \middle| \sum_{i=1}^{h} C_{k_i,\ell_i}^{(\alpha_i)} \right\} .$$

Now one can write

$$(4.5) \qquad E\underline{z}^k = \sum_{\ell=2}^{2k} \sum_{h=1}^{[\frac{\ell}{2}]} \sum{}''_{\lfloor \sum k_i = k,\ \sum \ell_i = \ell \rfloor} N\left(\sum_{i=1}^{h} C_{k_i,\ell_i}^{(\alpha_i)}\right)$$

$$E\left(\underline{z}^{(1)} \dots \underline{z}^{(k)} \middle| \sum_{i=1}^{h} C_{k_i,\ell_i}^{(\alpha_i)}\right) \sum{}^{*}\left\{ m^{(1)} \dots m^{(k)} \middle| \sum_{i=1}^{h} C_{k_i,\ell_i}^{(\alpha_i)} \right\} ,$$

where $\sum{}''_{\lfloor \sum_{i=1}^{h} k_i = k,\ \sum_{i=1}^{h} \ell_i = \ell \rfloor}$ means a summation over all configurations satisfying the indicated restrictions $\sum_{i=1}^{k} k_i = k$, and $\sum_{i=1}^{h} \ell_i = \ell$.

We define: $m_{i+} \overset{\text{def}}{=} \sum_j m_{ij}$

and $m_{++} \overset{\text{def}}{=} \sum_i m_{i+}$.

5. The moments of $\underline{x}_W$ and $\underline{x}_B$

The moments of $\underline{x}$ can be found by replacing r_1 or p_1 in the corresponding formulae for $\underline{x}_B$ by r_2 or p_2 respectively.

a) non-free sampling

$$(5.1) \qquad E\left(\underline{x}_B^{(1)} \dots \underline{x}_B^{(k)} \middle| \sum_{i=1}^{h} C_{k_i,\ell_i}^{(\alpha_i)}\right) = \frac{\binom{r_1}{\ell}}{\binom{n}{\ell}} ,$$

where $\sum_{i=1}^{h} \ell_i = \ell$.

From (4.5) we find after some simplifications

$$E\,\underline{x}_B = \frac{r_1(r_1-1)}{n(n-1)}\, m_{++} ,$$

$$\sigma^2 = E\,\underline{x}_B^2 - (E\,\underline{x}_B)^2 = 4\,\frac{r_1(r_1-1)(r_1-2)(n-r_1)}{n(n-1)(n-2)(n-3)}\sum_i (m_{i+} - \frac{1}{n}m_{++})^2$$

$$+\frac{2r_1(r_1-1)(n-r_1)(n-r_1-1)}{n^2(n-1)^2(n-2)(n-3)}\left\{n(n-1)\sum_{ij} m_{ij}{}^2 - m_{++}^2\right\}.$$

If m_{i+} does not depend on i, the first term of σ^2 is equal to zero.

b) free sampling

$$(5.2)\qquad E(\underline{x}_B^{(1)}\ldots\underline{x}_B^{(k)} \mid \sum_{i=1}^{h} C_{k_i,\ell_i}^{(\alpha_i)}) = p^{\ell},$$

so

$$E\,\underline{x}_B = p_1^2 \sum_{ij} m_{ij},$$

$$\sigma^2 = 2p_1^2(1-p_1)^2 \sum_{ij} m_{ij}^2 + 4p^3(1-p)\sum_i m_{i+}^2$$

6. The moments of y

In order to calculate

$$E\{\underline{y}^{(1)}\ldots\underline{y}^{(k)} \mid \sum_{i=1}^{h} C_{k_i,\ell_i}^{(\alpha_i)}\}$$

we first take a point P_i of the i-th connected component (i=1,.. ,h) as a reference point. Colour P_i white, next all points connected by a join to P_i are coloured black, then all points connected to these black points are coloured white. If in repeating this procedure one arrives at a point which has already been given one colour, but should be coloured by the just-mentioned rule in the other colour as well, then we conclude that the i-th connected component is not bichromatic.
If no such situation arises one arrives at a stage, where all points have been allotted a colour, viz τ_i points are white and $\ell_i - \tau_i$ black; we then say that the i-th connected component is bichromatic. The fact wether a graph is bichromatic or not does

not depend on the choice of the initial point.

Define

$$(6.1)\quad B\Big(\sum_{i=1}^{h} c_{k_i,\ell_i}^{(\alpha_i)}\Big) = \begin{cases} 1 & \text{if all connected components of } \sum_{i=1}^{h} c_{k_i,\ell_i}^{(\alpha_i)} \text{ are bichromatic} \\ 0 & \text{if not.} \end{cases}$$

a) <u>non free sampling</u>

$$(6.2)\quad E\Big\{\underline{y}^{(1)}\ldots\underline{y}^{(k)} \,\Big|\, \sum_{i=1}^{h} c_{k_i,\ell_i}^{\alpha_i}\Big\} = B\Big(\sum_{i=1}^{h} c_{k_i,\ell_i}^{(\alpha_i)}\Big)\,\frac{(n-\ell)!}{n!}\sum_{i=1}^{h}\sum_{\rho_i=0}^{1}$$

$$\frac{r_1!}{(r_1-\sum_i(1-\rho_i)\,\tau_i-\sum_i\rho_i(\ell_i-\tau_i))!}\;\frac{r_2!}{(r_2-\sum_i\rho_i\tau_i-\sum_i(\ell_i-\tau_i)(1-\rho_i))!}.$$

So

$$E\,\underline{y} = \frac{2r_1r_2}{n(n-1)}\sum_{ij} m_{ij},$$

$$\sigma^2 = \sum_i\Big(\sum_j m_{ij} - \frac{\sum_{ij} m_{ij}}{n}\Big)^2\Big\{\frac{r_1r_2(r_2-1)}{n(n-1)(n-2)} + \frac{r_2r_1(r_1-1)}{n(n-1)(n-2)} - \frac{4r_1r_2(r_1-1)(r_2-1)}{n(n-1)(n-2)(n-3)}\Big\}$$

$$+ 4\sum_{ij} m_{ij}^2\Big\{\frac{r_1r_2}{n(n-1)} - \frac{r_1r_2(r_2-1)}{n(n-1)(n-2)} - \frac{r_2r_1(r_1-1)}{n(n-1)(n-2)} + \frac{2r_1r_2(r_1-1)(r_2-1)}{n(n-1)(n-2)(n-3)}\Big\}$$

$$+ 4\Big(\sum_{ij} m_{ij}\Big)^2\,\frac{r_1r_2n(r_1+r_2+3)-r_1r_2(2r_1r_2+r_1+r_2+2)}{n^2(n-1)^2(n-2)(n-3)}.$$

b) <u>free sampling</u>

$$(6.3)\quad E\Big(\underline{y}^{(1)}\ldots\underline{y}^{(k)} \,\Big|\, \sum_{i=1}^{h} c_{k_i,\ell_i}^{(\alpha_i)}\Big) =$$

$$= B\Big(\sum_{i=1}^{h} c_{k_i,\ell_i}^{(\alpha_i)}\Big)\prod_{1}^{h}\big(p_1^{\tau_i}\,p_2^{\ell_i-\tau_i} + p_1^{\ell_i-\tau_i}\,p_2^{\tau_i}\big),$$

thus

$$E\,\underline{y} = 2p_1p_2\sum_{ij} m_{ij},$$

$$\sigma^2 = 4p_1p_2(p_1+p_2-4p_1p_2)\sum_{ijk} m_{ij}m_{ik} + 4p_1p_2(1-p_1-p_2+2p_1p_2)\sum_{ij} m_{ij}^2.$$

7. Tendency towards the normal distribution

The following theorems can be proved.

Theorem 7.1

i If r_1 and n tend to infinity such that

$$\lim \frac{r_1}{n} = \delta_1, \qquad 0 < \delta_1 < 1,$$

and if for all i

$$m_{i+} < C,$$

where C is a constant independent of i and n, then in the non-free sampling case the distribution of

$$\{\underline{x}_B - E\,\underline{x}_B\}\ \sigma(\underline{x}_B)^{-1}$$

tends to the standard normal one. $E\,\underline{x}_B$ and $\sigma(\underline{x}_B)^2$ have been given in section 5.

ii The same result holds for $\underline{x}_w$ if r_1 is replaced by r_2.

iii If the assumptions of part i of this theorem are satisfied and if moreover

$$\lim \frac{r_2}{n} = \delta_2, \qquad 0 < \delta_2 \leq 1 - \delta,$$

then in the non-free sampling case the distribution of

$$(\underline{y} - E\underline{y})\ \sigma(\underline{y})^{-1}$$

tends to the standard normal one. $E\underline{y}$ and $\sigma^2(y)$ have been given in section 6.

Theorem 7.2

i If n tends to infinity and p_1 to a limit $p_1^*(0 < p_1^* < 1)$, and if for all i

$$m_{i+} < C,$$

where C is a constant independent of i and n, then in the free sampling case the distribution of

$$\{\underline{x}_B - E\,\underline{x}_B\}\,\sigma(\underline{x}_B)^{-1}$$

tends to the standard normal one. $E\,\underline{x}_B$ and $\sigma^2(\underline{x}_B)$ have been given in section 5.

ii The same result holds for $\underline{x}_W$ if p_1 is replaced by p_2.

iii If the assumption of part i holds, and if also p_2 tends to a limit $p_2^*(0 < p_2^* < 1-p_1^*)$ then in the free sampling case the distribution of

$$(\underline{y} - E\underline{y})\, \sigma(\underline{y})^{-1}$$

tends to the standard normal one. $E\underline{y}$ and $\sigma^*(\underline{y})$ have been given in section 6.

8. Tendency towards the compound Poisson-distribution

Theorem 8.1

i If r_1 and n tend to infinity such that

$$\lim \frac{r_1}{n} = 0,$$

$$\lim \frac{r_1^2}{n^2} m_{++} = 2\lambda, \qquad 0 < \lambda < \infty,$$

and for all α and $k \geq 2$,

$$(8.1) \qquad \lim \left(\frac{r_1}{n}\right)^{k+1} \Sigma \left\{ m^{(1)} \ldots m^{(k)} \mid C_{k,k+1}^{(\alpha)} \right\} = 0,$$

and if for all i and j

$$m_{ij} < C_1,$$

where C_1 does not depend on i, j and n, then the distribution of $\frac{1}{2}\underline{x}_B$ tends to a compound Poisson distribution with moment generating function

$$\sum_{k=0}^{\infty} \frac{z^k}{k!} \lim E(\tfrac{1}{2}\underline{x}_B)^k = \exp\left\{\lambda \sum_{i=1}^{\infty} m_i^* \frac{z^i}{i!}\right\},$$

where

$$m_h^* = \lim \frac{\sum_{i,j} m_{ij}^h}{m_{++}}, \qquad h = 1,2,\ldots$$

Assumption (13.1) is satisfied if for all i $m_{i+} \leq C_2$, where C_2 does not depend on i and

$$\lim \frac{r_1}{n} C_2 = 0,$$

thus e.g. when C_2 is a constant not dependent on n.

ii The same result holds it $\underline{x}_B$ and r_1 is replaced by $\underline{x}_W$ and r_2 respectively.

iii If in part ii and iii $\frac{r_1}{n}$ is replaced by p_1, and $\frac{r_2}{n}$ by p_2, then the corresponding results for the free sampling case is obtained.

Theorem 13.2

i If r_1, r_2 and n tend to infinity, such that

$$\lim \frac{r_1}{n} = 0,$$

$$\lim \frac{r_2}{n} = 0,$$

$$\lim \frac{r_1 r_2}{n^2} m_{++} = \lambda , \qquad 0 < \lambda < \infty,$$

and for all α and $k = 2, \ldots .$

$$(13.2) \qquad \lim \frac{r_1^j r_2^{k-j+1}}{n^{k+1}} \sum \{ m^{(1)} \ldots m^{(k)} | C^{(\alpha)}_{k,k+1} \} = 0, \quad j=1, \ldots, k,$$

and if for all i and j

$$m_{ij} < C_1,$$

where C_1 does not depend on i, j and n, then in the non-free sampling case the distribution of $\frac{1}{2}\underline{y}$ tends to a compound Poisson-distribution with

$$\sum_{k=0}^{\infty} \frac{Z^k}{k!} \lim E(\tfrac{1}{2}\underline{y})^k = \exp\left\{ \lambda \sum_{i=1}^{\infty} \frac{m_i^* Z^i}{i!} \right\},$$

with

$$m_h^* = \lim \frac{\sum_{ij} m_{ij}^h}{m_{++}} , \qquad h = 1,2 \ldots .$$

Assumption (13.2) is satisfied if for all i $m_{i+} \leq C_2$, where C_2 does not depend on i and

$$\lim \frac{r_1}{n} C_2 = 0$$

$$\lim \frac{r_2}{n} C_2 = 0,$$

thus e.g. when C_2 is a constant not dependent on n.

ii If in part i $\frac{r_1}{n}$ is replaced by p_1, and $\frac{r_2}{n}$ by p_2, the corresponding result for the free sampling case is obtained.

9. References

C. BERGE (1958): Théorie des graphes et ses applications. Dunod, Paris.

C. van EEDEN and A.R. BLOEMENA, (1959): On probability distributions arising from points on a lattice. Report S 257, Statistics Dept. Mathematical Centre.

D.J. FINNEY (1947): The significance of associations in a square point lattice, Journ. Roy. Stat.Soc. Suppl.(9),99.

D. KOENIG (1936): Theorie der endlichen und unendlichen Graphen, Leipzig.

P.V. KRISHNA IYER (1947): Random association of points on a lattice, Nature, 160, 714.

P.V. KRISHNA IYER (1948): The theory of probability distributions of points on a line, JISAS 1, 173-195.

P.V. KRISHNA IYER (1949): The first and second moments of some probability distributions arising from points on a lattice and their applications, Biometrika, 36, 135-141.

P.V. KRISHNA IYER (1950a) The theory of probability distributions of points on a lattice. Ann.Math.Stat. 21, 198-217 & 22,310.

P.V. KRISHNA IYER (1950b,1951,1952) Further contributions to the theory of probability distributions of points on a line. JISAS, 2, 141-160; 3, 80-93; 4, 50-71.

P.A.P. MORAN (1947): Random associations on a lattice Proc.Cambr.Phil.Soc., 43, 321-328.

P.A.P. MORAN (1948): The interpretation of Statistical maps. Journ.Roy.Stat.Soc. B 10, 243-251.

H. TODD (1940): Note on the random association in a square point lattice, Journ.Roy.Stat. Soc.Suppl.7, 78.

JISAS = Journ. of the Indian Society for Agricultural Statistics.