MATHEMATICAL CENTRE TRACTS

2

SAMPLING FROM A GRAPH

BY

A. R. BLOEMENA

MATHEMATISCH CENTRUM AMSTERDAM

1964

FOREWORD

This book is the posthumous publication of the scientific work of A.R. BLOEMENA who died in 1960 at the early age of 32. He started his career as a mining engineer from the Technical University, Delft, Holland, but after practicing for a couple of years he switched to mathematical statistics. He joined the Mathematical Centre and developed rapidly into a very promising statistical research worker. Soon he was appointed deputy chief of the Statistical Department.

His thesis was nearly finished when he died. It had still, however, to be brought into its final form. This was done by his successor W.R. van ZWET. The Mathematical Centre is very grateful for his work which was far more than a mere editorial task, and it is glad to be able to present the main research results of BLOEMENA in the form of this book.

J. Hemelrijk

CONTENTS

CHAPTER 1

1.1. Introduction

Let be given a set of n points, numbered 1,2,...,n, and a $(n \times n)$-matrix M, with elements m_{ij}, satisfying

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

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

(1.1.3) $$\sum_j m_{ij}^2 \geq 1 \quad \text{for each } i,$$

(1.1.4) $$0 \leq m_{ij} < \infty .$$

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 of n points (cf. C. BERGE (1958), D. KOENIG (1936)), where the number of joins between points i and j is equal to m_{ij}. In this interpretation, $m_{ij} = 0$ means that there is no join between i and j; assumption (1.1.2) states that there are no loops, while assumption (1.1.3) implies that no point is isolated.

We shall sometimes indicate $\sum_j m_{ij}$ by m_{i+} and $\sum_i m_{i+}$ by m_{++}.

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

Case I. "non free sampling" : from the points 1,2,...,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*) $\underline{x}_{ij}^{(B)}, \underline{x}_{ij}^{(W)}, \underline{y}_{ij}$ (i,j=1,2,...,n) defined by**)

*) We shall distinguish random variables from numbers (e.g. the values they assume in an experiment) by underlining their symbols.

**) By writing spr α (salve probabilitate α) after a statement we shall indicate that the statement is true except for an event with probability smaller than or equal to α . Hence spr 0 corresponds to "with probability 1".

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

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

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

and for $i \neq j$

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

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

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

Obviously

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

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

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

Define

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

(1.1.5)
$$\underline{x}_W = \sum_{ij} m_{ij}\, \underline{x}_{ij}^{(W)},$$

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

Now $\underline{x}_B$ is twice the number of joins between black points, $\underline{x}_W$ is twice the number of joins between white points, whereas $\underline{y}$ is twice the number of joins between black and white points. Notice that if $m_{i,i+1} = m_{i+1,i} = 1$ for $i=1,2,\ldots,n-1$, and $m_{ij} = 0$ otherwise, and if $r_1+r_2=n$, then $\frac{1}{2}\underline{y}+1$ is the number of runs in a sequence of alternatives.

Define for $i,j=1,2,\ldots,n$

(1.1.6)
$$\underline{v}_{ij} = \underline{x}_{ij}^{(B)} + \underline{x}_{ij}^{(W)} - \underline{y}_{ij},$$

and

$$\underline{v} = \sum_{ij} m_{ij}\, \underline{v}_{ij}.$$

Obviously

(1.1.7)
$$\underline{v} = \underline{x}_B + \underline{x}_W - \underline{y}.$$

The statistic $\underline{v}$ is met in the study of the order - disorder problem (cf. A.R. BLOEMENA (1960))

A more general variable $\underline{z}$ of the form

$$(1.1.8) \qquad \underline{z} = \sum_{ij} m_{ij} \underline{z}_{ij} ,$$

where $\underline{z}_{ij}$ are random variables satisfying

$$\underline{z}_{ii} = 0 \quad \text{spr } 0$$

and for $i \neq j$

$$\underline{z}_{ij} = \underline{z}_{ji} ,$$

but not necessarily connected with points in a graph, will also be considered. A symmetry condition that is satisfied by $\underline{x}_B$, $\underline{x}_W$, $\underline{y}$ and $\underline{v}$ will also be imposed on $\underline{z}$ (cf. section 3.1), which then becomes a useful generalization of $\underline{x}_B$, $\underline{x}_W$, $\underline{y}$ and $\underline{v}$.

Obviously $\underline{z}$, as defined by (1.1.8), is a statistic belonging to the class $\mathcal{W}$ of statistics $\underline{w}$ that can be expressed as

$$(1.1.9) \qquad \underline{w} = \sum_{i=1}^{n} \sum_{j=1}^{n} \underline{w}_{ij} ,$$

where the random variables $\underline{w}_{ij}$ satisfy

$$\underline{w}_{ii} = 0 \quad \text{spr } 0$$

and for $i \neq j$

$$\underline{w}_{ij} = \underline{w}_{ji} .$$

The class $\mathcal{W}$ contains a well-known statistic. Let be given n pairs of random variables $(\underline{u}_i, \underline{v}_i)$, $(i=1,2,\ldots,n)$, and define for $i \neq j$

$$\underline{s}_{ij} = \begin{cases} +\frac{1}{2} & \text{if } (\underline{u}_i-\underline{u}_j)(\underline{v}_i-\underline{v}_j) > 0 \\ 0 & \text{if } (\underline{u}_i-\underline{u}_j)(\underline{v}_i-\underline{v}_j) = 0 \\ -\frac{1}{2} & \text{if } (\underline{u}_i-\underline{u}_j)(\underline{v}_i-\underline{v}_j) < 0 , \end{cases}$$

and

$$\underline{s}_{ii} = 0 \quad \text{spr } 0 ,$$

then the statistic $\underline{S}$, defined by

$$\underline{S} = \sum_{ij} \underline{s}_{ij}$$

is the statistic of M.G. KENDALL's rank correlation test (cf. M.G. KENDALL (1955)), and belongs to $\mathcal{W}$.

A review of records of previous work on the subject will be given at the end of this section. The next section gives some results on the probability distribution of $\underline{x}_B$ for small values of r_1.

To study the stochastic properties of $\underline{w}$, one approach is to study its moments. In chapter 2 we develop expressions for the reduced and unreduced moments of $\underline{w}$, using the theory of graphs. In chapters 3 and 4 these results are applied to $\underline{z}$, and to $\underline{x}_B$, $\underline{x}_W$, $\underline{y}$ and $\underline{v}$. Chapter 5 deals with an application to the problem of a test for randomness.

For results on runs in a sequence of alternatives we refer to H.A. KUIPERS (1957).

As far as we know the earliest results for the case of a rectangular lattice with $m_{ij} = 0$ or 1 are given by J.G. KIRKWOOD (1938) in a paper on the order-disorder problem. KIRKWOOD states the first moment and asymptotic expressions for the second and third cumulant of a simple transform of $\underline{x}_B$.

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, and the third and fourth moments for the case of free sampling. He proves the asymptotic normality of $\underline{x}_B$ and $\underline{y}$ (free sampling) for a rectangular two-dimensional lattice, where there are joins between neighbouring points in the direction of both axes (cf. also P.A.P. MORAN (1947)).

There exists a large number of papers on the subject by P.V. KRISHNA IYER (1947-1952), dealing with rectangular lattices, where either neighbouring points are joined in the direction of both axes, or neighbouring points are joined in the direction of both axes and in diagonal directions. The results of KRISHNA IYER are mostly on the first four moments or cumulants; statements are made about the asymptotic behaviour of the distributions of the statistics.

Results on the case of a rectangular two-dimensional lattice with vacancies (which is in fact a special case of MORAN's statistical map) are given by G.H. FREEMAN (1953).

A number of exact results for rectangular lattices (non free sampling) are described in a report by C. VAN EEDEN and A.R. BLOEMENA (1959). The present study is an outgrowth of this last-mentioned report, which arose from a study of the distribution of a statistic, obtained in a psychological test (cf. C.A.G. NASS (1960)).

Editor's note.
The author was aware of the existence of a great number of papers on the order-disorder problem and, given time, would have revised this review of previous work accordingly. The editor has refrained from

doing so because an excellent review of work on the order-disorder problem has since become available. The interested reader is referred to C. DOMB, On the theory of cooperative phenomena in crystals, Philos. Mag. Suppl., 9, No 34, p 149-361, 1960.

1.2. The exact distribution of $\underline{x}_B$ (non free sampling)

If r_1 is small and the m_{ij}'s assume only a few unequal values, the exact distribution of $\underline{x}_B$ can be obtained in a simple way. In principle the same procedure can be used for larger values of r_1 and for the case where the m_{ij}'s assume a larger number of unequal values, but the amount of simple algebra involved becomes rapidly prohibitive.

To demonstrate the method we deal with the case $r_1=3$ in some detail. For $i \neq j$ the m_{ij}'s are assumed to take three unequal values, b_1, b_2 and b_3. It follows that the random variable $\underline{x}_B$ can assume with non-zero probability the values $6b_1$, $6b_2$, $6b_3$, $4b_1+2b_2$, ... , $4b_3+2b_2$, $2b_1+2b_2+2b_3$.

To avoid notational difficulties we shall assume that these ten possible values for $\underline{x}_B$ are all unequal. However, if two or more of these values do happen to be equal the distribution of $\underline{x}_B$ may be calculated in exactly the same way by adding the corresponding probabilities.

It is only necessary to calculate three probabilities, viz. those of $\underline{x}_B$ taking the values $6b_1$, $4b_1+2b_2$ and $2b_1+2b_2+2b_3$. The other probabilities then follow by symmetry.

The probability that three given points i,j,k are chosen from n points is equal to $\{n(n-1)(n-2)\}^{-1}$. If points i,j,k are chosen, the event

$$"\underline{x}_B = 2b_1 + 2b_2 + 2b_3"$$

occurs if and only if one of the following statements is true:

I : $m_{ij} = b_1$, $m_{ik} = b_2$ and $m_{jk} = b_3$,

II: $m_{ij} = b_1$, $m_{ik} = b_3$ and $m_{jk} = b_2$,

⋮

VI: $m_{ij} = b_3$, $m_{ik} = b_1$ and $m_{jk} = b_2$.

At most one of these statements can be true for given i,j,k.

Let ν_1, ν_2, ν_3 be a permutation of the numbers 1, 2 and 3. An indicator which is 1 if $m_{ij} = b_{\nu_1}$, $m_{ik} = b_{\nu_2}$, and $m_{jk} = b_{\nu_3}$, and 0 otherwise is

$$\frac{(b_{\nu_2}-m_{ij})(b_{\nu_3}-m_{ij})(b_{\nu_1}-m_{ik})(b_{\nu_3}-m_{ik})(b_{\nu_1}-m_{jk})(b_{\nu_2}-m_{jk})}{(b_2-b_1)(b_3-b_1)(b_1-b_2)(b_3-b_2)(b_1-b_3)(b_2-b_3)} .$$

Thus if one of the statements I,II,...,VI is true

$$\frac{\sum\limits_{(\nu_1\nu_2\nu_3)} (b_{\nu_2}-m_{ij})(b_{\nu_3}-m_{ij})(b_{\nu_1}-m_{ik})(b_{\nu_3}-m_{ik})(b_{\nu_1}-m_{jk})(b_{\nu_2}-m_{jk})}{\prod\limits_{\omega=1}^{3} \prod\limits_{\substack{\mu=1\\ \mu\neq\omega}}^{3} (b_\omega-b_\mu)} = 1 ,$$

($\sum\limits_{(\nu_1\nu_2\nu_3)}$ means summing over all permutations of the numbers 1,2,3)
while this expression is zero if none of the six statements is true. Thus

$$P[\underline{x}_B = 2b_1+2b_2+2b_3 \wedge \text{ points } i,j,k \text{ are chosen} \mid r_1 = 3] =$$

$$= \frac{\sum\limits_{(\nu_1\nu_2\nu_3)} (b_{\nu_2}-m_{ij})(b_{\nu_3}-m_{ij})(b_{\nu_1}-m_{ik})(b_{\nu_3}-m_{ik})(b_{\nu_1}-m_{jk})(b_{\nu_2}-m_{jk})}{n(n-1)(n-2) \prod\limits_{\omega=1}^{3} \prod\limits_{\substack{\mu=1\\ \mu\neq\omega}}^{3} (b_\omega-b_\mu)} .$$

Summing over unequal values of i,j,k and interchanging the order of summation gives

$$(1.2.1) \qquad P[\underline{x}_B = 2b_1+2b_2+2b_3 \mid r_1 = 3] =$$

$$= \frac{\sum\limits_{(\nu_1\nu_2\nu_3)} \sum\limits_{(ijk)\neq} (b_{\nu_2}-m_{ij})(b_{\nu_3}-m_{ij})(b_{\nu_1}-m_{ik})(b_{\nu_3}-m_{ik})(b_{\nu_1}-m_{jk})(b_{\nu_2}-m_{jk})}{n(n-1)(n-2) \prod\limits_{\omega=1}^{3} \prod\limits_{\substack{\mu=1\\ \mu\neq\omega}}^{3} (b_\omega-b_\mu)} .$$

Suppose that among the n points there exists at least one triplet of points numbered $\lambda_1, \lambda_2, \lambda_3$, such that $m_{\lambda_1,\lambda_2} = b_1$, $m_{\lambda_1,\lambda_3} = b_2$ and $m_{\lambda_2,\lambda_3} = b_3$. Then each such triplet contributes 6 identical terms to (1.2.1) viz. one for each permutation ν_1,ν_2,ν_3. E.g. for $\nu_1=1$, $\nu_2=3$, $\nu_3=2$ (take $i=\lambda_2$, $j=\lambda_1$, $k=\lambda_3$) there is a term in the numerator of (1.2.1)

$$(b_3-m_{\lambda_2,\lambda_1})(b_2-m_{\lambda_2,\lambda_1})(b_1-m_{\lambda_2,\lambda_3})(b_2-m_{\lambda_2,\lambda_3})(b_1-m_{\lambda_1,\lambda_3})(b_3-m_{\lambda_1,\lambda_3})$$

while for $\nu_1=3$, $\nu_2=2$, $\nu_3=1$ we have (take $i=\lambda_3$, $j=\lambda_2$, $k=\lambda_1$)

$$(b_2-m_{\lambda_3,\lambda_2})(b_1-m_{\lambda_3,\lambda_2})(b_3-m_{\lambda_3,\lambda_1})(b_1-m_{\lambda_3,\lambda_1})(b_3-m_{\lambda_2,\lambda_1})(b_2-m_{\lambda_2,\lambda_1}) .$$

Therefore (1.2.1) can be written

(1.2.2) $$P[\underline{x}_B = 2b_1+2b_2+2b_3 \mid r_1 = 3] =$$

$$= \frac{6 \sum_{(ijk)\neq} (b_1-m_{ij})(b_2-m_{ij})(b_1-m_{ik})(b_3-m_{ik})(b_2-m_{jk})(b_3-m_{jk})}{n(n-1)(n-2) \prod_{\omega=1}^{3} \prod_{\substack{\mu=1\\ \mu\neq\omega}}^{3} (b_\omega-b_\mu)} .$$

If no triplet of points $\lambda_1,\lambda_2,\lambda_3$, can be found such that $m_{\lambda_1,\lambda_2} = b_1$, $m_{\lambda_1,\lambda_3} = b_2$, $m_{\lambda_2,\lambda_3} = b_3$, (1.2.1) is equal to zero, and so is (1.2.2).

In exactly the same way one obtains

(1.2.3) $$P[\underline{x}_B = 4b_1+2b_2 \mid r_1 = 3] =$$

$$= \frac{3 \sum_{(ijk)\neq} (b_2-m_{ij})(b_3-m_{ij})(b_2-m_{ik})(b_3-m_{ik})(b_1-m_{jk})(b_3-m_{jk})}{n(n-1)(n-2)(b_2-b_1)^2(b_3-b_1)^2(b_1-b_2)(b_3-b_2)} ,$$

and

(1.2.4) $$P[\underline{x}_B = 6b_1 \mid r_1 = 3] =$$

$$= \frac{\sum_{(ijk)\neq} (b_2-m_{ij})(b_3-m_{ij})(b_2-m_{ik})(b_3-m_{ik})(b_2-m_{jk})(b_3-m_{jk})}{n(n-1)(n-2)(b_2-b_1)^3(b_3-b_1)^3} .$$

The expressions (1.2.2), (1.2.3) and (1.2.4) may be simplified, e.g. the numerator of (1.2.4) can be written as

$$b_2^3b_3^3n(n-1)(n-2) - 3b_2^3b_3^2(b_2+b_3)(n-2) \sum_{(ij)\neq} m_{ij} +$$

$$+ 3b_2^2b_3^2(n-2) \sum_{(ij)\neq} m_{ij}^2 + 3b_2b_3(b_2+b_3)^2 \sum_{(ijk)\neq} m_{ij}\, m_{ik} +$$

$$- 6b_2b_3(b_2+b_3) \sum_{(ijk)\neq} m_{ij}^2\, m_{ik} + 3b_2b_3 \sum_{(ijk)\neq} m_{ij}^2\, m_{ik}^2 +$$

$$- (b_2+b_3)^3 \sum_{(ijk)\neq} m_{ij}\, m_{ik}\, m_{jk} + 3(b_2+b_3)^2 \sum_{(ijk)\neq} m_{ij}^2\, m_{ik}\, m_{jk} +$$

$$- 3(b_2+b_3) \sum_{(ijk)\neq} m_{ij}^2\, m_{ik}^2\, m_{jk} + \sum_{(ijk)\neq} m_{ij}^2\, m_{ik}^2\, m_{jk}^2 .$$

Now because of (1.1.2)

$$\sum_{(ij)\neq} m_{ij} = \sum_{ij} m_{ij}$$

$$\sum_{(ij)\neq} m_{ij}^2 = \sum_{ij} m_{ij}^2$$

$$\sum_{(ijk)\neq} m_{ij}\, m_{ik} = \sum_{(ij)\neq} \left[\sum_k m_{ij}\, m_{ik} - m_{ij}^2 \right] = \sum_{ijk} m_{ij}\, m_{ik} - \sum_{ij} m_{ij}^2$$

$$\sum_{(ijk)\neq} m_{ij}^2\, m_{ik} = \sum_{ijk} m_{ij}^2\, m_{ik} - \sum_{ij} m_{ij}^3$$

$$\sum_{(ijk)\neq} m_{ij}^2\, m_{ik}^2 = \sum_{ijk} m_{ij}^2\, m_{ik}^2 - \sum_{ij} m_{ij}^4$$

$$\sum_{(ijk)\neq} m_{ij}\, m_{ik}\, m_{jk} = \sum_{ijk} m_{ij}\, m_{ik}\, m_{jk}\,, \quad \text{etc.}$$

For $i \neq j$

$$(b_1 - m_{ij})(b_2 - m_{ij})(b_3 - m_{ij}) \equiv 0\,. \tag{1.2.5}$$

Summing this identity over i and j with $i \neq j$ gives

$$b_1 b_2 b_3 n(n-1) - (b_1b_2 + b_1b_3 + b_2b_3) \sum_{ij} m_{ij} + (b_1 + b_2 + b_3) \sum_{ij} m_{ij}^2 = \sum_{ij} m_{ij}^3\,,$$

while multiplying (1.2.5) by m_{ij} and summing gives

$$b_1 b_2 b_3 \sum_{ij} m_{ij} - (b_1b_2 + b_1b_3 + b_2b_3) \sum_{ij} m_{ij}^2 + (b_1 + b_2 + b_3) \sum_{ij} m_{ij}^3 = \sum_{ij} m_{ij}^4\,.$$

Using these identities the results of table 1.2.1 have been obtained. Since without loss of generality one of the numbers b_1, b_2, b_3 may be taken to be equal to zero, terms containing a factor $b_1b_2b_3$ have been omitted.

For $r_1 = 2$, $\underline{x}_B$ assumes the values $2b_1$, $2b_2$ and $2b_3$ with non-zero probability. It is sufficient to calculate $P[\underline{x}_B = 2b_1 \mid r_1 = 2]$; the other probabilities follow by symmetry. We find

$$(b_2 - b_1)(b_3 - b_1).P[\underline{x}_B = 2b_1 \mid r_1 = 2] = b_2 b_3 - \frac{(b_2 + b_3) \sum_{ij} m_{ij} - \sum_{ij} m_{ij}^2}{n(n-1)}\,. \tag{1.2.6}$$

For the case m_{ij} takes only two values, say 0 or 1 (MORAN's model) we have also considered the case $r_1 = 4$. The result is given in table 1.2.2.

Table 1.2.1

$n(n-1)(n-2).c_a.P[\underline{x}_B = a | r_1 = 3]$ for the case where the m_{ij} $(i \neq j)$ assume three unequal values b_1, b_2, b_3 and $b_1 b_2 b_3 = 0$. The table provides coefficients of the sums entered in the first column.

a	$2b_1+2b_2+2b_3$	$4b_1+2b_2$	$6b_1$
c_a	$\frac{1}{6}(b_1-b_2)^2(b_1-b_3)^2(b_2-b_3)^2$	$\frac{1}{3}(b_1-b_2)^3(b_1-b_3)^2(b_3-b_2)$	$(b_3-b_1)^3(b_2-b_1)^3$
1			$+b_2^3 b_3^3 n(n-1)(n-2)$
$\sum_{ij} m_{ij}$		$-b_2^2 b_3^3 n+b_1^2 b_3^2(b_1-b_3)$	$-3b_2^2 b_3^2(b_2+b_3)(n-1)$
$\sum_{ij} m_{ij}^2$		$+b_2^2 b_3^2 n-b_1^2 b_3(b_1-b_3)$	$+3b_2^2 b_3^2(n-1)$
$\sum_{ijk} m_{ij} m_{ik}$	$-b_1^2 b_2^2-b_1^2 b_3^2-b_2^2 b_3^2$	$+b_1 b_3^3+2b_2 b_3^2(b_2+b_3)$	$+3b_2 b_3(b_2+b_3)^2$
$\sum_{ijk} m_{ij}^2 m_{ik}$	$+(b_1+b_2+b_3)(b_1 b_2+b_1 b_3+b_2 b_3)$	$-2b_1 b_3^2-2b_2 b_3(b_2+2b_3)$	$-6b_2 b_3(b_2+b_3)$
$\sum_{ijk} m_{ij}^2 m_{ik}^2$	$-(b_1 b_2+b_1 b_3+b_2 b_3)$	$+b_3(b_1+2b_2)$	$+3b_2 b_3$
$\sum_{ijk} m_{ij} m_{ik} m_{jk}$	$+(b_1+b_2)(b_1+b_3)(b_2+b_3)$	$-(b_2+b_3)^2(b_1+b_3)$	$-(b_2+b_3)^3$
$\sum_{ijk} m_{ij}^2 m_{ik} m_{jk}$	$-(b_1^2+b_2^2+b_3^2)-3(b_1 b_2+b_1 b_3+b_2 b_3)$	$+(b_2+b_3)(2b_1+b_2+3b_3)$	$+3(b_2+b_3)^2$
$\sum_{ijk} m_{ij}^2 m_{ik}^2 m_{jk}$	$+2(b_1+b_2+b_3)$	$-(b_1+2b_2+3b_3)$	$-3(b_2+b_3)$
$\sum_{ijk} m_{ij}^2 m_{ik}^2 m_{jk}^2$	-1	+1	+1

Table 1.2.2

$n(n-1)(n-2)(n-3).P[\underline{x}_B = a | r_1 = 4]$ for the case where $m_{ij} = 0$ or 1.

The table provides coefficients of the sums entered in the first column.

a	0	2	4	6	8	10	12
1	$n(n-1)(n-2)(n-3)$	—	—	—	—	—	—
$\sum_{ij} m_{ij}$	$-6n^2+18n-14$	$+6n^2-6n+12$	$-12n-18$	+20	—	—	—
$\left(\sum_{ij} m_{ij}\right)^2$	+3	-6	+3	—	—	—	—
$\sum_{ijk} m_{ij} m_{ik}$	$+12n-12$	$-24n-12$	$+12n+60$	-36	—	—	—
$\sum_{ijk} m_{ij} m_{ik} m_{jk}$	$-4n$	$+12n+24$	$-12n-72$	$+4n+72$	-24	—	—
$\sum_{ijkl} m_{ij} m_{ik} m_{il}$	-4	+12	-12	+4	—	—	—
$\sum_{ijkl} m_{ij} m_{ik} m_{jl}$	-12	+36	-36	+12	—	—	—
$\sum_{ijkl} m_{ij} m_{ik} m_{jk} m_{il}$	+12	-48	+72	-48	+12	—	—
$\sum_{(ijkl)\neq} m_{ij} m_{ik} m_{jl} m_{kl}$	+3	-12	+18	-12	+3	—	—
$\sum_{(ijkl)\neq} m_{ij} m_{il} m_{jk} m_{kl} m_{ik}$	-6	+30	-60	+60	-30	+6	—
$\sum_{ijkl} m_{ij} m_{ik} m_{il} m_{jk} m_{jl} m_{kl}$	+1	-6	+15	-20	+15	-6	+1

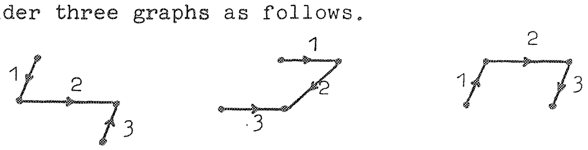
CHAPTER 2

2.1. Graphs

The notion of a graph is used in this study in two different ways. Firstly, as stated in section 1.1. in the case that all m_{ij} are non-negative integers, the internal structure of the set of points from which a sample is taken, is indicated by means of a graph. We shall call this graph from this point onwards the master-graph. The word "graph", without further indication, refers to the second type of graphs to be introduced in section 2.2. For this second purpose we use the word "graph" to denote k oriented joins, labelled 1,2,...,k between l points ($2 \leq l \leq 2k$), such that no point remains isolated (is not connected to at least one other point) and loops do not occur. Multiple joins are admitted. Such a graph will be called a (k,l)-graph.

Each join has a first and a second point, the orientation of the join being from the first to the second point. Each point is labelled by means of the labels of all joins beginning or ending in this point, with a suffix 1 or 2 to the join-label indicating whether the point is its first or second point.

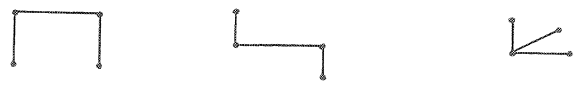
Two graphs are equivalent if they can be mapped on one another without changing the labelling of the joins and points. Non-equivalent graphs are called distinct. This gives a classification of graphs into equivalence classes. For the purposes of this study a class of equivalent graphs may and will be considered as one graph.
E.g. consider three graphs as follows.



The first two graphs are equivalent, the first and third one (and also the second and third one) are distinct.

Two distinct graphs, which can be made equivalent by means of a permutation of the join-labels and/or changing orientations in one of the two, are said to have the same configuration. This gives a further classification into configuration classes. Omitting the labelling and the orientation from a graph, a blank graph is obtained. Blank graphs obtained from distinct graphs with the same configuration will be considered as identical. Blank graphs, corresponding to different configuration classes, will be called different. Thus there is a 1-1 correspondence between blank graphs and configuration classes. The term "configuration of a blank graph" will be used to indicate the configuration class corresponding to the blank graph. By "configuration of a graph" is meant the configuration class to which the graph belongs.

As an example consider the following three blank graphs:



The first two are identical, i.e. they have the same configuration. The first and the third one (and also the second an the third one) are different (have a different configuration).

A blank graph is called connected if from every point of the graph every other point may be reached by travelling along the joins. Regardless of the orientation of the joins, a graph is called connected if its blank graph is. Every blank graph consists of one or more connected components, which have no connections between them. This decomposition is unique (cf. D. KOENIG, (1936), 15). This holds also for a graph. The components of the blank graph of a given graph are the blank graphs of its components.

For a connected (k,l) graph the relation

$$2 \leq l \leq k+1$$

holds and there is a finite number, say $q_{k,l}$, of different (k,l)-blank graphs or configurations. Let $C^{(\alpha)}_{k,l}$ be the α-th one of these ($\alpha=1,2,\ldots,q_{k,l}$). Notice that the symbol $C^{(\alpha)}_{k,l}$ always refers to a connected graph or a connected component of a graph. In case $q_{k,l}=1$, we shall write $C_{k,l}$ instead of $C^{(1)}_{k,l}$.

The configuration of a blank graph with h connected components ($1 \leq h \leq \left[\frac{l}{2}\right]$) can now be indicated symbolically by

$$\sum_{i=1}^{h} C^{(\alpha_i)}_{k_i,l_i},$$

if $C^{(\alpha_i)}_{k_i,l_i}$ is the configuration of its i-th component. If components with equal configuration occur, this may be written as

$$\sum_{j=1}^{s} g_j \, C^{(\alpha_j)}_{k_j,l_j},$$

where g_j indicates the number of components with configuration $C^{(\alpha_j)}_{k_j,l_j}$.

By $\mathcal{N}\left(\sum_{j=1}^{s} g_j C^{(\alpha_j)}_{k_j,l_j}\right)$ will be indicated the number of distinct graphs with configuration $\sum_{j=1}^{s} g_j C^{(\alpha_j)}_{k_j,l_j}$.

Theorem 2.1.1. Recursion formula for $\mathcal{N}$

$$\mathcal{N}\left(\sum_{j=1}^{s} g_j C_{k_j,l_j}^{(\alpha_j)}\right) = k! \prod_{j=1}^{s} \frac{1}{g_j!} \left\{\frac{\mathcal{N}(C_{k_j,l_j}^{(\alpha_j)})}{k_j!}\right\}^{g_j} \quad ; \quad \left(\sum_{j=1}^{s} g_j k_j = k\right).$$

Proof.

We first prove

$$\mathcal{N}(g_j C_{k_j,l_j}^{(\alpha_j)}) = \frac{(g_j k_j)!}{(g_j)!}\left\{\frac{\mathcal{N}(C_{k_j,l_j}^{(\alpha_j)})}{k_j!}\right\}^{g_j} \tag{2.1.1}$$

To obtain the number of distinct graphs with configuration $g_j C_{k_j,l_j}^{(\alpha_j)}$, the labels $1,\dots,g_j k_j$ have to be distributed over g_j identical (k_j,l_j)-components of the blank graph. This can be done in

$$\frac{(g_j k_j)!}{g_j!(k_j!)^{g_j}} \tag{2.1.2}$$

distinct ways, since each of the $g_j!$ permutations of the components leads to the same (i.e. equivalent) set of graphs in the end. In each component a set of k_j labels gives rise to $\mathcal{N}(C_{k_j,l_j}^{(\alpha_j)})$ distinct graphs; thus for each distribution of $k_j g_j$ labels over the g_j components there are

$$\left\{\mathcal{N}(C_{k_j,l_j}^{(\alpha_j)})\right\}^{g_j} \tag{2.1.3}$$

distinct graphs with the required configuration. The number of distinct graphs with configuration $g_j C_{k_j,l_j}^{(\alpha_j)}$ is now the product of (2.1.2) and (2.1.3), which proves (2.1.1).

Now consider $g_j C_{k_j,l_j}^{(\alpha_j)}$ as the j-th subgraph of $\sum_{j=1}^{s} g_j C_{k_j,l_j}^{(\alpha_j)}$. The number of ways in which k labels $1,\dots,k$ can be distributed over s different blank subgraphs such that the j-th one contains $g_j k_j$ labels $(j=1,2,\dots,s)$ is

$$\frac{k!}{(g_1k_1)!(g_2k_2)!\ \dots\ (g_sk_s)!}\ .$$

In the j-th subgraph each set of $g_j k_j$ labels gives rise to $\mathcal{N}(g_j C_{k_j,l_j}^{(\alpha_j)})$ distinct graphs. Thus

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

which upon application of (2.1.1) gives the result of the theorem.

In order to obtain a recursion formula for $\mathcal{N}(C_{k,l}^{(\alpha)})$ too, define

$$(2.1.4) \qquad \mathcal{N}\left(\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)}\right) = 0 \quad \text{if for at least one } i=1,2,\dots,h \text{ either } l_i \leq 1 \text{ or } l_i \geq k_i+2 .$$

Let

$$\mathcal{W}\left(\sum_{i=1}^{h'} C_{k_i',l_i'}^{(\alpha_i')} ; \sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)}\right)$$

be the number of distinct graphs having configuration $\sum_{i=1}^{h'} C_{k_i',l_i'}^{(\alpha_i')}$ that can arise in adding one oriented join to a given pattern consisting of a graph having configuration $\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)}$ and of $\sum_{i=1}^{h'} l_i' - \sum_{i=1}^{h} l_i$ unlabelled isolated points. As by means of Theorem (2.1.1) $\mathcal{N}\left(\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)}\right)$ can be calculated if $\mathcal{N}\left(C_{k_i,l_i}^{(\alpha_i)}\right)$, $(i=1,2,\dots,h)$ is known, we give only the recurrence relation:

$$(2.1.5) \qquad \mathcal{N}\left(C_{k+1,l+1}^{(\alpha)}\right) = \sum_{t=0}^{1} \sum_{\beta} \mathcal{W}\left(C_{k+1,l+1}^{(\alpha)} ; C_{k,l+t}^{(\beta)}\right) \mathcal{N}\left(C_{k,l+t}^{(\beta)}\right) +$$

$$+ \sum_{\substack{k_1,k_2 \\ k_1+k_2=k}} \sum_{\lfloor l_1,l_2;\ l_i=2,\dots,k_i+1,\ \sum l_i=l+1 \rfloor} \cdot$$

$$\cdot \sum_{\beta_1,\beta_2} \mathcal{W}\left(C_{k+1,l+1}^{(\alpha)} ; \sum_{i=1}^{2} C_{k_i,l_i}^{(\beta_i)}\right) \mathcal{N}\left(\sum_{i=1}^{2} C_{k_i,l_i}^{(\beta_i)}\right) .$$

We shall illustrate the use of these recursion formulae by calculating $\mathcal{N}$ for configurations up to $k = 3$.

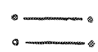
All graphs based on one join between two points are equivalent. Thus

$$(2.1.6) \qquad \mathcal{N}(C_{1,2}) = 1 \quad .$$

For $k = 2$, the following configurations are involved

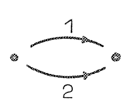
$C_{2,2}$

$C_{2,3}$

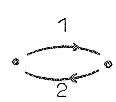


$2C_{1,2}$

Given one oriented join between two points, two distinct graphs may arise in adding one oriented join between the same points:



and



thus $\mathcal{W}(C_{2,2};C_{1,2}) = 2$, and from (2.1.5) $\mathcal{N}(C_{2,2}) = 2$.

Given one oriented join between two points, and an isolated point, four distinct graphs arise in adding one join in all possible ways to this pattern, such that a graph of configuration $C_{2,3}$ arises:

,

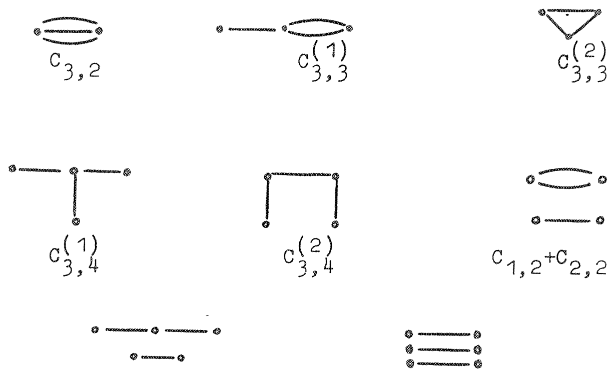
thus $\mathcal{W}(C_{2,3};C_{1,2}) = 4$ and $\mathcal{N}(C_{2,3}) = 4$.

Application of Theorem 2.1.1 with s=1, $g_1=2$, $k_1=1$, $l_1=2$, $\alpha=1$ and $\mathcal{N}(C_{1,2}) = 1$ gives $\mathcal{N}(2C_{1,2}) = 1$.

Summarizing we find for k=2

$$\mathcal{N}(C_{2,2}) = 2$$
$$(2.1.7) \qquad \mathcal{N}(C_{2,3}) = 4$$
$$\mathcal{N}(2C_{1,2}) = 1 \quad .$$

For k=3 the following configurations are involved.



$C_{3,2}$ $\quad C^{(1)}_{3,3}$ $\quad C^{(2)}_{3,3}$

$C^{(1)}_{3,4}$ $\quad C^{(2)}_{3,4}$ $\quad C_{1,2}+C_{2,2}$

$C_{1,2}+C_{2,3}$ $\quad 3C_{1,2}$

Thus e.g.

$$\mathcal{W}(C^{(1)}_{3,4};\ 2C_{1,2}) = 0 \qquad \mathcal{W}(C^{(1)}_{3,4};\ C_{2,3}) = 2 \qquad \mathcal{W}(C^{(1)}_{3,4};\ C_{2,2}) = 0$$
$$\mathcal{W}(C^{(2)}_{3,4};\ 2C_{1,2}) = 8 \qquad \mathcal{W}(C^{(2)}_{3,4};\ C_{2,3}) = 4 \qquad \mathcal{W}(C^{(2)}_{3,4};\ C_{2,2}) = 0 \ ,$$

and by (2.1.5)

$$\mathcal{N}(C^{(1)}_{3,4}) = \mathcal{N}(2C_{1,2})\mathcal{W}(C^{(1)}_{3,4};\ 2C_{1,2}) + \mathcal{N}(C_{2,3})\mathcal{W}(C^{(1)}_{3,4};\ C_{2,3}) = 8 \ .$$

Also

$$\mathcal{N}(C_{3,4}^{(2)}) = 8\mathcal{N}(2C_{1,2}) + 4\mathcal{N}(C_{2,3}) = 24 ,$$

while from Theorem 2.1.1

$$\mathcal{N}(C_{1,2} + C_{2,2}) = 6$$

In this way the following results for $k = 3$ are obtained

(2.1.8)
$$\begin{aligned} \mathcal{N}(C_{3,2}) &= 4 \\ \mathcal{N}(C_{3,3}^{(1)}) &= 24 \\ \mathcal{N}(C_{3,3}^{(2)}) &= 8 \\ \mathcal{N}(C_{3,4}^{(1)}) &= 8 \\ \mathcal{N}(C_{3,4}^{(2)}) &= 24 \\ \mathcal{N}(C_{1,2}+C_{2,2}) &= 6 \\ \mathcal{N}(C_{1,2}+C_{2,3}) &= 12 \\ \mathcal{N}(3C_{1,2}) &= 1 \quad . \end{aligned}$$

For reference purposes note that for $k=1,2,\ldots$

$$\mathcal{W}(C_{k+1,2};\, C_{k,2}) = 2$$

and therefore

(2.1.9)
$$\mathcal{N}(C_{k,2}) = 2^{k-1} \quad .$$

2.2. An expression for $(\sum w_{ij})^k$

Let w_{ij} $(i,j=1,2,\dots,n)$ be real numbers, satisfying

(2.2.1) $$w_{ii} = 0$$

and for $i \neq j$

(2.2.2) $$w_{ij} = w_{ji} .$$

From the sum

(2.2.3) $$(\sum_{ij} w_{ij})^k = \sum_{\tau_1=1}^{n} \dots \sum_{\tau_{2k}=1}^{n} w_{\tau_1,\tau_2} \dots w_{\tau_{2k-1},\tau_{2k}}$$

we consider one term

(2.2.4) $$t = w_{\tau_1,\tau_2} w_{\tau_3,\tau_4} \dots w_{\tau_{2k-1},\tau_{2k}} .$$

The subscripts occur in pairs : (τ_{2j-1},τ_{2j}), $(j=1,2,\dots,k)$. If for any j both subscripts assume the same value, t is equal to zero by (2.2.1). These terms can therefore be omitted from (2.2.3). Now consider a term t with $\tau_{2j-1} \neq \tau_{2j}$ $(j=1,2,\dots,k)$. Among the subscripts $\tau_1,\tau_2,\dots,\tau_{2k}$ l unequal numbers from $1,2,\dots,n$ occur $(2 \leq l \leq 2k)$. Call these $\lambda_1,\lambda_2,\dots,\lambda_l$ with

$$\lambda_1 < \lambda_2 < \dots < \lambda_l .$$

Each of the τ's is equal to one of the λ's; let

$$\tau_{2j-1} = \lambda_{\mu_j} \text{ and } \tau_{2j} = \lambda_{\nu_j} \quad (1 \leq \mu_j \leq l,\ 1 \leq \nu_j \leq l,\ j=1,2,\dots,k),$$

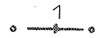
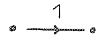
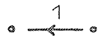
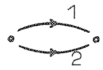
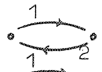
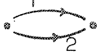
then

$$t = w_{\lambda_{\mu_1},\lambda_{\nu_1}} w_{\lambda_{\mu_2},\lambda_{\nu_2}} \dots w_{\lambda_{\mu_k},\lambda_{\nu_k}} \quad (\mu_j \neq \nu_j,\ j=1,2,\dots,k).$$

In the set of numbers $\mu_1, \mu_2, \dots, \mu_k, \nu_1, \nu_2, \dots, \nu_k$ all numbers $1,2,\dots,l$ are represented. To each such set of numbers $\mu_1, \mu_2, \dots, \mu_k, \nu_1,\nu_2,\dots,\nu_k$ - and therefore to each term t - there corresponds a (k,l)-graph in the following way. Take l points, numbered $1,2,\dots,l$. Join the points numbered μ_j and ν_j. Call this the j-th join with the points numbered μ_j and ν_j as first and second point respectively $(j=1,2,\dots,k)$. Omit the labels of the points.

Some examples of the correspondence between a term t and its (k,l)-graph are given in table 2.2.1.

Table 2.2.1. Some examples of the correspondence between terms (2.2.4) and graphs.

t	λ_1	λ_2	t	μ_1	ν_1	μ_2	ν_2	graph corresponding to t
w_{12}	1	2	w_{λ_1,λ_2}	1	2	-	-	(1)  (2)
w_{57}	5	7	w_{λ_1,λ_2}	1	2	-	-	(1)  (2)
w_{21}	1	2	w_{λ_2,λ_1}	2	1	-	-	(1)  (2)
$w_{12}w_{12}$	1	2	$w_{\lambda_1,\lambda_2}w_{\lambda_1,\lambda_2}$	1	2	1	2	(1)  (2)
$w_{12}w_{21}$	1	2	$w_{\lambda_1,\lambda_2}w_{\lambda_2,\lambda_1}$	1	2	2	1	(1)  (2)
$w_{35}w_{35}$	1	2	$w_{\lambda_1,\lambda_2}w_{\lambda_1,\lambda_2}$	1	2	1	2	(1)  (2)

From table 2.2.1 the following conclusions can be drawn

a) Two terms corresponding to equivalent graphs may or may not have the same value (cf. examples 1 and 2: terms with unequal values; examples 1 and 3: terms with equal values; in both cases equivalent graphs).

b) Two terms corresponding to distinct graphs but with the same configuration may or may not have the same value (cf. examples 4 and 5: terms with equal values because of (2.2.2); examples 5 and 6: terms with unequal values; in both cases distinct graphs with the same configuration).

Let $\theta_1,\theta_2,\ldots,\theta_l$ be l unequal numbers from $\{1,2,\ldots,n\}$ and let $\lambda_1,\lambda_2,\ldots,\lambda_l$ be the permutation of $\theta_1,\theta_2,\ldots,\theta_l$ satisfying

$$\lambda_1 < \lambda_2 < \ldots < \lambda_l \quad ,$$

then the terms

$$w_{\lambda_{\mu_1},\lambda_{\nu_1}} w_{\lambda_{\mu_2},\lambda_{\nu_2}} \cdots w_{\lambda_{\mu_k},\lambda_{\nu_k}}$$

and

$$w_{\theta_{\mu_1},\theta_{\nu_1}} w_{\theta_{\mu_2},\theta_{\nu_2}} \cdots w_{\theta_{\mu_k},\theta_{\nu_k}}$$

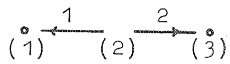
have equivalent graphs. This is evident from the way of construction of the graphs corresponding to each of the two terms; the labels of the points of one graph form a permutation of the labels of the points of the other one. As these labels are omitted anyway, both graphs are equivalent.

Example:

$l=3,\ \theta_1=7,\ \theta_2=4,\ \theta_3=9,$ thus $\lambda_1=4,\ \lambda_2=7,\ \lambda_3=9.$

$$w_{\lambda_1,\lambda_2}\, w_{\lambda_1,\lambda_3} = w_{4,7}\, w_{4,9} \longrightarrow$$

$$w_{\theta_1,\theta_2}\, w_{\theta_1,\theta_3} = w_{7,4}\, w_{7,9} \longrightarrow$$



From these considerations it follows that two terms then and only then correspond to equivalent graphs if they can be written in the form

$$w_{\theta_{\mu_1},\theta_{\nu_1}}\, w_{\theta_{\mu_2},\theta_{\nu_2}} \cdots w_{\theta_{\mu_k},\theta_{\nu_k}}$$

and

$$w_{\zeta_{\mu_1},\zeta_{\nu_1}}\, w_{\zeta_{\mu_2},\zeta_{\nu_2}} \cdots w_{\zeta_{\mu_k},\zeta_{\nu_k}} ,$$

where $\zeta_1, \zeta_2, \ldots, \zeta_l$ are also unequal numbers from $1,2,\ldots,n$, without interchanging factors w or subscripts to one w. Therefore the sum over all terms in $(\sum w_{ij})^k$, each of which corresponds to the same graph G can be written as

$$(2.2.5) \qquad \sum\left\{\prod_{j=1}^{k} w_{\tau_{2j-1},\tau_{2j}} \,\middle|\, G\right\} = \sum_{\substack{\theta_1=1 \\ (\theta_1,\ldots,\theta_l)\neq}}^{n} \cdots \sum_{\theta_l=1}^{n} w_{\theta_{\mu_1},\theta_{\nu_1}} \cdots w_{\theta_{\mu_k},\theta_{\nu_k}} ,$$

where l is the number of points of the graph G.

<u>Remark</u>: In general there are several sequences $\mu_1, \mu_2, \ldots, \mu_k,$ $\nu_1, \nu_2, \ldots, \nu_k$ corresponding to the same graph G, e.g. $\mu_1=1$ $\nu_1=2$ leads to $\circ \xrightarrow{1} \circ$, but also $\mu_1=2$ and $\nu_1=1$, again because the point labels are omitted. For use in the notation (2.2.5) one can choose one of these sequences. We shall use the one in which the μ's and ν's increase in order of appearance. Thus e.g. $w_{\theta_1,\theta_2}\, w_{\theta_1,\theta_3}$ rather than $w_{\theta_3,\theta_1}\, w_{\theta_3,\theta_2}$.

Now consider two distinct graphs G_1 and G_2 with the same configuration. This implies e.g. that the value of l is the same for both graphs. Given a set of numbers $\lambda_1, \lambda_2, \ldots, \lambda_l$ from $\{1,2,\ldots,n\}$ with $\lambda_1 < \lambda_2 < \ldots < \lambda_l$, there correspond l! terms of (2.2.3) to each of the graphs, viz. the number of permutations of the labels of the points. Because G_1 and G_2 have the same configuration, G_1 can be made equivalent to G_2 by means of a permutation of the labels of the joins (which corresponds to a permutation of the factors w in $\prod_j w_{\tau_{2j-1},\tau_{2j}}$) and by changing the direction of the orientation of some joins (which corresponds to interchanging pairs of subscripts to single factors w). A term of (2.2.3) corresponding to graph G_1 becomes therefore, with this permutation of the w's and this interchange of pairs of subscripts to single factors w, a term corresponding to graph G_2. The value of both terms is equal, because of (2.2.2). Thus to each of the l! terms with graph G_1 there corresponds one of the l! terms with graph G_2, having the same value. Thus

$$\Sigma\left\{\prod_{j=1}^{k} w_{\tau_{2j-1},\tau_{2j}} \mid G_1\right\} = \Sigma\left\{\prod_{j=1}^{k} w_{\tau_{2j-1},\tau_{2j}} \mid G_2\right\} , \tag{2.2.6}$$

which shows this sum to be equal for all distinct graphs having the same configuration. The contribution to $(\Sigma\, w_{ij})^k$ of terms corresponding to a given value of l, and to a given configuration $\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)}$, with $\sum_{i=1}^{h} k_i = k$, $\sum_{i=1}^{h} l_i = l$, is therefore

$$\mathcal{N}\left(\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)}\right) \underset{(\theta_1,\ldots,\theta_l)\neq}{\sum_{\theta_1=1}^{n} \cdots \sum_{\theta_l=1}^{n}} w_{\theta_{\mu_1},\theta_{\nu_1}} \cdots w_{\theta_{\mu_k},\theta_{\nu_k}} , \tag{2.2.7}$$

where $\mathcal{N}\left(\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)}\right)$ is the number of distinct graphs having the configuration $\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)}$ (cf. section 2.1), while the graph corresponding to the set of numbers $\mu_1, \mu_2, \ldots, \mu_k, \nu_1, \nu_2, \ldots, \nu_k$ is one of the graphs having this configuration. Symbolically we write (2.2.7) as

$$\mathcal{N}\left(\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)}\right) \Sigma^{*}\left\{w^{(1)}\ldots w^{(k)} \mid \sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)}\right\} . \tag{2.2.8}$$

Summing over all configurations with h connected components, k joins and l points - this summation will be indicated by $\Sigma''_{\lfloor \sum_{i=1}^{h} k_i = k,\ \sum_{i=1}^{h} l_i = l \rfloor}$ - summing over all h and l gives now

$$\left(\sum_{ij} w_{ij}\right)^k = \sum_{l=2}^{2k} \sum_{h=1}^{[\frac{l}{2}]} \Sigma''_{\lfloor \sum_{i=1}^{h} k_i = k,\ \sum_{i=1}^{h} l_i = l \rfloor} \mathcal{N}\left(\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)}\right) \cdot \Sigma^{*}\left\{w^{(1)}\ldots w^{(k)} \,\middle|\, \sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)}\right\} . \tag{2.2.9}$$

2.3. The quantities $\sum^{*}\{ w^{(1)}\dots w^{(k)} \mid \dots \}$.

Consider the sum

$$\sum_{\substack{\theta_1=1 \\ (\theta_1,\dots,\theta_l)\neq}}^{n} \dots \sum_{\theta_l=1}^{n} w_{\theta_{\mu_1},\theta_{\nu_1}} \dots w_{\theta_{\mu_k},\theta_{\nu_k}} \quad , \tag{2.3.1}$$

where (cf. section 2.2) in the set of numbers $\mu_1,\mu_2,\dots,\mu_k,\nu_1,\nu_2,\dots,\nu_k$ all numbers $1,2,\dots,l$ are represented in such a way that the configuration of the graph corresponding to <u>each term</u> of (2.3.1) is $\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)}$.

We now introduce a graph corresponding to a <u>sum</u>

$$\underbrace{\sum_{\theta_1=1}^{n} \dots \sum_{\theta_l=1}^{n}}_{D^*} w_{\theta_{\mu_1},\theta_{\nu_1}} \dots w_{\theta_{\mu_k},\theta_{\nu_k}} \quad , \tag{2.3.2}$$

where D^* is a restriction yet to be specified and $\mu_1,\mu_2,\dots,\mu_k,\nu_1,\nu_2,\dots,\nu_k$ are the same numbers as in (2.3.1). The graph corresponding to the sum (2.3.2) is obtained as follows. Take a set of l points and label them $1,2,\dots,l$. Join the points labelled μ_j and ν_j , $(j=1,2,\dots,k)$. A graph is then obtained <u>with labelled points</u>.

If the restriction D^* is: "$\theta_1, \theta_2,\dots, \theta_l$ assume unequal values", the blank graph corresponding to the sum is identical to the blank graph corresponding to each of the terms of the sum.
If the restriction D^* is defined otherwise, this is not necessarily so.

The value of the sum remains unchanged if the order of the w's is changed, and (or) the pair of subscripts to one w is interchanged for some w's. Therefore the value of the sum depends only on the blank graph corresponding to the sum, the restriction D^*, and the point-labels, as far as they are needed to specify D^*.

If the restriction D^* is "$\theta_1, \theta_2,\dots, \theta_l$ assume unequal values" the notation for (2.3.2) has been introduced already:

$$\sum^{*}\left\{ w^{(1)}\dots w^{(k)} \mid \sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)} \right\} . \tag{2.3.3}$$

If the restriction D^* is empty the sum will be indicated as

$$\sum \left\{ w^{(1)}\dots w^{(k)} \mid \sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)} \right\} , \tag{2.3.4}$$

where $\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)}$ is the configuration of the blank graph corresponding to the sum. Finally if the restriction D^* is that the subscripts

$\theta_{l_1+\ldots+l_u+1}$, $\theta_{l_1+\ldots+l_u+2}$, ..., θ_l assume unequal values and no restriction is imposed on $\theta_1, \theta_2, \ldots, \theta_{l_1+\ldots+l_u}$ $(1 \leq u < h)$, we need only distinguish between these two sets of point-labels to specify D^*, and write

$$(2.3.5) \qquad \sum{}^* \left\{ w^{(1)} \ldots w^{(k)} \;\middle|\; \sum_{i=1}^{u} C_{k_i,l_i}^{(\alpha_i)} + \left(\sum_{i=u+1}^{h} C_{k_i,l_i}^{(\alpha_i)} \right) \right\}.$$

Note that

$$(2.3.6) \qquad \sum{}^* \left\{ w^{(1)} \ldots w^{(k)} \;\middle|\; \sum_{i=1}^{u} C_{k_i,l_i}^{(\alpha_i)} + \left(\sum_{i=u+1}^{h} C_{k_i,l_i}^{(\alpha_i)} \right) \right\} =$$

$$\sum \left\{ w^{(1)} \ldots w^{(k_1+\ldots+k_u)} \;\middle|\; \sum_{i=1}^{u} C_{k_i,l_i}^{(\alpha_i)} \right\} .$$

$$. \sum{}^* \left\{ w^{(k_1+\ldots+k_u+1)} \ldots w^{(k)} \;\middle|\; \sum_{i=u+1}^{h} C_{k_i,l_i}^{(\alpha_i)} \right\} .$$

We now remove the restrictions imposed on the sum (2.3.1). Ignoring the inequalities on e.g. θ_l leads to

$$(2.3.7) \qquad \underset{(\theta_1,\ldots,\theta_{l-1})\neq}{\sum_{\theta_1=1}^{n} \cdots \sum_{\theta_l=1}^{n}} w_{\theta_{\mu_1},\theta_{\nu_1}} \ldots w_{\theta_{\mu_k},\theta_{\nu_k}} ,$$

where the blank-graph corresponding to the sum still has the configuration $\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)}$. The difference between (2.3.7) and (2.3.1) consists of $(l-1)$ sums like (2.3.1), but where in the j-th one θ_l is replaced by θ_j $(j=1,\ldots,l-1)$. The configuration of the graphs corresponding to the j-th sum can be obtained from the graph (with the labelled points) corresponding to the sum (2.3.1) by making points l and j coincide. The following three cases may arise.

1) In the graph with configuration $\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)}$ points l and j are connected by a join. This means that θ_l and θ_j occur as subscripts to one w. Taking $\theta_l=\theta_j$ the sum concerned is equal to zero by (2.2.1) and is therefore omitted.
2) Points l and j belong to one connected component, say to the one

with configuration $C_{k_1,l_1}^{(\alpha_1)}$, but are not connected by a join. Making points l and j coincide, the configuration thus arising is of the form $C_{k_1,l_1-1}^{(\beta)} + \sum_{i=2}^{h} C_{k_i,l_i}^{(\alpha_i)}$.

3) Points l and j belong to different connected components, say to $C_{k_1,l_1}^{(\alpha_1)}$ and $C_{k_2,l_2}^{(\alpha_2)}$ respectively.
The configuration of the graph obtained by making points l and j coincide is of the form $C_{k_1+k_2,\ l_1+l_2-1}^{(\beta)} + \sum_{i=3}^{h} C_{k_i,l_i}^{(\alpha_i)}$.

Thus (2.3.1) in which l θ's had to be unequal, is expressed as a number of sums in each of which (l-1) θ's have to be unequal. Repeating this procedure of removing inequalities imposed on summation indices leads to

(2.3.8) $\sum^{*}\left\{ w^{(1)}\dots w^{(k)} \mid \sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)} \right\} =$

$$\sum_{h'=1}^{h} \sum_{l'=2h'}^{l} \sum\nolimits^{*}_{\lfloor \sum_{i=1}^{h'} k'_i=k,\ \sum_{i=1}^{h'} l'_i=l' \rfloor} \mathcal{A}\left(\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)} \ ; \ \sum_{i=1}^{h'} C_{k'_i,l'_i}^{(\alpha'_i)} \right) .$$

$$. \sum \left\{ w^{(1)}\dots w^{(k)} \mid \sum_{i=1}^{h'} C_{k'_i,l'_i}^{(\alpha'_i)} \right\} .$$

The coefficients $\mathcal{A}\left(\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)} \ ; \ \sum_{i=1}^{h'} C_{k'_i,l'_i}^{(\alpha'_i)} \right)$ are defined by (2.3.8) only for the case

$$h \geqq h' ,$$

(2.3.9) $$\sum_{i=1}^{h} k_i = \sum_{i=1}^{h'} k'_i = k ,$$

$$\sum_{i=1}^{h} l_i = l \geqq l' = \sum_{i=1}^{h'} l'_i .$$

If for two configurations $\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)}$ and $\sum_{i=1}^{h'} C_{k'_i,l'_i}^{(\alpha'_i)}$ any of the relations (2.3.9) does not hold we define

(2.3.10) $$\mathcal{A}\left(\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)} \ ; \ \sum_{i=1}^{h'} C_{k'_i,l'_i}^{(\alpha'_i)} \right) = 0 .$$

It is evident from the description given that even if (2.3.9) is satisfied, (2.3.10) continues to hold if the blank graphs with configurations $\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)}$ and $\sum_{i=1}^{h'} C_{k_i',l_i'}^{(\alpha_i')}$ are different, and the latter cannot be obtained from the former by making points coincide. This is certainly the case if either $\sum_{i=1}^{h'} C_{k_i',l_i'}^{(\alpha_i')}$ has a component containing less than min $(k_1, k_2,\dots,k_h)$ joins or $\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)}$ has a component containing more than max $(k_1', k_2',\dots,k_{h'}')$ joins. This proves

Lemma 2.3.1

If for two configurations $\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)}$ and $\sum_{i=1}^{h'} C_{k_i',l_i'}^{(\alpha_i')}$ (2.3.9) is satisfied, then

$$\mathcal{A}\left(\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)} \; ; \; \sum_{i=1}^{h'} C_{k_i',l_i'}^{(\alpha_i')}\right) = 0$$

if either

(2.3.11) $\quad \min (k_1,k_2,\dots,k_h) > \min (k_1', k_2',\dots,k_{h'}')$

or

(2.3.12) $\quad \max (k_1, k_2,\dots,k_h) > \max (k_1', k_2',\dots,k_{h'}')$.

Consider again the sum (2.3.1) with k replaced by (k+g) and l by (l+2g). Suppose that the graph corresponding to the sum has configuration

$$\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)} + gC_{1,2} \;,$$

where $k_i \geq 2$ for $i=1,2,\dots,h$, $\sum_{i=1}^{h} k_i=k$, $\sum_{i=1}^{h} l_i=l$. Let the labelling of the points of the graph be as follows. The points of the g connected components with configuration $C_{1,2}$ carry the numbers 1 and 2; 3 and 4;.. ...; 2g - 1 and 2g respectively.

This means that e.g. the summation-subscripts θ_1 and θ_2 occur as subscripts to one w, and both occur only once in the product. The other l points of the graph carry the numbers 2g+1, 2g+2, ..., 2g+l. In our shorthand notation (2.3.1) is written

(2.3.13) $$\sum{}^{*} \left\{ w^{(1)}\dots w^{(k+g)} \;\middle|\; \sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)} + gC_{1,2} \right\} .$$

Now remove the inequalities imposed on θ_1 and θ_2 .

The inequality $\theta_1 \neq \theta_2$ imposed on (2.3.13) can be dropped without (2.3.13) changing in value, as θ_1 and θ_2 occur as subscripts to one w.

Ignoring the inequalities on θ_1 and θ_2 leads to

$$(2.3.14)\qquad \sum{}^{*}\left\{ w^{(1)}\ldots w^{(k+g)} \;\Big|\; \left(\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)} + (g-1)C_{1,2} \right) + C_{1,2} \right\},$$

which by (2.3.6) is equal to

$$(2.3.15)\qquad \sum_{ij} w_{ij}\cdot \sum{}^{*}\left\{ w^{(1)}\ldots w^{(k+g-1)} \;\Big|\; \sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)} + (g-1)C_{1,2} \right\}.$$

The difference between (2.3.15) and (2.3.13) consists of sums like (2.3.1) for each of the following cases

i	$\theta_1 \neq \theta_3, \theta_4, \ldots, \theta_{l+2g}$ and $\theta_2 = \theta_j$	$(j = 3,4,\ldots, l+2g)$
ii	$\theta_1 = \theta_j$ and $\theta_2 \neq \theta_3, \theta_4, \ldots, \theta_{l+2g}$	$(j = 3,4,\ldots, l+2g)$
iii	$\theta_1 = \theta_{j_1}$ and $\theta_2 = \theta_{j_2}$	$(j_1, j_2 = 3,4,\ldots, l+2g)$.

We start with the $(l+2g-2)^2$ sums corresponding to iii. The $(l+2g-2)$ sums with

$$\theta_1 = \theta_2 = \theta_j \qquad (j = 3,4,\ldots, l+2g)$$

are identically equal to zero by (2.2.1). The $2(g-1)$ sums corresponding to

$$\begin{array}{ll}
\theta_1 = \theta_3 & \theta_2 = \theta_4 \\
\theta_1 = \theta_4 & \theta_2 = \theta_3 \\
\theta_1 = \theta_5 & \theta_2 = \theta_6 \\
\vdots & \vdots \\
\theta_1 = \theta_{2g} & \theta_2 = \theta_{2g-1}
\end{array}$$

all have a graph with the configuration $\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)} + C_{2,2} + (g-2)C_{1,2}$, thus giving as a contribution to the difference between (2.3.15) and (2.3.13)

$$2(g-1)\sum{}^{*}\left\{ w^{(1)}\ldots w^{(k+g)} \;\Big|\; \sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)} + C_{2,2} + (g-2)C_{1,2} \right\}$$

There are $4(g-1)(g-2)$ sums, viz. corresponding to

$$\begin{array}{ll}
\theta_1 = \theta_3\,, & \theta_2 = \theta_5, \theta_6, \ldots, \theta_{2g}, \\
\theta_1 = \theta_4\,, & \theta_2 = \theta_5, \theta_6, \ldots, \theta_{2g}, \\
\theta_1 = \theta_5\,, & \theta_2 = \theta_3, \theta_4, \theta_7, \ldots, \theta_{2g}, \\
\vdots & \vdots \\
\theta_1 = \theta_{2g}, & \theta_2 = \theta_3, \theta_4, \ldots, \theta_{2g-2},
\end{array}$$

all corresponding to graphs with configuration $\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)} + C_{3,4}^{(2)} + (g-3)C_{1,2}$, thus giving as a contribution to the difference between (2.3.15) and (2.3.13)

$$4(g-1)(g-2)\ \Sigma^*\left\{ w^{(1)}\dots w^{(k+g)}\ \middle|\ \sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)} + C_{3,4}^{(2)} + (g-3)C_{1,2}\right\} .$$

There are $4(g-1)l$ sums in iii, viz.

$$\theta_1 = \theta_3,\ \theta_4,\dots,\ \theta_{2g} \quad ,\ \theta_2 = \theta_{2g+1},\ \theta_{2g+2},\dots,\ \theta_{2g+l}$$

$$\theta_1 = \theta_{2g+1},\ \theta_{2g+2},\dots,\ \theta_{2g+l} \ ,\ \theta_1 = \theta_3,\ \theta_4,\dots,\ \theta_{2g} \quad ,$$

corresponding to graphs with a configuration $\sum_{i=1}^{h} C_{k_i^*,l_i^*}^{(\alpha_i^*)} + (g-2)C_{1,2}$,

where $k_i^* \geq 2$, $\sum_{i=1}^{h} k_i^* = k+2$ and $\sum_{i=1}^{h} l_i^* = l+2$.

The multiplicity with which sums corresponding to graphs with configuration $\sum_{i=1}^{h} C_{k_i^*,l_i^*}^{(\alpha_i^*)} + (g-2)C_{1,2}$ occur may be denoted by

$4(g-1)\ \mathcal{M}_o\left(\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)}\ ;\ \sum_{i=1}^{h} C_{k_i^*,l_i^*}^{(\alpha_i^*)}\right)$, where $\mathcal{M}_o(\ ;\ )$ does not depend on g.

Finally there are $l(l-1)$ sums in iii, viz.

$$\theta_1 = \theta_{2g+1},\ \theta_{2g+2},\dots,\ \theta_{2g+l},\quad \theta_2 = \theta_{2g+1},\ \theta_{2g+2},\dots,\ \theta_{2g+l},\qquad \theta_1 \neq \theta_2 ,$$

corresponding to graphs with a configuration $\sum_{i=1}^{h^*} C_{k_i^*,l_i^*}^{(\alpha_i^*)} + (g-1)C_{1,2}$,

where $h-1 \leq h^* \leq h$, $k_i^* \geq 2$, $\sum_{i=1}^{h^*} k_i^* = k+1$ and $\sum_{i=1}^{h^*} l_i^* = l$. The multiplicity

$2.\mathcal{M}_1\left(\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)}\ ;\ \sum_{i=1}^{h^*} C_{k_i^*,l_i^*}^{(\alpha_i^*)}\right)$ with which sums corresponding to graphs

with configuration $\sum_{i=1}^{h^*} C_{k_i^*,l_i^*}^{(\alpha_i^*)} + (g-1)C_{1,2}$ occur does not depend on g.

With regard to i and ii the $4(g-1)$ sums with

$$\theta_1 \neq \theta_3,\ \theta_4,\dots,\ \theta_{l+2g}\ ,\ \theta_2 = \theta_3,\ \theta_4,\dots,\ \theta_{2g}$$

$$\theta_1 = \theta_3,\ \theta_4,\dots,\ \theta_{2g}\quad ,\ \theta_2 \neq \theta_3,\ \theta_4,\dots,\ \theta_{l+2g}$$

correspond to graphs with configuration $\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)} + C_{2,3} + (g-2)C_{1,2}$,

thus contributing to the difference between (2.3.15) and (2.3.13)

$$4(g-1)\ \Sigma^*\left\{ w^{(1)}\dots w^{(k+g)}\ \middle|\ \sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)} + C_{2,3} + (g-2)C_{1,2}\right\} .$$

The 21 sums corresponding to the remaining cases in i and ii, viz.

$\theta_1 \neq \theta_3, \theta_4, \dots, \theta_{1+2g}$, $\theta_2 = \theta_{2g+1}, \theta_{2g+2}, \dots, \theta_{2g+1}$

$\theta_1 = \theta_{2g+1}, \theta_{2g+2}, \dots, \theta_{2g+1}$, $\theta_2 \neq \theta_3, \theta_4, \dots, \theta_{1+2g}$,

correspond to graphs with a configuration

$\sum_{i=1}^{h} C_{k_i^*,l_i^*}^{(\alpha_i^*)} + (g-1)C_{1,2}$, where $k_i^* \geq 2$, $\sum_{i=1}^{h} k_i^* = k+1$, $\sum_{i=1}^{h} l_i^* = l+1$,

while again the multiplicity $2.\mathcal{M}_2\left(\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)} ; \sum_{i=1}^{h} C_{k_i^*,l_i^*}^{(\alpha_i^*)}\right)$ with which sums corresponding to graphs with configuration $\sum_{i=1}^{h} C_{k_i^*,l_i^*}^{(\alpha_i^*)} + (g-1)C_{1,2}$ occur does not depend on g.

We have now established for a configuration $\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)}$ with $\sum_{i=1}^{h} k_i = k$, $\sum_{i=1}^{h} l_i = l$ and $g > 0$

$$(2.3.16)\quad \sum{}^{*}\left\{ w^{(1)}\dots w^{(k+g)} \mid \sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)} + gC_{1,2} \right\} =$$

$$\sum_{ij} w_{ij}.\sum{}^{*}\left\{ w^{(1)}\dots w^{(k+g-1)} \mid \sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)} + (g-1)C_{1,2} \right\} +$$

$$-\ 2(g-1)\sum{}^{*}\left\{ w^{(1)}\dots w^{(k+g)} \mid \sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)} + C_{2,2} + (g-2)C_{1,2} \right\} +$$

$$-\ 4(g-1)\sum{}^{*}\left\{ w^{(1)}\dots w^{(k+g)} \mid \sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)} + C_{2,3} + (g-2)C_{1,2} \right\} +$$

$$-\ 4(g-1)(g-2)\sum{}^{*}\left\{ w^{(1)}\dots w^{(k+g)} \mid \sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)} + C_{3,4}^{(2)} + (g-3)C_{1,2} \right\} +$$

$$-\ 4(g-1)\sum{}''_{\lfloor k_i^* \geq 2,\ \sum_{i=1}^{h} k_i^* = k+2,\ \sum_{i=1}^{h} l_i^* = l+2 \rfloor} \mathcal{M}_0\left(\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)} ; \sum_{i=1}^{h} C_{k_i^*,l_i^*}^{(\alpha_i^*)}\right).$$

$$.\sum{}^{*}\left\{ w^{(1)}\dots w^{(k+g)} \mid \sum_{i=1}^{h} C_{k_i^*,l_i^*}^{(\alpha_i^*)} + (g-2)C_{1,2} \right\} +$$

$$-2\sum_{s=1}^{2}\ \sum_{h^*=h+s-2}^{h}\ \sum{}''_{\lfloor k_i^* \geq 2,\ \sum_{i=1}^{h^*} k_i^* = k+1,\ \sum_{i=1}^{h^*} l_i^* = l+s-1 \rfloor} \mathcal{M}_s\left(\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)} ; \sum_{i=1}^{h^*} C_{k_i^*,l_i^*}^{(\alpha_i^*)}\right).$$

$$.\sum{}^{*}\left\{ w^{(1)}\dots w^{(k+g)} \mid \sum_{i=1}^{h^*} C_{k_i^*,l_i^*}^{(\alpha_i^*)} + (g-1)C_{1,2} \right\}.$$

We now prove

<u>Lemma 2.3.2</u>

Given two configurations $\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)}$ and $\sum_{i=1}^{h'} C_{k_i',l_i'}^{(\alpha_i')}$ and two non-negative integers g and g', such that

a) $k_i \geq 2$, $i = 1,2,\dots, h$ and $k_i' \geq 2$, $i = 1,2,\dots, h'$,

b) $g + \sum_{i=1}^{h} k_i = g' + \sum_{i=1}^{h'} k_i' = k + g$,

c) $2g + \sum_{i=1}^{h} l_i \geq 2g' + \sum_{i=1}^{h'} l_i'$,

then

$$\mathcal{A}\left(\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)} + gC_{1,2} \; ; \; \sum_{i=1}^{h'} C_{k_i',l_i'}^{(\alpha_i')} + g'C_{1,2}\right) =$$

$$= \begin{cases} 0 & \text{if } 0 \leq g < g' \\ \binom{g}{g'} \mathcal{A}\left(\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)} + (g-g')C_{1,2} \; ; \; \sum_{i=1}^{h'} C_{k_i',l_i'}^{(\alpha_i')}\right) , & \text{if } g \geq g' \geq 0 . \end{cases}$$

<u>Proof</u>

The lemma is obvious for $0 \leq g < g'$ since the number of components of a blank graph having configuration $C_{1,2}$ cannot increase when points are made to coincide (cf. the remark leading to lemma 2.3.1). As the lemma is trivial for $g'=0$ we consider $g \geq g' > 0$. Apply (2.3.8) to both sides of (2.3.16) and consider the coefficients of

$$\sum \left\{ w^{(1)} \dots w^{(k+g)} \mid \sum_{i=1}^{h'} C_{k_i',l_i'}^{(\alpha_i')} + g'C_{1,2} \right\} . \tag{2.3.17}$$

As (2.3.17) is equal to

$$\sum_{ij} w_{ij} \cdot \sum \left\{ w^{(1)} \dots w^{(k+g-1)} \mid \sum_{i=1}^{h'} C_{k_i',l_i'}^{(\alpha_i')} + (g'-1)C_{1,2} \right\} ,$$

the coefficient of (2.3.17) in the expression obtained by applying (2.3.8) to the first sum in the righthand member of (2.3.16) is equal to

$$\mathcal{A}\left(\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)} + (g-1)C_{1,2} \; ; \; \sum_{i=1}^{h'} C_{k_i',l_i'}^{(\alpha_i')} + (g'-1)C_{1,2}\right) .$$

Thus one obtains the recurrence relation

(2.3.18)
$$\mathcal{A}\Big(\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)} + gC_{1,2} \;;\; \sum_{i=1}^{h'} C_{k'_i,l'_i}^{(\alpha'_i)} + g'C_{1,2}\Big) =$$
$$\mathcal{A}\Big(\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)} + (g-1)C_{1,2} \;;\; \sum_{i=1}^{h'} C_{k'_i,l'_i}^{(\alpha'_i)} + (g'-1)C_{1,2}\Big) +$$
$$-2(g-1)\mathcal{A}\Big(\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)} + C_{2,2} + (g-2)C_{1,2} \;;\; \sum_{i=1}^{h'} C_{k'_i,l'_i}^{(\alpha'_i)} + g'C_{1,2}\Big) +$$
$$-4(g-1)\mathcal{A}\Big(\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)} + C_{2,3} + (g-2)C_{1,2} \;;\; \sum_{i=1}^{h'} C_{k'_i,l'_i}^{(\alpha'_i)} + g'C_{1,2}\Big) +$$
$$-4(g-1)(g-2)\mathcal{A}\Big(\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)} + C_{3,4}^{(2)} + (g-3)C_{1,2} \;;\; \sum_{i=1}^{h'} C_{k'_i,l'_i}^{(\alpha'_i)} + g'C_{1,2}\Big) +$$
$$-4(g-1) \sum\nolimits''_{\lfloor k^*_i \geq 2,\ \sum_{i=1}^{h} k^*_i = k+2,\ \sum_{i=1}^{h} l^*_i = l+2 \rfloor} \mathcal{M}_0\Big(\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)} \;;\; \sum_{i=1}^{h} C_{k^*_i,l^*_i}^{(\alpha^*_i)}\Big) \,.$$
$$.\mathcal{A}\Big(\sum_{i=1}^{h} C_{k^*_i,l^*_i}^{(\alpha^*_i)} + (g-2)C_{1,2} \;;\; \sum_{i=1}^{h'} C_{k'_i,l'_i}^{(\alpha'_i)} + g'C_{1,2}\Big) +$$
$$-2\sum_{s=1}^{2} \sum_{h^*=h+s-2}^{h} \sum\nolimits''_{\lfloor k^*_i \geq 2,\ \sum_{i=1}^{h^*} k^*_i = k+1,\ \sum_{i=1}^{h^*} l^*_i = l+s-1 \rfloor} \mathcal{M}_s\Big(\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)} \;;\; \sum_{i=1}^{h^*} C_{k^*_i,l^*_i}^{(\alpha^*_i)}\Big) .$$
$$.\mathcal{A}\Big(\sum_{i=1}^{h^*} C_{k^*_i,l^*_i}^{(\alpha^*_i)} + (g-1)C_{1,2} \;;\; \sum_{i=1}^{h'} C_{k'_i,l'_i}^{(\alpha'_i)} + g'C_{1,2}\Big) .$$

For $0 < g' = g$ one obtains from (2.3.18) using the first part of the lemma (i.e. the case where $0 \leq g < g'$) and (2.3.10)
$$\mathcal{A}\Big(\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)} + g'C_{1,2} \;;\; \sum_{i=1}^{h'} C_{k'_i,l'_i}^{(\alpha'_i)} + g'C_{1,2}\Big) =$$
$$= \mathcal{A}\Big(\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)} + (g'-1)C_{1,2} \;;\; \sum_{i=1}^{h'} C_{k'_i,l'_i}^{(\alpha'_i)} + (g'-1)C_{1,2}\Big) =$$
repeating this procedure
$$= \mathcal{A}\Big(\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)} \;;\; \sum_{i=1}^{h'} C_{k'_i,l'_i}^{(\alpha'_i)}\Big) \quad ,$$
which proves the second part of the lemma for $g=g'$. Suppose the lemma is true for $0 < g' \leq g \leq g'+r-1$ $(r \geq 1)$; we shall then prove it to be true for $g=g'+r$ too.

For $g' \leq g \leq g'+r-1$ it follows from the lemma (which was supposed to be true for these values of g) :

(2.3.19)
$$\mathcal{A}\Big(\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)} + gC_{1,2} \;;\; \sum_{i=1}^{h'} C_{k'_i,l'_i}^{(\alpha'_i)} + g'C_{1,2}\Big) =$$
$$= \frac{g}{g'}\, \mathcal{A}\Big(\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)} + (g-1)C_{1,2} \;;\; \sum_{i=1}^{h'} C_{k'_i,l'_i}^{(\alpha'_i)} + (g'-1)C_{1,2}\Big).$$

Now take $g=g'+r$ in (2.3.18) and apply (2.3.19) to all but the first term of the righthand member of (2.3.18). This leads to

$$(2.3.20)\qquad \mathcal{A}\Big(\sum_{i=1}^{h} C^{(\alpha_i)}_{k_i,l_i} + (g'+r)C_{1,2}\ ;\ \sum_{i=1}^{h'} C^{(\alpha_i')}_{k_i',l_i'} + g'C_{1,2}\Big) =$$

$$\mathcal{A}\Big(\sum_{i=1}^{h} C^{(\alpha_i)}_{k_i,l_i} + (g'+r-1)C_{1,2}\ ;\ \sum_{i=1}^{h'} C^{(\alpha_i')}_{k_i',l_i'} + (g'-1)C_{1,2}\Big) +$$

$$+\frac{g'+r-1}{g'}\Big[-2(g'+r-2)\mathcal{A}\Big(\sum_{i=1}^{h} C^{(\alpha_i)}_{k_i,l_i} + C_{2,2} + (g'+r-3)C_{1,2}\ ;\ \sum_{i=1}^{h'} C^{(\alpha_i')}_{k_i',l_i'} + (g'-1)C_{1,2}\Big) +$$

$$-4(g'+r-2)\mathcal{A}\Big(\sum_{i=1}^{h} C^{(\alpha_i)}_{k_i,l_i} + C_{2,3} + (g'+r-3)C_{1,2}\ ;\ \sum_{i=1}^{h'} C^{(\alpha_i')}_{k_i',l_i'} + (g'-1)C_{1,2}\Big) +$$

$$-4(g'+r-2)(g'+r-3)\mathcal{A}\Big(\sum_{i=1}^{h} C^{(\alpha_i)}_{k_i,l_i} + C^{(2)}_{3,4} + (g'+r-4)C_{1,2}\ ;\ \sum_{i=1}^{h'} C^{(\alpha_i')}_{k_i',l_i'} + (g'-1)C_{1,2}\Big) +$$

$$-4(g'+r-2)\sum{}''_{\lfloor k_i^*\geq 2,\ \sum_{i=1}^{h} k_i^*=k+2,\ \sum_{i=1}^{h} l_i^*=l+2\rfloor}\ \mathcal{M}_0\Big(\sum_{i=1}^{h} C^{(\alpha_i)}_{k_i,l_i}\ ;\ \sum_{i=1}^{h} C^{(\alpha_i^*)}_{k_i^*,l_i^*}\Big)\ .$$

$$.\mathcal{A}\Big(\sum_{i=1}^{h} C^{(\alpha_i^*)}_{k_i^*,l_i^*} + (g'+r-3)C_{1,2}\ ;\ \sum_{i=1}^{h'} C^{(\alpha_i')}_{k_i',l_i'} + (g'-1)C_{1,2}\Big) +$$

$$-2\sum_{s=1}^{2}\ \sum_{h^*=h+s-2}^{h}\ \sum{}''_{\lfloor k_i^*\geq 2,\ \sum_{i=1}^{h^*} k_i^*=k+1,\ \sum_{i=1}^{h^*} l_i^*=l+s-1\rfloor}\ \mathcal{M}_s\Big(\sum_{i=1}^{h} C^{(\alpha_i)}_{k_i,l_i}\ ;\ \sum_{i=1}^{h^*} C^{(\alpha_i^*)}_{k_i^*,l_i^*}\Big).$$

$$.\mathcal{A}\Big(\sum_{i=1}^{h^*} C^{(\alpha_i^*)}_{k_i^*,l_i^*} + (g'+r-2)C_{1,2}\ ;\ \sum_{i=1}^{h'} C^{(\alpha_i')}_{k_i',l_i'} + (g'-1)C_{1,2}\Big)\Big]\ .$$

For $g' > 1$ the terms inside the square brackets in (2.3.20) are by (2.3.18) equal to

$$\mathcal{A}\Big(\sum_{i=1}^{h} C^{(\alpha_i)}_{k_i,l_i} + (g'+r-1)C_{1,2}\ ;\ \sum_{i=1}^{h'} C^{(\alpha_i')}_{k_i',l_i'} + (g'-1)C_{1,2}\Big) +$$

$$-\mathcal{A}\Big(\sum_{i=1}^{h} C^{(\alpha_i)}_{k_i,l_i} + (g'+r-2)C_{1,2}\ ;\ \sum_{i=1}^{h'} C^{(\alpha_i')}_{k_i',l_i'} + (g'-2)C_{1,2}\Big)\ ,$$

thus from (2.3.20) the following recursion relation is obtained

$$\mathcal{A}\left(\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)} + (g'+r)C_{1,2} \; ; \; \sum_{i=1}^{h'} C_{k_i',l_i'}^{(\alpha_i')} + g'C_{1,2}\right) + \tag{2.3.21}$$

$$- \left(1+ \frac{g'+r-1}{g'}\right) \mathcal{A}\left(\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)} + (g'+r-1)C_{1,2} \; ; \; \sum_{i=1}^{h'} C_{k_i',l_i'}^{(\alpha_i')} + (g'-1)C_{1,2}\right) +$$

$$+ \frac{g'+r-1}{g'} \mathcal{A}\left(\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)} + (g'+r-2)C_{1,2} \; ; \; \sum_{i=1}^{h'} C_{k_i',l_i'}^{(\alpha_i')} + (g'-2)C_{1,2}\right) = 0 \; .$$

For $g'=1$, (2.3.21) remains valid if the last term is simply omitted, thus giving

$$\mathcal{A}\left(\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)} + (r+1)C_{1,2} \; ; \; \sum_{i=1}^{h'} C_{k_i',l_i'}^{(\alpha_i')} + C_{1,2}\right) =$$

$$(r+1) \mathcal{A}\left(\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)} + rC_{1,2} \; ; \; \sum_{i=1}^{h'} C_{k_i',l_i'}^{(\alpha_i')}\right) .$$

Taking $g'=2,3,\ldots$ successively, the result of the second part of the lemma is found for $g=g'+r$. By induction the second part of lemma 2.3.2 then follows.

Lemma 2.3.3

$$\sum_{u=0}^{r+s} \mathcal{A}\left(2rC_{1,2} + tC_{2,2} + sC_{2,3} \; ; \; (t+u)C_{2,2} + (r+s-u)C_{2,3}\right) y^u =$$

$$= (-1)^r \frac{(2r)!}{r!} (1-y)^s (2-y)^r \; .$$

Proof

Consider the sum

$$\sum{}^{*} \left\{ w^{(1)} \ldots w^{(2r+2s+2t)} \mid 2rC_{1,2} + tC_{2,2} + sC_{2,3} \right\} . \tag{2.3.22}$$

In the summation $4r+3s+2t$ summation-subscripts are involved, viz. the number of points of the graph corresponding to the sum. Let the points of the graph be numbered as follows. The first and second points of the connected components with configuration $C_{1,2}$ are 1 and 2; 3 and 4; .. ...; $4r-1$ and $4r$ respectively; the points of the connected components with configuration $C_{2,3}$ are $4r+1$, $4r+2$ and $4r+3$; ...; $4r+3s-2$, $4r+3s-1$ and $4r+3s$, where each time the second number refers to the point where the two joins of the component meet. Finally the points of the connected components with configuration $C_{2,2}$ are $4r+3s+1$ and $4r+3s+2$; ...; $4r+3s+2t-1$ and $4r+3s+2t$.

In (2.3.22) remove the inequalities on the summation-indices $\theta_{4r+3s+2t-1}$ and $\theta_{4r+3s+2t}$.

Then

$$(2.3.23) \qquad \Sigma^*\left\{w^{(1)}\ldots w^{(2r+2s+2t)} \mid 2rC_{1,2} + tC_{2,2} + sC_{2,3}\right\} =$$

$$= \sum_{ij} w_{ij}^2 \, \Sigma^*\left\{w^{(1)}\ldots w^{(2r+2s+2t-2)} \mid 2rC_{1,2} + (t-1)C_{2,2} + sC_{2,3}\right\} +$$

- sums corresponding to graphs having at least one connected component with three or more joins.

Apply (2.3.16) to both sides of (2.3.23), collect the coefficients of

$$\Sigma\left\{w^{(1)}\ldots w^{(2r+2s+2t)} \mid (t+u)C_{2,2} + (r+s-u)C_{2,3}\right\}$$

or (what is just the same)

$$\sum_{ij} w_{ij}^2 \, \Sigma\left\{w^{(1)}\ldots w^{(2r+2s+2t-2)} \mid (t+u-1)C_{2,2} + (r+s-u)C_{2,3}\right\} ,$$

and apply (2.3.12) to obtain

$$(2.3.24) \qquad \mathcal{A}\left(2rC_{1,2} + tC_{2,2} + sC_{2,3} \,;\, (t+u)C_{2,2} + (r+s-u)C_{2,3}\right) =$$

$$= \mathcal{A}\left(2rC_{1,2} + (t-1)C_{2,2} + sC_{2,3} \,;\, (t+u-1)C_{2,2} + (r+s-u)C_{2,3}\right) ,$$

repeating this procedure,

$$= \mathcal{A}\left(2rC_{1,2} + sC_{2,3} \,;\, uC_{2,2} + (r+s-u)C_{2,3}\right) .$$

Consider now

$$(2.3.25) \qquad \Sigma^*\left\{w^{(1)}\ldots w^{(2r+2s)} \mid 2rC_{1,2} + sC_{2,3}\right\} ,$$

where the points of the connected components with configuration $C_{1,2}$ and $C_{2,3}$ are numbered as before. Removing the inequalities on $\theta_{4r+3s-2}$, $\theta_{4r+3s-1}$ and θ_{4r+3s} leads to

$$(2.3.26) \qquad \Sigma^*\left\{w^{(1)}\ldots w^{(2r+2s)} \mid 2rC_{1,2} + sC_{2,3}\right\} =$$

$$= \sum_{ijk} m_{ij}\, m_{ik} \cdot \Sigma^*\left\{w^{(1)}\ldots w^{(2r+2s-2)} \mid 2rC_{1,2} + (s-1)C_{2,3}\right\} +$$

$$- \quad \Sigma^*\left\{w^{(1)}\ldots w^{(2r+2s)} \mid 2rC_{1,2} + C_{2,2} + (s-1)C_{2,3}\right\} +$$

- sums corresponding to graphs having at least one connected component with three or more joins.

The second term in the righthand member of (2.3.26) arises in taking $\theta_{4r+3s-2} = \theta_{4r+3s}$ in (2.3.25). Apply (2.3.16) to both sides of (2.3.26) and collect the coefficients of

$$\sum \left\{ w^{(1)} \ldots w^{(2r+2s)} \mid uC_{2,2} + (r+s-u)C_{2,3} \right\} \quad ,$$

or, what amounts to the same, of

$$\sum_{ijk} m_{ij} m_{ik} \cdot \sum \left\{ w^{(1)} \ldots w^{(2r+2s-2)} \mid uC_{2,2} + (r+s-u-1)C_{2,3} \right\}$$

and apply (2.3.12):

$$\mathcal{A}\left(2rC_{1,2} + sC_{2,3} \; ; \; uC_{2,2} + (r+s-u)C_{2,3} \right) = \tag{2.3.27}$$

$$= \mathcal{A}\left(2rC_{1,2} + (s-1)C_{2,3} \; ; \; uC_{2,2} + (r+s-u-1)C_{2,3} \right) +$$

$$- \mathcal{A}\left(2rC_{1,2} + C_{2,2} + (s-1)C_{2,3} \; ; \; uC_{2,2} + (r+s-u)C_{2,3} \right) =$$

by (2.3.24)

$$= \mathcal{A}\left(2rC_{1,2} + (s-1)C_{2,3} \; ; \; uC_{2,2} + (r+s-u-1)C_{2,3} \right) +$$

$$- \mathcal{A}\left(2rC_{1,2} + (s-1)C_{2,3} \; ; \; (u-1)C_{2,2} + (r+s-u)C_{2,3} \right) \quad .$$

Introducing the abbreviation

$$f_r(s,u) = \mathcal{A}\left(2rC_{1,2} + sC_{2,3} \; ; \; uC_{2,2} + (r+s-u)C_{2,3} \right)$$

(2.3.27) can be written (with the obvious modifications for $u=0$ and $u=r+s$):

$$\left\{ \begin{array}{lll} f_r(s,u) & = f_r(s-1,u) - f_r(s-1,u-1) & \text{for } 0 < u < r+s \; , \\ f_r(s,0) & = f_r(s-1,0) \; , & \\ f_r(s,r+s) & = \qquad\qquad - f_r(s-1, r+s-1) \; . & \end{array} \right. \tag{2.3.28}$$

It follows from (2.3.28) that

$$\sum_{u=0}^{r+s} f_r(s,u) y^u = \sum_{u=0}^{r+s-1} f_r(s-1,u) y^u - \sum_{u=1}^{r+s} f_r(s-1,u-1) y^u =$$

$$= (1-y) \sum_{u=0}^{r+s-1} f_r(s-1,u) y^u \; .$$

Repeating this procedure gives

$$\sum_{u=0}^{r+s} f_r(s,u) y^u = (1-y)^s \sum_{u=0}^{r} f_r(0,u) y^u \quad . \tag{2.3.29}$$

Finally consider

$$\sum^* \left\{ w^{(1)} \ldots w^{(2r)} \mid 2rC_{1,2} \right\} \tag{2.3.30}$$

where the points of the connected components with configuration $C_{1,2}$ are numbered as before. Similar to (2.3.23) we find if the inequalities on θ_{2r-1} and θ_{2r} are removed

$$\begin{aligned}\sum^* \left\{ w^{(1)} \ldots w^{(2r)} \mid 2rC_{1,2} \right\} = & \sum_{ij} w_{ij} . \sum^* \left\{ w^{(1)} \ldots w^{(2r-2)} \mid (2r-1)C_{1,2} \right\} + \\ & - 2(2r-1) \sum^* \left\{ w^{(1)} \ldots w^{(2r)} \mid (2r-2)C_{1,2} + C_{2,2} \right\} + \\ & - 4(2r-1) \sum^* \left\{ w^{(1)} \ldots w^{(2r)} \mid (2r-2)C_{1,2} + C_{2,3} \right\} + \\ & - \text{sums corresponding to graphs with at least} \\ & \quad \text{one connected component with three or more joins.}\end{aligned}$$

Apply (2.3.16) to both sides of this equation and collect the coefficients of

$$\sum \left\{ w^{(1)} \ldots w^{(2r)} \mid uC_{2,2} + (r-u)C_{2,3} \right\} .$$

Thus for $0 < u < r$, using (2.3.12)

$$\begin{aligned} f_r(0,u) & \equiv \mathcal{A} \left(2rC_{1,2} \; ; \; uC_{2,2} + (r-u)C_{2,3} \right) = \\ & = - 2(2r-1) \; \mathcal{A} \left((2r-2)C_{1,2} + C_{2,2} \; ; \; uC_{2,2} + (r-u)C_{2,3} \right) + \\ & \quad - 4(2r-1) \; \mathcal{A} \left((2r-2)C_{1,2} + C_{2,3} \; ; \; uC_{2,2} + (r-u)C_{2,3} \right) = \\ & = - 2(2r-1) \; \mathcal{A} \left((2r-2)C_{1,2} \; ; \; (u-1)C_{2,2} + (r-u)C_{2,3} \right) + \\ & \quad - 4(2r-1) \; \mathcal{A} \left((2r-2)C_{1,2} + C_{2,3} \; ; \; uC_{2,2} + (r-u)C_{2,3} \right. = \\ & \equiv - 2(2r-1) f_{r-1}(0,u-1) \; - \; 4(2r-1) f_{r-1}(1,u) = \\ \text{by (2.3.28)} \quad & = + 2(2r-1) f_{r-1}(0,u-1) \; - \; 4(2r-1) f_{r-1}(0,u) \quad . \end{aligned}$$

For u=0 and r

$$\begin{aligned} f_r(0,0) & = \qquad\qquad - 4(2r-1) f_{r-1}(0,0) \; , \\ f_r(0,r) & = 2(2r-1) f_{r-1}(0,r-1) \; . \end{aligned}$$

Thus

$$\sum_{u=0}^{r} f_r(0,u) y^u = -2(2r-1)(2-y) \sum_{u=0}^{r-1} f_{r-1}(0,u) y^u \quad .$$

Repeating this procedure gives

$$\sum_{u=0}^{r} f_r(0,u)y^u = (-1)^{r-1}\, 2^{r-1}\, (2r-1)(2r-3)\ldots 3(2-y)^{r-1} \sum_{u=0}^{1} f_1(0,u)y^u .$$

Now $f_1(0,0)$ and $f_1(0,1)$ are found easily by applying (2.2.9) and (2.1.7) with k=2

$$\left(\sum w_{ij}\right)^2 = \sum_{(ijkl)\neq} w_{ij}w_{kl} + \mathcal{N}(C_{2,3}) \sum_{(ijk)\neq} w_{ij}w_{ik} + \mathcal{N}(C_{2,2}) \sum_{(ij)\neq} w_{ij}^2 =$$

$$= \sum_{(ijkl)\neq} w_{ij}w_{kl} + 4 \sum_{(ijk)\neq} w_{ij}w_{ik} + 2 \sum_{(ij)\neq} w_{ij}^2 .$$

Now

$$\sum_{(ij)\neq} w_{ij}^2 = \sum_{ij} w_{ij}^2$$

$$\sum_{(ijk)\neq} w_{ij}w_{ik} = \sum_{ijk} w_{ij}w_{ik} - \sum_{ij} w_{ij}^2$$

thus

$$\sum_{(ijkl)\neq} w_{ij}w_{kl} = \left(\sum_{ij} w_{ij}\right)^2 - 4 \sum_{ijk} w_{ij}w_{ik} + 2 \sum_{ij} w_{ij}^2 .$$

Comparing with (2.3.8) we find

$$f_1(0,0) = \mathcal{A}(2C_{1,2}\ ;\ C_{2,3}) = -4$$
$$f_1(0,1) = \mathcal{A}(2C_{1,2}\ ;\ C_{2,2}) = +2 .$$

Thus

(2.3.31)
$$\sum_{u=0}^{r} f_r(0,u)y^u = (-1)^r .\ 2^r(2r-1)(2r-3)\ldots 3.1\ (2-y)^r =$$
$$= (-1)^r \frac{(2r)!}{r!} (2-y)^r$$

The lemma now follows from (2.3.24), (2.3.29) and (2.3.31).

<u>Lemma 2.3.4</u>

If for $i,j=1,2,\ldots,n$, $w_{ij} \geq 0$ and $\sum_j w_{ij} \leq c$, where c is a constant not depending on i and n, then

$$\sum \left\{ w^{(1)} \ldots w^{(k)} \;\middle|\; \sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)} \right\} = \mathcal{O}(n^h) .$$

<u>Proof</u>.

As

$$\sum \left\{ w^{(1)} \ldots w^{(k)} \;\middle|\; \sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)} \right\} = \prod_{i=1}^{h} \sum \left\{ w^{(k_1+\ldots+k_{i-1}+1)} \ldots w^{(k_1+\ldots+k_i)} \;\middle|\; C_{k_i,l_i}^{(\alpha_i)} \right\}$$

the lemma is proved if it is shown that

$$\sum \left\{ w^{(1)} \ldots w^{(k)} \mid C_{k,l}^{(\alpha)} \right\} = \mathcal{O}(n) \tag{2.3.32}$$

for every k, l and α .

For k=1 and l=2 (2.3.32) is true, because

$$\sum \left\{ w^{(1)} \mid C_{1,2} \right\} = \sum_{ij} w_{ij} \leq cn \quad .$$

For $k > 1$ consider first the case $l=k+1$. Now $C_{k,k+1}^{(\alpha)}$ is a configuration of a graph containing no circuits; it is a tree (cf. D.KOENIG (1936), 53). Each tree contains at least two points to which only one join is connected (cf. D.KOENIG (1936), 49). Let one of these points be labelled k+1. This means that the summation-subscript Θ_{k+1} occurs only once as a subscript to the w's. Summing over Θ_{k+1} gives

$$\sum \left\{ w^{(1)} \ldots w^{(k)} \mid C_{k,k+1}^{(\alpha)} \right\} \leq c \sum \left\{ w^{(1)} \ldots w^{(k-1)} \mid C_{k-1,k}^{(\beta)} \right\} ,$$

where the graph with configuration $C_{k-1,k}^{(\beta)}$, is obtained from $C_{k,k+1}^{(\alpha)}$ by taking the point labelled k+1 away together with the join connected to it. The remaining graph is again a tree. Repeating this procedure leads in k-1 steps to

$$\sum \left\{ w^{(1)} \ldots w^{(k)} \mid C_{k,k+1}^{(\alpha)} \right\} \leq c^{k-1} \sum \left\{ w^{(1)} \mid C_{1,2} \right\} \leq c^{k} n , \tag{2.3.33}$$

which proves (2.3.32) for l=k+1.

Now consider a graph with configuration $C_{k,l}^{(\alpha)}$, with $2 \leq l < k+1$. In this graph one can take k-l+1 joins away, such that a tree $C_{l-1,l}^{(\beta)}$ remains with l-1 joins and l points (D.KOENIG (1936), 53). Let the joins that can be taken away correspond to $w^{(l)},\ldots,w^{(k)}$, then

$$w^{(1)} \ldots w^{(k)} \leq c^{k-l+1} \, w^{(1)} \ldots w^{(l-1)} \quad .$$

Summing over the l summation-subscripts involved gives

$$\sum \left\{ w^{(1)} \ldots w^{(k)} \mid C_{k,l}^{(\alpha)} \right\} \leq c^{k-l+1} \sum \left\{ w^{(1)} \ldots w^{(l-1)} \mid C_{l-1,l}^{(\beta)} \right\} ,$$

which by (2.3.33) proves (2.3.32).

CHAPTER 3

3.1. The moments of $\underline{z}$

In section 2.2 the starting-point has been the set of numbers w_{ij}. However, the development given in section 2.2 does not change if the w_{ij} are random variables, thus $\underline{w}_{ij}$. Only (2.2.1) should be replaced by

(3.1.1) $$\underline{w}_{ii} = 0 \quad , \text{ spr } 0 \quad .$$

Thus analogous to (2.2.3) we have

(3.1.2) $$\Big(\sum_{ij} \underline{w}_{ij}\Big)^k = \sum_{\tau_1=1}^{n} \cdots \sum_{\tau_{2k}=1}^{n} \underline{w}_{\tau_1,\tau_2} \cdots \underline{w}_{\tau_{2k-1},\tau_{2k}} \quad .$$

Now take

(3.1.3) $$\underline{w}_{ij} \equiv m_{ij}\, \underline{z}_{ij} \qquad (i,j=1,2,..,n) \quad ,$$

where $\underline{z}_{ij}$ are the random variables introduced in section 1, then (cf. 1.1.8)

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

Consider one term of (3.1.2)

(3.1.4) $$\underline{w}_{\tau_1,\tau_2}\, \underline{w}_{\tau_3,\tau_4} \cdots \underline{w}_{\tau_{2k-1},\tau_{2k}}$$

which by (3.1.3) is equal to

(3.1.4') $$m_{\tau_1,\tau_2} \cdots m_{\tau_{2k-1},\tau_{2k}} \cdot \underline{z}_{\tau_1,\tau_2} \cdots \underline{z}_{\tau_{2k-1},\tau_{2k}} \quad ,$$

where we assume $\tau_{2j-1} \neq \tau_{2j}$ $(j=1,2,\ldots,k)$. Among the $2k$ subscripts $\tau_1,\tau_2,\ldots,\tau_{2k}$ l unequal numbers from $1,2,\ldots,l$ occur $(2 \leq l \leq 2k)$. Call these $\lambda_1,\lambda_2,\ldots,\lambda_l$ with

$$\lambda_1 < \lambda_2 < \ldots < \lambda_l \quad .$$

Each of the τ's is equal to one of the λ's. Let

$$\tau_{2j-1} = \lambda_{\mu_j} \quad \text{and} \quad \tau_{2j} = \lambda_{\nu_j} \qquad (j=1,2,\ldots,k)$$

then (3.1.4') is equal to

(3.1.5) $$m_{\lambda_{\mu_1},\lambda_{\nu_1}} \cdots m_{\lambda_{\mu_k},\lambda_{\nu_k}} \cdot \underline{z}_{\lambda_{\mu_1},\lambda_{\nu_1}} \cdots \underline{z}_{\lambda_{\mu_k},\lambda_{\nu_k}} \quad .$$

The product

$$m_{\lambda_{\mu_1},\lambda_{\nu_1}} \cdots m_{\lambda_{\mu_k},\lambda_{\nu_k}}$$

will be called the m-part of (3.1.5),
the product

$$\underline{z}_{\lambda_{\mu_1},\lambda_{\nu_1}} \cdots \underline{z}_{\lambda_{\mu_k},\lambda_{\nu_k}}$$

the z-part.

To the set of numbers $\mu_1, \mu_2, \ldots, \mu_k,\ \nu_1, \nu_2, \ldots, \nu_k$ - and therefore to the m-part and the z-part of (3.1.5) - there corresponds a graph, as indicated in section 2.2.

As shown by examples 1,2 and 3 and 4,5 and 6 of Table 2.2.1 (replace w by m), the m-parts of two terms of (3.1.2) corresponding to equivalent graphs may or may not have the same value. Also the m-parts of two terms of (3.1.2) corresponding to distinct graphs, but with the same configuration, may or may not have the same value.

Regarding the z-part of the terms of (3.1.2) we introduce the following

symmetry assumption:

> For k=1,2,..., the expectation of the z-part of a term of (3.1.2) depends only on the configuration of the graph which corresponds to that term.

Examples (cf. Table 2.2.1):

1) $E\underline{z}_{12} = E\underline{z}_{57}$, as both $\underline{z}_{12}$ and $\underline{z}_{57}$ have a graph with configuration .

2) $E\underline{z}_{12}^2 = E\underline{z}_{35}^2$, as both $\underline{z}_{12}^2 \equiv \underline{z}_{12}\,\underline{z}_{12}$ and $\underline{z}_{35}^2 \equiv \underline{z}_{35}\,\underline{z}_{35}$ have a graph with configuration .

It will be shown in later sections that in fact $\underline{x}_{ij}^{(B)}$, $\underline{x}_{ij}^{(W)}$, $\underline{y}_{ij}$ and $\underline{v}_{ij}$ satisfy this symmetry assumption.

If the configuration of the graph corresponding to (3.1.5) is $\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)}$, the expectation of the z-part will be written symbolically as

$$E\left(\underline{z}^{(1)} \ldots \underline{z}^{(k)} \mid \sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)}\right) .$$

The contribution to $\underline{E}\underline{z}^k = E\left(\sum_{ij}\underline{w}_{ij}\right)^k$ of terms corresponding to a given value of l, and to a given configuration $\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)}$, $(\sum_{i=1}^{h} k_i = k, \sum_{i=1}^{h} l_i = l)$, is now by (2.2.7), (2.2.8) and the symmetry assumption equal to

$$\mathcal{N}\left(\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)}\right) \sum_{\substack{\theta_1=1 \\ (\theta_1,\ldots,\theta_l)\neq}}^{n} \cdots \sum_{\theta_l=1}^{n} E\, \underline{w}_{\theta_{\mu_1},\theta_{\nu_1}} \cdots \underline{w}_{\theta_{\mu_k},\theta_{\nu_k}} =$$

$$= \mathcal{N}\left(\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)}\right) \sum_{\substack{\theta_1=1 \\ (\theta_1,\ldots,\theta_l)\neq}}^{n} \cdots \sum_{\theta_l=1}^{n} m_{\theta_{\mu_1},\theta_{\nu_1}} \cdots m_{\theta_{\mu_k},\theta_{\nu_k}}\, E\, \underline{z}_{\theta_{\mu_1},\theta_{\nu_1}} \cdots \underline{z}_{\theta_{\mu_k},\theta_{\nu_k}} =$$

$$= \mathcal{N}\left(\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)}\right) E\left(\underline{z}^{(1)} \cdots \underline{z}^{(k)} \mid \sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)}\right) \sum_{\substack{\theta_1=1 \\ (\theta_1,\ldots,\theta_l)\neq}}^{n} \cdots \sum_{\theta_l=1}^{n} m_{\theta_{\mu_1},\theta_{\nu_1}} \cdots m_{\theta_{\mu_k},\theta_{\nu_k}} =$$

$$= \mathcal{N}\left(\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)}\right) E\left(\underline{z}^{(1)} \cdots \underline{z}^{(k)} \mid \sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)}\right) \Sigma^{*}\left\{ m^{(1)} \cdots m^{(k)} \mid \sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)} \right\},$$

where the graph corresponding to the numbers $\mu_1,\mu_2,\ldots,\mu_k$, $\nu_1,\nu_2,\ldots,\nu_k$ is one of the graphs having the configuration $\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)}$.

Summing over all configurations with k joins gives

$$(3.1.6) \qquad \underline{E}\underline{z}^k = \sum_{l=2}^{2k} \sum_{h=1}^{\left[\frac{l}{2}\right]} \Sigma''_{\left\lfloor \sum_{i=1}^{h} k_i=k,\ \sum_{i=1}^{h} l_i=l \right\rfloor} \mathcal{N}\left(\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)}\right) \cdot$$

$$\cdot\, E\left(\underline{z}^{(1)} \cdots \underline{z}^{(k)} \mid \sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)}\right) \Sigma^{*}\left\{ m^{(1)} \cdots m^{(k)} \mid \sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)} \right\}.$$

Applying (2.3.8) and defining for a configuration $\sum_{i=1}^{h'} C_{k_i',l_i'}^{(\alpha_i')}$, with $\sum_{i=1}^{h'} l_i' = l'$, $\sum_{i=1}^{h'} k_i' = k$ and $l \geq l'$

$$(3.1.7)\quad \mathscr{D}^{(k)}_{l,l'}\left(\sum_{i=1}^{h'} C^{(\alpha_i')}_{k_i',l_i'}\right) = \sum_{h=h'}^{\left[\frac{l}{2}\right]} \sum{}'' \underset{\sum_{i=1}^{h} k_i=k,\ \sum_{i=1}^{h} l_i=l}{} \mathscr{N}\left(\sum_{i=1}^{h} C^{(\alpha_i)}_{k_i,l_i}\right) \cdot$$

$$\cdot\ E\left(\underline{z}^{(1)}\ldots\underline{z}^{(k)}\mid \sum_{i=1}^{h} C^{(\alpha_i)}_{k_i,l_i}\right) \mathscr{A}\left(\sum_{i=1}^{h} C^{(\alpha_i)}_{k_i,l_i};\ \sum_{i=1}^{h'} C^{(\alpha_i')}_{k_i',l_i'}\right),$$

and for $l < l'$ (cf. (2.3.9) and (2.3.10))

$$(3.1.8)\quad \mathscr{D}^{(k)}_{l,l'}\left(\sum_{i=1}^{h'} C^{(\alpha_i')}_{k_i',l_i'}\right) = 0 ,$$

it follows from (3.1.6) that

$$(3.1.9)\quad E\underline{z}^{k} = \sum_{h'=1}^{k} \sum_{l'=2h'}^{2k} \sum{}'' \underset{\sum_{i=1}^{h'} k_i'=k,\ \sum_{i=1}^{h'} l_i'=l'}{} \sum_{l=2}^{2k} \mathscr{D}^{(k)}_{l,l'}\left(\sum_{i=1}^{h'} C^{(\alpha_i')}_{k_i',l_i'}\right) \cdot$$

$$\cdot \sum \left\{ m^{(1)}\ldots m^{(k)} \mid \sum_{i=1}^{h'} C^{(\alpha_i')}_{k_i',l_i'}\right\}.$$

3.2. The reduced moments of $\underline{z}$

From (3.1.9) it follows that

$$E\underline{z}^{k} = \sum_{g'=0}^{k-2} \sum_{h'=1}^{\left[\frac{k-g'}{2}\right]} \sum_{l'=2h'+2g'}^{k+g'+h'} \sum{}'' \underset{\sum_{i=1}^{h} k_i'=k-g',\ k_i'\geq 2,\ \sum_{i=1}^{h'} l_i'=l'-2g'}{} \cdot$$

$$\cdot \sum_{l=2}^{2k} \mathscr{D}^{(k)}_{l,l'}\left(\sum_{i=1}^{h'} C^{(\alpha_i')}_{k_i',l_i'} + g'C_{1,2}\right) \sum \left\{ m^{(1)}\ldots m^{(k)} \mid \sum_{i=1}^{h'} C^{(\alpha_i')}_{k_i',l_i'} + g'C_{1,2}\right\} +$$

$$+\ \mathscr{D}^{(k)}_{2k,2k}\left(kC_{1,2}\right) \sum\left\{ m^{(1)}\ldots m^{(k)} \mid kC_{1,2}\right\}.$$

Thus

$$\mu_k \equiv E(\underline{z}-E\underline{z})^k = \sum_{\nu=0}^{k} (-1)^{\nu}\binom{k}{\nu} (E\underline{z})^{\nu} E\underline{z}^{(k-\nu)} =$$

$$= \sum_{\nu=0}^{k-2} (-1)^{\nu} \binom{k}{\nu} \left\{ E\left(\underline{z}^{(1)} \mid C_{1,2}\right)\right\}^{\nu} \sum \left\{ m^{(1)}\ldots m^{(\nu)} \mid \nu C_{1,2}\right\} \sum_{g'=0}^{k-\nu-2} \sum_{h'=1}^{\left[\frac{k-\nu-g'}{2}\right]}$$

$$\sum_{l'=2h'+2g'}^{k-\nu+g'+h'} \sum{}'' \underset{\sum_{i=1}^{h'} k_i'=k-\nu-g',\ k_i'\geq 2,\ \sum_{i=1}^{h'} l_i'=l'-2g'}{} \sum_{l=2}^{2k-2\nu} \mathscr{D}^{(k-\nu)}_{l,l'}\left(\sum_{i=1}^{h'} C^{(\alpha_i')}_{k_i',l_i'} + g'C_{1,2}\right).$$

$$. \sum \left\{ m^{(1)} \ldots m^{(k-\nu)} \mid \sum_{i=1}^{h'} C_{k_i',l_i'}^{(\alpha_i')} + g'C_{1,2} \right\} +$$

$$+ \sum_{\nu=0}^{k} (-1)^\nu \binom{k}{\nu} \left\{ E\left(\underline{z}^{(1)} \mid C_{1,2}\right)\right\}^\nu \sum \left\{ m^{(1)}\ldots m^{(\nu)} \mid \nu C_{1,2}\right\} .$$

$$. \mathcal{D}^{(k-\nu)}_{2k-2\nu,2k-2\nu}\left((k-\nu)C_{1,2}\right) \sum\left\{m^{(1)}\ldots m^{(k-\nu)} \mid (k-\nu)C_{1,2}\right\} .$$

In the first part of the righthand member change

g' into $g'^{*}-\nu$, and g'^{*} into g' ,

then l' into $l'^{*}-2\nu$, and l'^{*} into l', and

finally l into $l^{*}-2\nu$ and l^{*} into l, and interchange the order of summation over ν and g'. Then:

(3.2.1) $$\mu_k = \sum_{g'=0}^{k-2}\sum_{\nu=0}^{g'} (-1)^\nu \binom{k}{\nu} \left\{E\left(\underline{z}^{(1)} \mid C_{1,2}\right)\right\}^\nu \sum_{h'=1}^{\left[\frac{k-g'}{2}\right]} \sum_{l'=2h'+2g'}^{k+g'+h'} .$$

$$. \sum{}''_{\left\lfloor \sum_{i=1}^{h'} k_i'=k-g';\,k_i'\geq 2,\ \sum_{i=1}^{h'} l_i'=l'-2g' \right\rfloor} \sum_{l=2\nu+2}^{2k} \mathcal{D}^{(k-\nu)}_{l-2\nu,l'-2\nu}\left(\sum_{i=1}^{h'} C^{(\alpha_i')}_{k_i',l_i'} + (g'-\nu)C_{1,2}\right).$$

$$. \sum\left\{m^{(1)}\ldots m^{(k)} \mid \sum_{i=1}^{h'} C^{(\alpha_i')}_{k_i',l_i'} + g'C_{1,2}\right\} + \sum_{\nu=0}^{k}(-1)^\nu\binom{k}{\nu}\left\{E\left(\underline{z}^{(1)}\mid C_{1,2}\right)\right\}^\nu .$$

$$. \mathcal{D}^{(k-\nu)}_{2k-2\nu,2k-2\nu}\left((k-\nu)C_{1,2}\right) . \sum\left\{m^{(1)}\ldots m^{(k)} \mid kC_{1,2}\right\}.$$

For $l \geq l'$

(3.2.2) $$\mathcal{D}^{(k-\nu)}_{l-2\nu,l'-2\nu}\left(\sum_{i=1}^{h'} C^{(\alpha_i')}_{k_i',l_i'} + (g'-\nu)C_{1,2}\right) =$$

$$= \sum_{h=h'+g'-\nu}^{\left[\frac{l}{2}\right]-\nu} \sum{}''_{\left\lfloor \sum_{i=1}^h k_i = k-\nu,\ \sum_{i=1}^h l_i = l-2\nu\right\rfloor} \mathcal{N}\left(\sum_{i=1}^h C^{(\alpha_i)}_{k_i,l_i}\right) E\left(\underline{z}^{(1)}\ldots\underline{z}^{(k-\nu)} \mid \sum_{i=1}^h C^{(\alpha_i)}_{k_i,l_i}\right).$$

$$.\mathcal{A}\left(\sum_{i=1}^h C^{(\alpha_i)}_{k_i,l_i} ; \sum_{i=1}^{h'} C^{(\alpha_i')}_{k_i',l_i'} + (g'-\nu)C_{1,2}\right) =$$

$$= \sum_{g=g'}^{\left[\frac{l}{2}\right]} \sum_{h=h'-\left[\frac{g-g'}{2}\right]}^{\left[\frac{k-g}{2}\right]} \sum{}''_{\left\lfloor \sum_{i=1}^h k_i=k-g,\, k_i\geq 2,\ \sum_{i=1}^h l_i = l-2g\right\rfloor} \mathcal{N}\left(\sum_{i=1}^h C^{(\alpha_i)}_{k_i,l_i} + (g-\nu)C_{1,2}\right).$$

$$. E\left(\underline{z}^{(1)}\ldots\underline{z}^{(k-\nu)} \mid \sum_{i=1}^h C^{(\alpha_i)}_{k_i,l_i} + (g-\nu)C_{1,2}\right).$$

$$\cdot \mathcal{A}\left(\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)} + (g-\nu)C_{1,2} \; ; \; \sum_{i=1}^{h'} C_{k'_i,l'_i}^{(\alpha'_i)} + (g'-\nu)C_{1,2}\right) .$$

The lower summation-limit for g is determined by the fact that for $g < g'$ the coefficient $\mathcal{A}$ is equal to zero by lemma 2.3.2. The lower summation-limit for h is determined as follows. We recall from section 2.3 that the coefficient $\mathcal{A} \neq 0$ only if the configuration $\sum_{i=1}^{h'} C_{k'_i,l'_i}^{(\alpha'_i)} + (g'-\nu)C_{1,2}$ can be obtained from $\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)} + (g-\nu)C_{1,2}$ by making points coincide. In this process the total number of components must have decreased by at least $\left[\frac{g-g'+1}{2}\right]$,

so $h'+g'-\nu \leq h+g-\nu-\left[\frac{g-g'+1}{2}\right]$, or $h \geq h'+\left[\frac{g-g'+1}{2}\right]-(g-g') = h'-\left[\frac{g-g'}{2}\right]$.

By theorem 2.1.1, because of $k_i \geq 2$ and $\sum_{i=1}^{h} k_i = k-g$,

$$\mathcal{N}\left(\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)} + (g-\nu)C_{1,2}\right) = \frac{(k-\nu)!}{(g-\nu)!(k-g)!}\mathcal{N}\left(\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)}\right) ,$$

$$\mathcal{N}\left(\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)} + (g-g')C_{1,2}\right) = \frac{(k-g')!}{(g-g')!(k-g)!}\mathcal{N}\left(\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)}\right) ,$$

thus

$$\mathcal{N}\left(\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)} + (g-\nu)C_{1,2}\right) = \frac{(k-\nu)!(g-g')!}{(k-g')!(g-\nu)!}\mathcal{N}\left(\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)} + (g-g')C_{1,2}\right) .$$

Applying this and lemma 2.3.2 to (3.2.2), changing g into g^*+g', and g^* into g, gives

$$\mathcal{D}_{1-2\nu,l'-2\nu}^{(k-\nu)}\left(\sum_{i=1}^{h'} C_{k'_i,l'_i}^{(\alpha'_i)} + (g'-\nu)C_{1,2}\right) = \sum_{g=0}^{\left[\frac{l}{2}\right]-g'} \sum_{h=h'-\left[\frac{g}{2}\right]}^{\left[\frac{k-g-g'}{2}\right]} \binom{k-\nu}{k-g'} \cdot$$

$$\cdot \sum{}''_{\left\lfloor \sum_{i=1}^{h} k_i=k-g-g',\, k_i \geq 2,\, \sum_{i=1}^{h} l_i=l-2g-2g' \right\rfloor} \mathcal{N}\left(\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)} + gC_{1,2}\right) .$$

$$\cdot \mathcal{A}\left(\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)} + gC_{1,2} ; \sum_{i=1}^{h'} C_{k'_i,l'_i}^{(\alpha'_i)}\right) . \; E\left(\underline{z}^{(1)}\ldots\underline{z}^{(k-\nu)} \,\middle|\, \sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)} + (g+g'-\nu)C_{1,2}\right) .$$

In the same way one finds

$$\mathcal{D}_{2k-2\nu,2k-2\nu}^{(k-\nu)}\left((k-\nu)C_{1,2}\right) = E\left(\underline{z}^{(1)}\ldots\underline{z}^{(k-\nu)} \mid (k-\nu)C_{1,2}\right) .$$

By applying these results to (3.2.1), it follows after some rearrangement that

$$(3.2.3)\quad \mu_k = \sum_{g'=0}^{k-2} \sum_{h'=1}^{\left[\frac{k-g'}{2}\right]} \sum_{l'=2h'+2g'}^{k+g'+h'} \sum''_{\left\lfloor \sum_{i=1}^{h'} k_i'=k-g',\, k_i' \geq 2,\, \sum_{i=1}^{h'} l_i'=l'-2g' \right\rfloor} \binom{k}{g'} .$$

$$. \sum \left\{ m^{(1)} \ldots m^{(k)} \mid \sum_{i=1}^{h'} C_{k_i',l_i'}^{(\alpha_i')} + g'C_{1,2} \right\} \sum_{g=0}^{k-g'} \sum_{h=h'-\left[\frac{g}{2}\right]}^{\left[\frac{k-g-g'}{2}\right]} \sum_{l=2h+2g+2g'}^{2k}$$

$$\sum''_{\left\lfloor \sum_{i=1}^{h} k_i=k-g-g',\, k_i \geq 2,\, \sum_{i=1}^{h} l_i=l-2g-2g' \right\rfloor} \mathcal{N}\left(\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)} + gC_{1,2} \right) .$$

$$. \mathcal{A}\left(\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)} + gC_{1,2} \,;\, \sum_{i=1}^{h'} C_{k_i',l_i'}^{(\alpha_i')} \right) . \sum_{\nu=0}^{g'} (-1)^{\nu} \binom{g'}{\nu} . \left\{ E\left(\underline{z}^{(1)} \mid C_{1,2}\right)\right\}^{\nu} .$$

$$. E\left(\underline{z}^{(1)} \ldots \underline{z}^{(k-\nu)} \mid \sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)} + (g'+g-\nu)C_{1,2} \right) +$$

$$+ \sum\left\{ m^{(1)} \ldots m^{(k)} \mid kC_{1,2} \right\} \sum_{\nu=0}^{k} (-1)^{\nu} \binom{k}{\nu} \left\{ E\left(\underline{z}^{(1)} \mid C_{1,2}\right)\right\}^{\nu} E\left(\underline{z}^{(1)} \ldots \underline{z}^{(k-\nu)} \mid (k-\nu)C_{1,2} \right) .$$

This expression for μ_k will prove to be useful in studying the asymptotic behaviour of the distribution of $\underline{x}_B$, $\underline{x}_W$ and $\underline{y}$.

3.3. The moments of $\underline{x}_B$ and $\underline{x}_W$ - non free sampling

It is obvious that the moments of $\underline{x}_W$ can be obtained from those of $\underline{x}_B$ by changing r_1 into r_2. It is therefore sufficient to consider the moments of $\underline{x}_B$. In this section we write $\underline{x}$ for $\underline{x}_B$ and r for r_1.

For $i \neq j$

$$E\underline{x}_{ij} = P\left[\underline{x}_{ij}=1\right] = P\left[\text{points } i \text{ and } j \text{ are chosen}\right] =$$

$$= P\left[\text{point } i \text{ is chosen}\right] . P\left[\text{point } j \text{ is chosen} \mid i \text{ is chosen}\right] =$$

$$= \frac{r}{n} . \frac{r-1}{n-1} = \frac{r^{!2}}{n^{!2}} .$$

Thus

$$(3.3.1)\quad E\underline{x} = E \sum_{ij} m_{ij}\underline{x}_{ij} = \sum_{i \neq j} m_{ij}E\underline{x}_{ij} = \frac{r^{!2}}{n^{!2}} \sum_{i \neq j} m_{ij} = \frac{r^{!2}}{n^{!2}} \sum_{ij} m_{ij} .$$

In general for a configuration $\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)}$ with $\sum_{i=1}^{h} k_i = k$, $\sum_{i=1}^{h} l_i = l$,

$$(3.3.2) \qquad E\left(\underline{x}^{(1)} \ldots \underline{x}^{(k)} \mid \sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)}\right) = P\left[l \text{ given points are chosen}\right] = \frac{r(r-1)\ldots(r-l+1)}{n(n-1)\ldots(n-l+1)} = \frac{r^{!l}}{n^{!l}} \quad ,$$

which shows that the symmetry assumption introduced in section 3.1 holds for $\underline{x}$. In this chapter we assume that r always satisfies $r \geq l$.

$E\underline{x}^2$ is obtained by applying (3.1.6). There are 3 configurations involved for k=2.

1)  $= 2C_{1,2}$

By (3.3.2) $E\left(\underline{x}^{(1)}\ \underline{x}^{(2)} \mid 2C_{1,2}\right) = \frac{r^{!4}}{n^{!4}}$; by (2.1.7) $\mathcal{N}(2C_{1,2}) = 1$,

thus giving a contribution to $E\underline{x}^2$ of

$$\frac{r^{!4}}{n^{!4}} \sum{}^{*}\left\{ m^{(1)}\ m^{(2)} \mid 2C_{1,2} \right\} = \frac{r^{!4}}{n^{!4}} \sum_{(ijkl)\neq} m_{ij} m_{kl} \quad .$$

2) $= C_{2,3}$

By (3.3.2) $E\left(\underline{x}^{(1)}\ \underline{x}^{(2)} \mid C_{2,3}\right) = \frac{r^{!3}}{n^{!3}}$; by (2.1.7) $\mathcal{N}(C_{2,3}) = 4$,

thus giving a contribution to $E\underline{x}^2$ of

$$4\ \frac{r^{!3}}{n^{!3}} \sum{}^{*}\left\{ m^{(1)}\ m^{(2)} \mid C_{2,3} \right\} = 4\ \frac{r^{!3}}{n^{!3}} \sum_{(ijk)\neq} m_{ij} m_{ik} \quad .$$

3)  $= C_{2,2}$

By (3.3.2) $E\left(\underline{x}^{(1)}\ \underline{x}^{(2)} \mid C_{2,2}\right) = \frac{r^{!2}}{n^{!2}}$; by (2.1.7) $\mathcal{N}(C_{2,2}) = 2$,

thus giving a contribution to $E\underline{x}^2$ of

$$2\ \frac{r^{!2}}{n^{!2}} \sum{}^{*}\left\{ m^{(1)}\ m^{(2)} \mid C_{2,2} \right\} = 2\ \frac{r^{!2}}{n^{!2}} \sum_{(ij)\neq} m_{ij}^2 \quad .$$

Therefore

$$(3.3.3) \qquad E\underline{x}^2 = \frac{r^{!4}}{n^{!4}} \sum_{(ijkl)\neq} m_{ij} m_{kl} + 4\ \frac{r^{!3}}{n^{!3}} \sum_{(ijk)\neq} m_{ij} m_{ik} + 2\ \frac{r^{!2}}{n^{!2}} \sum_{(ij)\neq} m_{ij}^2 \quad .$$

In order to illustrate the use of (3.1.9), we calculate first by means of lemma 2.3.3.

$$\mathcal{A}(C_{2,2}\ ;\ C_{2,2}) = 1$$

$$\mathcal{A}(C_{2,3}\ ;\ C_{2,2}) = -1 \qquad \mathcal{A}(C_{2,3}\ ;\ C_{2,3}) = 1$$

$$\mathcal{A}(2C_{1,2}\ ;\ C_{2,2}) = +2 \qquad \mathcal{A}(2C_{1,2}\ ;\ C_{2,3}) = -4 \quad ,$$

while by lemma 2.3.2 $\mathcal{A}(2C_{1,2}\ ;\ 2C_{1,2}) = 1$.

Thus

$$\mathcal{D}^{(2)}_{2,2}(C_{2,2}) = \mathcal{N}(C_{2,2})\, E\left(\underline{z}^{(1)}\, \underline{z}^{(2)} \mid C_{2,2}\right)\mathcal{A}(C_{2,2}\ ;\ C_{2,2}) = 2\,\frac{r^{!2}}{n^{!2}}$$

$$\mathcal{D}^{(2)}_{3,2}(C_{2,2}) = -\,4\,\frac{r^{!3}}{n^{!3}} \quad ,$$

$$\mathcal{D}^{(2)}_{4,2}(C_{2,2}) = 2\,\frac{r^{!4}}{n^{!4}} \quad ,$$

$$\mathcal{D}^{(2)}_{3,3}(C_{2,3}) = 4\,\frac{r^{!3}}{n^{!3}} \quad ,$$

$$\mathcal{D}^{(2)}_{4,3}(C_{2,3}) = -\,4\,\frac{r^{!4}}{n^{!4}} \quad ,$$

$$\mathcal{D}^{(2)}_{4,4}(2C_{1,2}) = \frac{r^{!4}}{n^{!4}} \quad ,$$

and all other $\mathcal{D}^{(2)}_{\dots,\dots}(\dots)$ are equal to zero.

Finally apply (3.1.9):

$$(3.3.4)\qquad E\underline{x}^2 = \frac{r^{!4}}{n^{!4}}\, m_{++}^2 + 4\left(\frac{r^{!3}}{n^{!3}} - \frac{r^{!4}}{n^{!4}}\right)\sum_i m_{i+}^2 + 2\left(\frac{r^{!2}}{n^{!2}} - 2\,\frac{r^{!3}}{n^{!3}} + \frac{r^{!4}}{n^{!4}}\right)\sum_{ij} m_{ij}^2 \ .$$

This result could have been obtained from (3.3.3) by simple reasoning. From (2.2.9) it follows for k=2

$$(3.3.5)\qquad m_{++}^2 = \left(\sum_{ij} m_{ij}\right)^2 = \sum_{(ijkl)\neq} m_{ij}m_{kl} + 4\sum_{(ijk)\neq} m_{ij}m_{ik} + 2\sum_{(ij)\neq} m_{ij}^2$$

while

$$m_{i+}^2 = \left(\sum_j m_{ij}\right)^2 = \sum_{(jk)\neq} m_{ij}m_{ik} + \sum_j m_{ij}^2 \quad .$$

Summing over i gives

$$\sum_i m_{i+}^2 = \sum_i \sum_{(jk)\neq} m_{ij}m_{ik} + \sum_{ij} m_{ij}^2 \quad ,$$

and because $m_{ii} = 0$

(3.3.6) $$\sum_i m_{i+}^2 = \sum_{(ijk)\neq} m_{ij}m_{ik} + \sum_{(ij)\neq} m_{ij}^2 \quad .$$

Also

(3.3.7) $$\sum_{(ij)\neq} m_{ij}^2 = \sum_{ij} m_{ij}^2 \quad .$$

Applying (3.3.5), (3.3.6) and (3.3.7) to (3.3.3) leads directly to (3.3.4).

To derive an expression of the variance of $\underline{x}$, one could use (3.2.3); in this case it is easier to obtain σ^2 directly from (3.3.4) and (3.3.1). Easy algebra gives

(3.3.8) $$\sigma^2 = \frac{4r^{!3}(n-r)}{n^{!4}} \sum_i \left(m_{i+} - \frac{m_{++}}{n}\right)^2 + 2\,\frac{r^{!2}(n-r)^{!2}}{n^{!2}n^{!4}} \left(n^{!2}\sum_{ij} m_{ij}^2 - m_{++}^2\right) .$$

If for all i m_{i+} is a constant, the first term of σ^2 is equal to zero.

From (3.1.6), (3.3.2) and (2.1.8) the third moment is easily found to be

(3.3.9) $$E\underline{x}^3 = \frac{r^{!6}}{n^{!6}} \sum_{(ijkluv)\neq} m_{ij}m_{kl}m_{uv} + 12\,\frac{r^{!5}}{n^{!5}} \sum_{(ijklu)\neq} m_{ij}m_{ik}m_{lu} +$$

$$+ 6\,\frac{r^{!4}}{n^{!4}} \sum_{(ijkl)\neq} m_{ij}^2 m_{kl} + 8\,\frac{r^{!4}}{n^{!4}} \sum_{(ijkl)\neq} m_{ij}m_{ik}m_{il} + 24\,\frac{r^{!4}}{n^{!4}} \sum_{(ijkl)\neq} m_{ij}m_{ik}m_{jl} +$$

$$+ 8\,\frac{r^{!3}}{n^{!3}} \sum_{(ijk)\neq} m_{ij}m_{ik}m_{jk} + 24\,\frac{r^{!3}}{n^{!3}} \sum_{(ijk)\neq} m_{ij}^2 m_{ik} + 4 \sum_{(ij)\neq} m_{ij}^3 \quad .$$

Removing the inequalities on the summations one obtains the result tabled in the left-hand part of table 3.3.1.

Table 3.3.1 : Two expressions for $E\underline{x}^3$.

	$\frac{r^{!2}}{n^{!2}}$	$\frac{r^{!3}}{n^{!3}}$	$\frac{r^{!4}}{n^{!4}}$	$\frac{r^{!5}}{n^{!5}}$	$\frac{r^{!6}}{n^{!6}}$		$\frac{r^{!2}(n-r)^{!4}}{n^{!6}}$	$\frac{r^{!3}(n-r)^{!3}}{n^{!6}}$	$\frac{r^{!4}(n-r)^{!2}}{n^{!6}}$	$\frac{r^{!5}(n-r)}{n^{!6}}$	$\frac{r^{!6}}{n^{!6}}$
m_{++}^3					+ 1						+ 1
$m_{++} \sum_i m_{i+}^2$				+ 12	- 12					+ 12	
$m_{++} \sum_{ij} m_{ij}^2$			+ 6	- 12	+ 6				+ 6		
$\sum_i m_{i+}^3$			+ 8	- 24	+ 16				+ 8	- 8	
$\sum_{ij} m_{ij}\ m_{i+}\ m_{j+}$			+ 24	- 48	+ 24				+ 24		
$\sum_{ijk} m_{ij}\ m_{ik}\ m_{jk}$		+ 8	- 24	+ 24	- 8			+ 8			
$\sum_{ij} m_{ij}^2\ m_{i+}$		+ 24	- 96	+ 120	- 48			+ 24	- 24		
$\sum_{ij} m_{ij}^3$	+ 4	- 24	+ 52	- 48	+ 16		+ 4	- 8	+ 4		

1322 48/49 753/832

Table 3.3.2 : An expression for Ex^4.

	$16\,\frac{r!^2(n-r)!^6}{n!^8}$	$16\,\frac{r!^3(n-r)!^5}{n!^8}$	$16\,\frac{r!^4(n-r)!^4}{n!^8}$	$16\,\frac{r!^5(n-r)!^3}{n!^8}$	$16\,\frac{r!^6(n-r)!^2}{n!^8}$	$16\,\frac{r!^7(n-r)!}{n!^8}$	$16\,\frac{r!^8}{n!^8}$
m_{++}^4							+ 1/16
$m_{++}^2 \sum_i m_{i+}^2$						+ 3/2	
$m_{++}^2 \sum_{ij} m_{ij}^2$					+ 3/4		
$\left(\sum_i m_{i+}^2\right)^2$					+ 3		
$m_{++} \sum_i m_{i+}^3$					+ 2	- 2	
$m_{++} \sum_{ij} m_{ij}\, m_{i+}\, m_{j+}$					+ 6		
$m_{++} \sum_{ij} m_{ij}^2\, m_{i+}$				+ 6	- 6		
$m_{++} \sum_{ijk} m_{ij}\, m_{ik}\, m_{jk}$				+ 2			
$\sum_i m_{i+}^2 \sum_{jk} m_{jk}^2$				+ 3			
$\sum_{ijk} m_{ij}\, m_{ik}\, m_{j+}\, m_{k+}$				+ 12			
$\sum_{ij} m_{ij}\, m_{i+}^2\, m_{j+}$				+ 12	- 12		
$\sum_i m_{i+}^4$				+ 1	- 4	+ 1	
$m_{++} \sum_{ij} m_{ij}^3$			+ 1	- 2	+ 1		
$\left(\sum_{ij} m_{ij}^2\right)^2$			+ 3/4				
$\sum_{ijk} m_{ij}\, m_{ik}\, m_{jk}\, m_{i+}$			+ 12	- 12			
$\sum_{ijkl} m_{ij}\, m_{ik}\, m_{jl}\, m_{kl}$			+ 3				
$\sum_{ij} m_{ij}^2\, m_{i+}^2$			+ 6	- 24	+ 6		
$\sum_{ij} m_{ij}^2\, m_{i+}\, m_{j+}$			+ 6	- 12	+ 6		
$\sum_{ijk} m_{ij}^2\, m_{ik}\, m_{k+}$			+ 12	- 12			
$\sum_{ijk} m_{ij}^2\, m_{ik}\, m_{jk}$		+ 6	- 12	+ 6			
$\sum_{ij} m_{ij}^3\, m_{i+}$		+ 4	- 20	+ 20	- 4		
$\sum_{ijk} m_{ij}^2\, m_{ik}^2$		+ 3	- 12	+ 3			
$\sum_{ij} m_{ij}^4$	+ 1/2	- 4	+ 9	- 4	+ 1/2		

114

The right-hand part of the table is obtained by some simplifications. The third reduced moment is now only a matter of easy algebra.

$$(3.3.10) \quad \mu_3 = 8\,\frac{r^{!4}(n-r)}{n^{!6}}\,(n-2r+3)\sum_i\left(m_{i+}-\frac{m_{++}}{n}\right)^3 +$$
$$-\,24\,\frac{r^{!3}(n-r)^{!2}}{n^{!2}\,n^{!6}}\left\{(2r-3)n-8r+9\right\}m_{++}\sum_i\left(m_{i+}-\frac{m_{++}}{n}\right)^2 +$$
$$+\,24\,\frac{r^{!3}(n-r)^{!2}}{n^{!6}}\,(n-2r+1)\sum_{ij} m_{ij}^2\left(m_{i+}-\frac{m_{++}}{n}\right) +$$
$$+\,8\,\frac{r^{!3}(n-r)^{!3}}{n^{!6}}\sum_{ijk} m_{ij}\,m_{ik}\,m_{jk} +$$
$$+\,4\,\frac{r^{!2}(n-r)^{!2}}{n^{!6}}\left\{n^2-(4r+1)n+4r^2+4\right\}\sum_{ij} m_{ij}^3 +$$
$$+\,24\,\frac{r^{!4}(n-r)^{!2}}{n^{!6}}\sum_{ij} m_{ij}\,m_{i+}\,m_{j+} +$$
$$-\,12\,\frac{r^{!2}(n-r)^{!2}}{n^{!2}\,n^{!6}}\left\{n^2-3(2r-1)n+6r^2-4\right\}m_{++}\sum_{ij} m_{ij}^2 +$$
$$-\,8\,\frac{r^{!2}(n-r)^{!2}}{(n^{!2})^2\,n^{!6}}\left\{(3r^2-14r+15)n^2-(7r^2-35r+33)n-2r^2-15r+18\right\}m_{++}^3 \quad .$$

For the special case: m_{i+} independent of i, μ_3 reduces to a simpler expression:

$$(3.3.11) \quad \mu_3 = 8\,\frac{r^{!3}(n-r)^{!3}}{n^{!6}}\sum_{ijk} m_{ij}\,m_{ik}\,m_{jk} +$$
$$+\,4\,\frac{r^{!2}(n-r)^{!2}}{n^{!6}}\left\{n^2-(4r+1)n+4r^2+4\right\}\sum_{ij} m_{ij}^3 +$$
$$-\,12\,\frac{r^{!2}(n-r)^{!2}}{n^{!2}\,n^{!6}}\left\{n^2-3(2r-1)n+6r^2-4\right\}m_{++}\sum_{ij} m_{ij}^2 +$$
$$-\,8\,\frac{r^{!2}(n-r)^{!2}}{(n^{!2})^2\,n^{!6}}\left\{(r-3)n^2-(r^2-5r-3)n-5r^2\right\}m_{++}^3 \quad .$$

The fourth moment of $\underline{x}$ has been calculated and tabled in table 3.3.2.

3.4. The moments of $\underline{x}_B$ and $\underline{x}_W$ - free sampling

Again the moments of $\underline{x}_W$ can be obtained from those of $\underline{x}_B$ by changing p_1 into p_2. It is sufficient therefore to consider only the moments of $\underline{x}_B$. We write $\underline{x}$ for $\underline{x}_B$ and p for p_1.

For $i \neq j$

(3.4.1) $$E\underline{x}_{ij} = P[\underline{x}_{ij}=1] = P[\text{points } i \text{ and } j \text{ are chosen}] = $$
$$= P[\text{point } i \text{ is chosen}] \cdot P[\text{point } j \text{ is chosen}] = $$
$$= p.p = p^2 .$$

In general

(3.4.2) $$E\left(\underline{x}^{(1)} \ldots \underline{x}^{(k)} \mid \sum_{i=1}^{h} C_{k_i, l_i}^{(\alpha_i)}\right) = P[l \text{ given points are chosen}] = p^l ,$$

where $\sum_{i=1}^{h} l_i = l$.

So

(3.4.3) $$E\underline{x} = m_{++}p^2$$
$$E\underline{x}^2 = m_{++}^2 p^4 + 2 \sum_{ij} m_{ij}^2 . p^2(1-p)^2 + 4 \sum_i m_{i+}^2 \cdot p^3(1-p)$$

thus

(3.4.4) $$\sigma^2 = 2p^2(1-p) \left\{ (1-p) \sum_{ij} m_{ij}^2 + 2p \sum_i m_{i+}^2 \right\} .$$

In the same way

$$\mu_3 = 4p^2(1-p) \left\{ 2p^2(1-2p) \sum_i m_{i+}^3 + 6p^2(1-p) \sum_{ij} m_{ij} m_{i+} m_{j+} + \right.$$
$$+ 2p(1-p)^2 \sum_{ijk} m_{ij} m_{ik} m_{jk} + 6p(1-p)(1-2p) \sum_{ij} m_{ij}^2 m_{i+} +$$
$$\left. + (1-p)(1-2p)^2 \sum_{ij} m_{ij}^3 \right\} .$$

3.5. The moments of $\underline{y}$ - non free sampling

In order to calculate

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

take a point P_i of the i-th connected component 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 by a join to these black points are coloured white. Repeat this procedure. If 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 non-bichromatic.

If no such situation arises one arrives at a stage where all points of the i-th component have been allotted to one of the colours, viz. τ_i points are of the same colour as the reference point, and $l_i - \tau_i$ points are of the other colour. In this case the i-th component is called bichromatic. If all components of a graph are bichromatic, the graph is called bichromatic; a graph is non-bichromatic if at least one of its components is non-bichromatic.

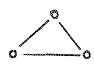
The decision whether a graph is bichromatic or not does not depend on a particular choice of the reference point, but only on the fact whether a graph contains cycles of odd length (in which case it is non-bichromatic, cf. C. BERGE, 1958, 31) or not. It is therefore a property of the configuration, and we shall speak of bichromatic and non-bichromatic configurations accordingly.

Define

$$\mathcal{B}\left(\sum_{i=1}^{h} C_{k_i, l_i}^{(\alpha_i)}\right) = \begin{cases} 1 & \text{if } \sum_{i=1}^{h} C_{k_i, l_i}^{(\alpha_i)} \text{ is bichromatic,} \\ 0 & \text{if not.} \end{cases}$$

Obviously

$$\mathcal{B}\left(\sum_{i=1}^{h} C_{k_i, l_i}^{(\alpha_i)}\right) = \prod_{i=1}^{h} \mathcal{B}\left(C_{k_i, l_i}^{(\alpha_i)}\right). \tag{3.5.1}$$

An example of a non-bichromatic configuration is $C_{3,3}^{(2)}$ =  .

Now consider the expectation of a product of $\underline{y}$'s, whose graph has the configuration $C_{3,3}^{(2)}$

$$E\, \underline{y}_{12}\, \underline{y}_{13}\, \underline{y}_{23} = P\left[\underline{y}_{12}=1 \cap \underline{y}_{13}=1 \cap \underline{y}_{23}=1\right] .$$

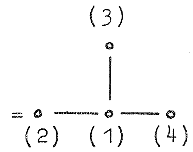
However the event "$\underline{y}_{12}=1 \cap \underline{y}_{13}=1 \cap \underline{y}_{23}=1$" is impossible, because of the fact that $C_{3,3}^{(2)}$ is non-bichromatic, thus

$$E\left(\underline{y}^{(1)}\,\underline{y}^{(2)}\,\underline{y}^{(3)} \mid C_{3,3}^{(2)}\right) = 0 \quad ,$$

and in general for a non-bichromatic configuration $\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)}$

$$(3.5.2) \qquad E\left(\underline{y}^{(1)}\ldots\underline{y}^{(k)} \mid \sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)}\right) = 0 \quad .$$

Consider as an example of a bichromatic configuration $C_{3,4}^{(1)}$ =



Choose the point indicated as (1) as the reference point. Colour this point white and then colour the other points according to the above-mentioned procedure. To the points (2), (3) and (4) the colour black is allotted. Thus $\tau=1$, $l-\tau=3$. One might also colour the reference point black, then again one arrives at $\tau=1$, $l-\tau=3$. Now the expectation of a product of $\underline{y}$'s, whose graph has the configuration $C_{3,4}^{(1)}$ is

$$E\,\underline{y}_{12}\,\underline{y}_{13}\,\underline{y}_{14} = P\left[\underline{y}_{12}=1 \cap \underline{y}_{13}=1 \cap \underline{y}_{14}=1\right] =$$

$$= P\left[\underline{y}_{12}=1 \cap \underline{y}_{13}=1 \cap \underline{y}_{14}=1 \cap \text{point 1 is white}\right] +$$

$$+ P\left[\underline{y}_{12}=1 \cap \underline{y}_{13}=1 \cap \underline{y}_{14}=1 \cap \text{point 1 is black}\right] \quad .$$

The event "$\underline{y}_{12}=1 \cap \underline{y}_{13}=1 \cap \underline{y}_{14}=1 \cap$ point 1 is white" occurs only when in the random sample of size r_1 from n points $l-\tau=3$ given points (viz. 2, 3 and 4) are included, while in the sample of size r_2 $\tau=1$ given point (viz. point 1) is included. Thus

$$P\left[\underline{y}_{12}=1 \cap \underline{y}_{13}=1 \cap \underline{y}_{14}=1 \cap \text{point 1 is white}\right] = \frac{r_1^{!3}\,r_2}{n^{!4}} \quad .$$

The event "$\underline{y}_{12}=1 \cap \underline{y}_{13}=1 \cap \underline{y}_{14}=1 \cap$ point 1 is black" occurs only when in the sample of size r_1 $\tau=1$ given point is included, and in the sample of size r_2 $l-\tau=3$ given points. Thus

$$P\left[\underline{y}_{12}=1 \cap \underline{y}_{13}=1 \cap \underline{y}_{14}=1 \cap \text{point 1 is black}\right] = \frac{r_1\,r_2^{!3}}{n^{!4}} \quad .$$

Therefore

$$E\left(\underline{y}^{(1)}\, \underline{y}^{(2)}\, \underline{y}^{(3)} \mid C_{3,4}^{(1)} \right) = \frac{r_1^{!3} r_2 + r_1 r_2^{!3}}{n^{!4}} ,$$

and in general for a bichromatic configuration $C_{k,l}^{(\alpha)}$

$$(3.5.3) \qquad E\left(\underline{y}^{(1)} \ldots \underline{y}^{(k)} \mid C_{k,l}^{(\alpha)} \right) = \frac{r_1^{!\tau} r_2^{!l-\tau} + r_1^{!l-\tau} r_2^{!\tau}}{n^{!l}} .$$

The corresponding result for a bichromatic configuration $\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)}$ is slightly more complex. Let the statement "P_i=black" (c.q. white) mean that the colouring procedure applied to the i-th connected component starts with colouring P_i black (c.q. white); it ends up with τ_i black (white) and $l_i-\tau_i$ white (black) points.

Let $\rho_i = 0$ or 1 $(i=1,2,\ldots,h)$ and let the statement

$$(3.5.4) \qquad \text{"} \sum_{i=1}^{h} \rho_i \; (P_i\text{=black}) \; + \; \sum_{i=1}^{h} (1-\rho_i) \; (P_i\text{=white}) \text{"}$$

mean that for an i for which $\rho_i = 1$ the statement "P_i=black" holds, and for an i for which $\rho_i = 0$ the statement "P_i=white". Then (3.5.4) means that after colouring the points of the configuration

$$\sum_{i=1}^{h} \rho_i \, \tau_i + \sum_{i=1}^{h} (1-\rho_i)(l_i-\tau_i) \qquad \text{points are black, and}$$

$$\sum_{i=1}^{h} (1-\rho_i)\tau_i + \sum_{i=1}^{h} \rho_i(l_i-\tau_i) \qquad \text{points are white.}$$

Now consider the event

$$\underline{y}^{(1)} \ldots \underline{y}^{(k)} = 1 ,$$

where the configuration of the graph corresponding to the product of $\underline{y}$'s is $\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)}$.

The event

$$\underline{y}^{(1)} \ldots \underline{y}^{(k)} = 1 \cap (3.5.4)$$

occurs only if in the sample of r_1 black points

$$\sum_{i=1}^{h} \rho_i \tau_i + \sum_{i=1}^{h} (1-\rho_i)(l_i-\tau_i)$$

given points are included, and in the sample of r_2 white points

$$\sum_{i=1}^{h} (1-\rho_i)\tau_i + \sum_{i=1}^{h} \rho_i(l_i-\tau_i)$$

given points are included.

Thus

$$P\left[\underline{y}^{(1)}\ldots\underline{y}^{(k)} = 1 \cap (3.5.4)\right] =$$

$$= \frac{{}^{!}\left\{\sum_{i=1}^{h}\rho_i\tau_i + \sum_{i=1}^{h}(1-\rho_i)(l_i-\tau_i)\right\}_{r_1} \cdot {}^{!}\left\{\sum_{i=1}^{h}(1-\rho_i)\tau_i + \sum_{i=1}^{h}\rho_i(l_i-\tau_i)\right\}_{r_2}}{n^{!l}} .$$

Summing over all values of $\rho_1, \rho_2, \ldots, \rho_h$ leads to the expectation of $\underline{y}^{(1)}\ldots\underline{y}^{(k)}$.

Incorporating (3.5.2) we thus obtain for an arbitrary configuration $\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)}$, with $\sum_{i=1}^{h} k_i = k$, $\sum_{i=1}^{h} l_i = l$,

$$(3.5.5) \qquad E\left\{\underline{y}^{(1)}\ldots\underline{y}^{(k)} \mid \sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)}\right\} = \mathcal{B}\left(\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)}\right) \sum_{\rho_1=0}^{1} \cdots \sum_{\rho_h=0}^{1}$$

$$\frac{{}^{!}\left\{\sum_{i=1}^{h}\rho_i\tau_i + \sum_{i=1}^{h}(1-\rho_i)(l_i-\tau_i)\right\}_{r_1} \cdot {}^{!}\left\{\sum_{i=1}^{h}(1-\rho_i)\tau_i + \sum_{i=1}^{h}\rho_i(l_i-\tau_i)\right\}_{r_2}}{n^{!l}} .$$

This result shows that the symmetry assumption introduced in section 3.1 holds for $\underline{y}$.

Thus e.g.

$$E\,\underline{y}_{ij} = \frac{2r_1r_2}{n^{!2}} ,$$

and

$$(3.5.6) \qquad E\,\underline{y} = \frac{2r_1r_2}{n^{!2}}\, m_{++} .$$

Also

$$E\,\underline{y}_{ij}\,\underline{y}_{kl} = \frac{4r_1^{!2}r_2^{!2}}{n^{!4}} \qquad (i,j,k,l) \neq$$

$$E\,\underline{y}_{ij}\,\underline{y}_{ik} = \frac{r_1r_2^{!2} + r_1^{!2}r_2}{n^{!3}} \qquad (i,j,k) \neq$$

$$E\,\underline{y}_{ij}^2 = \frac{2r_1r_2}{n^{!2}} \quad .$$

$E\underline{y}^2$ can readily be calculated in the same way as $E\underline{x}^2$ in section 3.3 . From $E\underline{y}^2$ and (3.5.6) it follows that

$$(3.5.7)\quad \sigma^2 = \frac{4r_1r_2}{n^{!4}}\left[\sum_i\left(m_{i+} - \frac{m_{++}}{n}\right)^2\left\{n(r_1+r_2-2) - 4r_1r_2 + r_1 + r_2 + 2\right\} + \right.$$
$$+ \sum_{ij} m_{ij}^2\left\{(n-r_1-r_2)(n-3) + 2(r_1-1)(r_2-1)\right\} +$$
$$\left. + \; m_{++}^2\left\{\frac{-n^2 + n(r_1+r_2+3) - 2r_1r_2 - r_1 - r_2 - 2}{n(n-1)}\right\}\right] .$$

For reference purposes note that for a connected component $C_{1,2}$ $l_i - \tau_i = \tau_i = 1$.

Thus

$$(3.5.8)\quad E\left(\underline{y}^{(1)}\ldots\underline{y}^{(k+g)} \mid \sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)} + gC_{1,2}\right) = 2^g\mathcal{B}\left(\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)}\right)\sum_{\rho_1=0}^{1}\ldots\sum_{\rho_h=0}^{1}$$

$$\frac{r_1^{!\left\{\sum_{i=1}^{h}\rho_i\tau_i + \sum_{i=1}^{h}(1-\rho_i)(l_i-\tau_i) + g\right\}} \cdot r_2^{!\left\{\sum_{i=1}^{h}(1-\rho_i)\tau_i + \sum_{i=1}^{h}\rho_i(l_i-\tau_i) + g\right\}}}{n^{!(l+2g)}} ,$$

where $\sum_{i=1}^{h} k_i = k$ and $\sum_{i=1}^{h} l_i = l$.

3.6. The moments of $\underline{y}$ - free sampling

For a configuration $C_{k,l}^{(\alpha)}$

$$E\left(\underline{y}^{(1)}\ldots\underline{y}^{(k)} \mid C_{k,l}^{(\alpha)}\right) = \mathcal{B}\left(C_{k,l}^{(\alpha)}\right)\left(p_1^{\tau}p_2^{l-\tau} + p_1^{l-\tau}p_2^{\tau}\right) .$$

Since for free sampling

$$E\left(\underline{y}^{(1)}\ldots\underline{y}^{(k)} \mid \sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)}\right) = \prod_{i=1}^{h} E\left(\underline{y}^{(1)}\ldots\underline{y}^{(k_i)} \mid C_{k_i,l_i}^{(\alpha_i)}\right) , \qquad \sum_{i=1}^{h} k_i = k ,$$

we have

$$(3.6.1)\quad E\left(\underline{y}^{(1)}\ldots\underline{y}^{(k)} \mid \sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)}\right) = \mathcal{B}\left(\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)}\right) \prod_{i=1}^{h}\left\{p_1^{\tau_i}p_2^{l_i-\tau_i} + p_1^{l_i-\tau_i}p_2^{\tau_i}\right\} .$$

E.g.

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

thus

$$(3.6.2)\qquad E\,\underline{y} = 2p_1p_2m_{++} .$$

Also

$$(3.6.3)\qquad \sigma^2 = 4p_1p_2(p_1+p_2-4p_1p_2)\sum_i m_{i+}^2 + 4p_1p_2(1-p_1-p_2+2p_1p_2)\sum_{ij} m_{ij}^2 .$$

3.7. A property of the random variables $\underline{v}_{ij}$

If in a blank graph to each point an even number of joins are connected, the graph will be said to have an even-joined configuration. If a configuration is not even-joined, it will be called odd-joined.

Theorem 3.7.1.

Let $r_2 = n-r_1$, then in the case of non free sampling,

$$\sum_{r_1=0}^{n}\binom{n}{r_1}E\left(\underline{v}^{(1)}\ldots\underline{v}^{(k)} \mid \sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)}\right) = \begin{cases} 0 & \text{if } \sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)} \text{ is odd-joined,} \\ & \text{or (and) } l=\sum_{i=1}^{h} l_i > n, \\ 2^n & \text{if } \sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)} \text{ is even-joined} \\ & \text{and } l=\sum_{i=1}^{h} l_i \leq n. \end{cases}$$

Proof.

Consider

(3.7.1) $$E\,\underline{v}_{\zeta_{\mu_1'},\zeta_{\nu_1'}}\cdots\underline{v}_{\zeta_{\mu_k'},\zeta_{\nu_k'}}\ ,$$

where in the set of numbers $\{\mu_1',\mu_2',\ldots,\mu_k',\nu_1',\nu_2',\ldots,\nu_k'\}$ $(\mu_j'\neq\nu_j', j=1,2,..,k)$ all numbers $1,2,\ldots,l$ are represented and where $\zeta_1,\zeta_2,\ldots,\zeta_l$ are l unequal numbers from $\{1,2,\ldots,n\}$. We suppose $l\leq n$. As shown in section 2.2. there corresponds a graph to (3.7.1). Let the configuration of this graph be odd-joined, which means that there is at least one point of the graph, which is connected by an odd number, say t, joins to s other points $(1\leq t\leq k,\ 1\leq s\leq \min(l-1,t))$. This means with regard to (3.7.1) that there is at least one ζ, say ζ_{α_1}, occurring as a subscript to t factors $\underline{v}$, while s other ζ's, say $\zeta_{\alpha_2},\zeta_{\alpha_3},\ldots,\zeta_{\alpha_{s+1}}$ occur each together with ζ_{α_1} as a subscript to at least one factor $\underline{v}$ $(\alpha_1,\alpha_2,\ldots,\alpha_{s+1}$ are unequal numbers from $\{1,2,\ldots,l\})$. As the expectation (3.7.1) depends only on the configuration, it is equal to

(3.7.2) $$E\,\underline{v}_{\theta_{\mu_1'},\theta_{\nu_1'}}\cdots\underline{v}_{\theta_{\mu_k'},\theta_{\nu_k'}}\ ,$$

where $\theta_1,\theta_2,\ldots,\theta_l$ is some permutation of $\zeta_1,\zeta_2,\ldots,\zeta_l$ (cf. section 2.2.). Choose

$$\theta_1=\zeta_{\alpha_1},\ \theta_2=\zeta_{\alpha_2},\ \ldots\ ,\ \theta_{s+1}=\zeta_{\alpha_{s+1}}\ ,$$

and for $\theta_{s+2},\ \theta_{s+3},\ \ldots\ ,\ \theta_l$ a permutation of the remaining ζ's. With this choice of the permutation, θ_1 occurs in (3.7.2) as a subscript to t factors $\underline{v}$, the other subscripts of these $\underline{v}$'s being $\theta_2,\ldots,\theta_{s+1}$, each at least once.

Now permuting factors $\underline{v}$ and interchanging subscripts to one $\underline{v}$ does not change the value of (3.7.2). By carrying out this procedure in a suitable way, one can write for (3.7.2)

(3.7.3) $$E\,\underline{v}_{\theta_1,\theta_{\nu_1}}\cdots\underline{v}_{\theta_1,\theta_{\nu_t}}\,\underline{v}_{\theta_{\mu_{t+1}},\theta_{\nu_{t+1}}}\cdots\underline{v}_{\theta_{\mu_k},\theta_{\nu_k}}\ ,$$

where among the set of numbers $\{\nu_1,\nu_2,\ldots,\nu_t\}$ all numbers $2,3,\ldots,(s+1)$ occur and among $\{\mu_{t+1},\mu_{t+2},\ldots,\mu_k,\ \nu_1,\nu_2,\ldots,\nu_k\}$ all numbers $2,3,\ldots,l$.

Let Θ be the set with elements $\theta_2,\theta_3,\ldots,\theta_l$ and let $\Theta^{(1)}$ and $\Theta^{(2)}$ be two given subsets of Θ, of l_1 and $l-l_1-1$ elements respectively such that $\Theta^{(1)} \cap \Theta^{(2)} = \emptyset$ and $\Theta^{(1)} \cup \Theta^{(2)} = \Theta$. We remind of the fact that $\theta_2,\theta_3,\ldots,\theta_l$ are unequal numbers from $\{1,2,\ldots,n\}$, and refer to points of the master graph (cf. section 1.1.). Consider the event E: the points of the master graph, whose numbers are elements of $\Theta^{(1)}$ are black and those whose numbers are elements of $\Theta^{(2)}$ are white. As r_1 points of the master graph are chosen at random from n points and are coloured black, while the remaining ones are coloured white

$$(3.7.4) \qquad P\left[E \mid \Theta^{(1)}, \Theta^{(2)}\right] = \frac{r_1^{!l_1}(n-r_1)^{!(l-l_1-1)}}{n^{!(l-1)}} \quad .$$

For $j = t+1,t+2,\ldots,k$

$$\underline{v}_{\theta_{\mu_j},\theta_{\nu_j}} = \underline{x}^{(B)}_{\theta_{\mu_j},\theta_{\nu_j}} + \underline{x}^{(W)}_{\theta_{\mu_j},\theta_{\nu_j}} - \underline{y}_{\theta_{\mu_j},\theta_{\nu_j}}$$

takes, conditionally on E, the value $+1$ if both θ_{μ_j} and θ_{ν_j} belong to $\Theta^{(1)}$, or both to $\Theta^{(2)}$, and -1 otherwise.

Conditionally on E

$$(3.7.5) \qquad \prod_{j=t+1}^{k} \underline{v}_{\theta_{\mu_j},\theta_{\nu_j}}$$

takes the value $+1$ or -1, depending on whether an even or an odd number of the $\underline{v}$'s take a value -1.
Suppose (3.7.5) takes the value $+1$.

Let out of $\theta_2,\theta_3,\ldots,\theta_{s+1}$ l_2 θ's, say $\theta_2,\theta_3,\ldots,\theta_{l_2+1}$ belong to $\Theta^{(1)}$, while $\theta_{l_2+2},\theta_{l_2+3},\ldots,\theta_{s+1}$ belong to $\Theta^{(2)}$.

If the point of the master graph with number θ_1 belongs to the black points, the probability of this event being conditionally on E

$$\frac{r_1-l_1}{n-l+1} \quad ,$$

then also conditionally on E

$$(3.7.6) \qquad \underline{v}_{\theta_1,\theta_2},\ldots,\underline{v}_{\theta_1,\theta_{l_2+1}}$$

take the value $+1$, and

(3.7.7) $$\underline{v}_{\Theta_1,\Theta_{l_2+2}},\ldots,\underline{v}_{\Theta_1,\Theta_{s+1}}$$

take the value -1 . The value of

(3.7.8) $$\prod_{i=1}^{t} \underline{v}_{\Theta_1,\Theta_{\nu_i}}$$

is therefore completely determined on the said condition. Suppose this value is +1 .

If, however, Θ_1 belongs to the white points, the probability of this event being on the condition E

$$\frac{n-r_1-l+l_1+1}{n-l+1} \quad ,$$

then each of the $\underline{v}$'s in (3.7.6) is equal to -1 , and each of those in (3.7.7) is +1 . The value of (3.7.8) is therefore $(-1)^t$ times the value it assumes in the case where Θ_1 is black. As t is odd, the value of (3.7.8) is now -1 . Thus conditionally on E

(3.7.9) $$\underline{v}_{\Theta_1,\Theta_{\nu_1}}\cdots\underline{v}_{\Theta_1,\Theta_{\nu_t}}\ \underline{v}_{\Theta_{\mu_{t+1}},\Theta_{\nu_{t+1}}}\cdots\underline{v}_{\Theta_{\mu_k},\Theta_{\nu_k}}$$

is equal to +1 if and only if Θ_1 is black, and is equal to -1 if and only if Θ_1 is white. Hence the expectation of (3.7.9), conditional on E is

$$P\left[\Theta_1 \text{ is black} \mid E\right] - P\left[\Theta_2 \text{ is white} \mid E\right] = \frac{r_1-l_1}{n-l+1} - \frac{n-r_1-l+l_1+1}{n-l+1} \ .$$

Multiplication with $P\left[E \mid \Theta^{(1)}, \Theta^{(2)}\right]$ gives as the contribution to the expectation of (3.7.9) for the given sets $\Theta^{(1)}$ and $\Theta^{(2)}$

$$\frac{r_1^{!(l_1+1)}(n-r_1)^{!(l-l_1-1)}}{n^{!l}} - \frac{r_1^{!l_1}(n-r_1)^{!(l-l_1)}}{n^{!l}} \ .$$

Multiplying by $\binom{n}{r_1}$ and summing gives the contribution to

$$\sum_{r_1=0}^{n} \binom{n}{r_1} E\left(\underline{v}^{(1)}\ldots\underline{v}^{(k)} \mid \sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)}\right)$$ for the given sets $\Theta^{(1)}$ and $\Theta^{(2)}$:

$$\sum_{r_1=0}^{n} \binom{n}{r_1} \frac{r_1^{!(l_1+1)} (n-r_1)^{!(l-l_1-1)}}{n^{!l}} - \sum_{r_1=0}^{n} \binom{n}{r_1} \frac{r_1^{!l_1} (n-r_1)^{!(l-l_1)}}{n^{!l}} =$$

$$= \sum_{r_1=l_1+1}^{n-l+l_1+1} \binom{n-l}{r_1-l_1-1} - \sum_{r_1=l_1}^{n-l+l_1} \binom{n-l}{r_1-l_1} = 2^{n-l} - 2^{n-l} = 0 \ .$$

The same result is obtained if (3.7.8) is -1 , and also if (3.7.5) is supposed to be -1 , and (3.7.8) either -1 or $+1$. It holds therefore for every pair of sets $\Theta^{(1)}, \Theta^{(2)}$ with $\Theta^{(1)} \cap \Theta^{(2)} = \emptyset$ and $\Theta^{(1)} \cup \Theta^{(2)} = \Theta$, and thus holds in general. As the part of theorem 3.7.1 relating to the case $l = \sum_{i=1}^{h} l_i > n$ is trivial, this proves the first part of the theorem.

To prove the second part note that a graph with an even-joined configuration can be decomposed into elementary cycles, which are cycles which may have points but no joins in common. (cf. D.KOENIG, (1936)). Such a decomposition does not need to be unique.

Consider again (3.7.1) and suppose the graph (in which the point-labels are retained) corresponding to it to have an even-joined configuration. Decompose the graph into elementary cycles. Permute factors $\underline{v}$ in (3.7.1) - which means also: permute the join-labels of the graph - in such a way that the joins of one of the elementary cycles are numbered $1,2,\dots,l_1$ in this order (joins 1 and l_1 being connected to one point). This permutation procedure leads to the expectation (if necessary with interchanging subscripts to single factors v)

$$\text{(3.7.10)} \quad E\, \underline{v}_{\zeta_{\mu_1''},\zeta_{\nu_1''}} \underline{v}_{\zeta_{\nu_1''},\zeta_{\nu_2''}} \cdots \underline{v}_{\zeta_{\nu_{l_1-1}''},\zeta_{\mu_1''}} \underline{v}_{\zeta_{\mu_{l_1+1}''},\zeta_{\nu_{l_1+1}''}} \cdots \underline{v}_{\zeta_{\mu_k''},\zeta_{\nu_k''}}$$

which is equal to (3.7.1). The value of (3.7.10) is not changed if $\zeta_1,\zeta_2,\dots,\zeta_l$ are permuted. Let $\Theta_1,\Theta_2,\dots,\Theta_l$ be the permutation of $\zeta_1,\zeta_2,\dots,\zeta_l$ with $\Theta_1 = \zeta_{\mu_1''}$, $\Theta_2 = \zeta_{\nu_1''}$, ... , $\Theta_{l_1} = \zeta_{\nu_{l_1-1}''}$, while $\Theta_{l_1+1},\Theta_{l_1+2},\dots,\Theta_k$ form some permutation of the remaining ζ's. Then (3.7.1) is equal to

$$E\, \underline{v}_{\Theta_1,\Theta_2} \underline{v}_{\Theta_2,\Theta_3} \cdots \underline{v}_{\Theta_{l_1},\Theta_1} \underline{v}_{\Theta_{\mu_{l_1+1}},\Theta_{\nu_{l_1+1}}} \cdots \underline{v}_{\Theta_{\mu_k},\Theta_{\nu_k}} \ .$$

To prove the second part of the lemma it is sufficient to show that this expectation is equal to one. Again this is true if

$$(3.7.11)\quad P\left[\underline{v}_{\theta_1,\theta_2}\,\underline{v}_{\theta_2,\theta_3}\cdots\underline{v}_{\theta_{l_1},\theta_1}\,\underline{v}_{\theta_{\mu_{l_1+1}},\theta_{\nu_{l_1+1}}}\cdots\underline{v}_{\theta_{\mu_k},\theta_{\nu_k}} = 1\right] = 1.$$

Now r_1 points of the master graph are chosen from the n points and are coloured black, while the remaining ones are coloured white. Let $\Theta^{(1)}$ and $\Theta^{(2)}$ be the two subsets of $\Theta = \{\theta_1,\ldots,\theta_l\}$, such that those θ's which correspond to black points are elements of $\Theta^{(1)}$, and those which correspond to white points are elements of $\Theta^{(2)}$. We now prove that

$$(3.7.12)\qquad \underline{v}_{\theta_1,\theta_2}\,\underline{v}_{\theta_2,\theta_3}\cdots\underline{v}_{\theta_t,\theta_{t+1}}$$

takes the value $+1$ if θ_1 and θ_{t+1} belong both to $\Theta^{(1)}$, or both to $\Theta^{(2)}$, and -1 otherwise. For t=1 this is evident from

$$\underline{v}_{\theta_1,\theta_2} = \underline{x}^{(B)}_{\theta_1,\theta_2} + \underline{x}^{(W)}_{\theta_1,\theta_2} - \underline{y}_{\theta_1,\theta_2} .$$

Assume (3.7.12) to be true for all positive values of t smaller than or equal to t_1 ($t_1 \geq 1$). We then show it to hold for t_1+1 as well.
For $t=t_1+1$

$$(3.7.13)\quad \underline{v}_{\theta_1,\theta_2}\,\underline{v}_{\theta_2,\theta_3}\cdots\underline{v}_{\theta_{t_1+1},\theta_{t_1+2}} = \left[\underline{v}_{\theta_1,\theta_2}\cdots\underline{v}_{\theta_{t_1},\theta_{t_1+1}}\right]\cdot\underline{v}_{\theta_{t_1+1},\theta_{t_1+2}}.$$

Now if $\theta_1 \in \Theta^{(1)}$ and $\theta_{t_1+1} \in \Theta^{(1)}$ the first part of the second member of (3.7.13) is $+1$ by assumption. The second part is $+1$ if $\theta_{t_1+1} \in \Theta^{(1)}$ and $\theta_{t_1+2} \in \Theta^{(1)}$, or if $\theta_{t_1+1} \in \Theta^{(2)}$ and $\theta_{t_1+2} \in \Theta^{(2)}$. Thus if $\theta_1 \in \Theta^{(1)}$ (3.7.13) is equal to $+1$ if $\theta_{t+2} \in \Theta^{(1)}$. In the same way one shows that if $\theta_1 \in \Theta^{(2)}$, (3.7.13) is equal to $+1$ if also $\theta_{t_1+2} \in \Theta^{(2)}$. This proves that (3.7.13) is equal to $+1$ if θ_1 and θ_{t_1+2} belong to the same set $\Theta^{(i)}$.

In the same way the -1 part of the statement regarding (3.7.13) is proved.

By induction the result connected with (3.7.12) follows. In particular

$$\underline{v}_{\theta_1,\theta_2}\,\underline{v}_{\theta_2,\theta_3}\cdots\underline{v}_{\theta_{l_1},\theta_1}$$

takes the value +1 spr 0, because θ_1 appears at the beginning and the end of the sequence of subscripts. Thus the products of $\underline{v}$'s corresponding to each elementary cycle of the graph take the value +1 spr 0, conditionally on $\Theta^{(1)}$ and $\Theta^{(2)}$. This holds for every pair $\Theta^{(1)}$ and $\Theta^{(2)}$ satisfying $\Theta^{(1)} \cap \Theta^{(2)} = \emptyset$ and $\Theta^{(1)} \cup \Theta^{(2)} = \Theta$. Thus (3.7.11) follows. This finishes the proof of Theorem 3.7.1.

The author stated this property of the random variables $\underline{v}_{ij}$ in a study concerning the ISING model of ferromagnetism (A.R. BLOEMENA (1960)) where it was used to derive the high temperature expansion of the partition function.

CHAPTER 4

4.1. Tendency towards the normal distribution (non free sampling)

Theorem 4.1.1.

i If r_1 and n tend to infinity in such a way that

$$\frac{r_1^2}{n} \longrightarrow \infty \quad \text{and} \quad \frac{r_1}{n} \leq 1 - \varepsilon$$

for some $\varepsilon > 0$, and if for all i

$$m_{i+} \leq c \quad ,$$

where c is a constant independent of i and n, then in the case of non free sampling 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 $\sigma^2(\underline{x}_B)$ are given by (3.3.1) and (3.3.8) with r replaced by r_1.

ii If in part i r_1 is replaced by r_2, and $\underline{x}_B$ by $\underline{x}_W$, the corresponding result for $\underline{x}_W$ is obtained.

Theorem 4.1.2.

If r_1 , r_2 and n tend to infinity in such a way that

$$\frac{r_1 r_2}{n} \longrightarrow \infty \quad \text{and} \quad \frac{r_1}{n} \leq 1 - \varepsilon \; , \quad \frac{r_2}{n} \leq 1 - \varepsilon$$

for some $\varepsilon > 0$, and if for all i

$$m_{i+} \leq c \quad ,$$

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

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

tends to the standard normal one. $E\underline{y}$ and $\sigma^2(\underline{y})$ are given by (3.5.6) and (3.5.7).

Note that the assumption $m_{i+} \leq c$ may be combined with (1.1.3) and (1.1.4) to give

$$1 \leq \sum_j m_{ij}^2 \leq m_{i+}^2 \leq c^2 \qquad \text{for all } i, \text{ so}$$

$$n \leq \sum_{ij} m_{ij}^2 \leq \sum_i m_{i+}^2 \leq c^2 n \quad ;$$

from $m_{i+}^2 \geq 1$ we also have

$$n \leq \sum_i m_{i+} = m_{++} \leq cn \quad .$$

We introduce the abbreviations

(4.1.1)
$$m_1(n) = \frac{1}{n} m_{++} \quad , \qquad 1 \leq m_1(n) \leq c \quad ,$$
$$m_2(n) = \frac{1}{n} \sum_{ij} m_{ij}^2 \quad , \qquad 1 \leq m_2(n) \leq c^2 \quad ,$$
$$m_3(n) = \frac{1}{n} \sum_i m_{i+}^2 \quad , \qquad 1 \leq m_3(n) \leq c^2 \quad .$$

We also note that $m_3(n) - m_1^2(n) = \frac{1}{n} \sum_i \left(m_{i+} - \frac{m_{++}}{n} \right)^2 \geq 0$.

To prove theorem 4.1.1 we start from (3.2.3) and show that $\mu_k \sigma^{-k}$ tend to the moments of the standard normal distribution. First we consider (writing $\underline{x}$ and r instead of $\underline{x}_B$ and r_1)

(4.1.2)
$$A = \sum_{\nu=0}^{g'} (-1)^{\nu} \binom{g'}{\nu} \left\{ E\left(\underline{x}^{(1)} \mid C_{1,2}\right) \right\}^{\nu} \cdot$$
$$\cdot \; E\left(\underline{x}^{(1)} \ldots \underline{x}^{(k-\nu)} \,\middle|\, \sum_{i=1}^{h} C_{k_i, l_i}^{(\alpha_i)} + (g+g'-\nu) C_{1,2} \right),$$

where $\sum_{i=1}^{h} k_i = k-g-g'$, $k_i \geq 2$ and $\sum_{i=1}^{h} l_i = l-2g-2g'$. In the appendix asymptotic expressions for A are given for the case where $\frac{r(n-r)}{n} \longrightarrow \infty$. Since the assumptions of the theorem also imply $\varepsilon \leq \frac{n-r}{n} \leq 1$ we may simplify these expressions slightly to obtain

(4.1.3) $$A = \frac{g'!}{(\frac{1}{2}g')!}\left(\frac{r}{n}\right)^{1}\left(-2.\frac{n-r}{rn}\right)^{\frac{1}{2}g'} + \mathcal{O}\left(r^{1-\frac{1}{2}g'-1}.n^{-1}\right) = \mathcal{O}\left(r^{1-\frac{1}{2}g'}.n^{-1}\right),$$

if g' is even, and

(4.1.4) $$A = \mathcal{O}\left(r^{1-\frac{1}{2}g'-\frac{1}{2}}.n^{-1}\right)$$

if g' is odd.

Also, from (3.3.8) we find that

(4.1.5) $$\sigma^2 = 2\,\frac{r^2(n-r)}{n^3}\left\{(n-r)m_2(n) + 2r\left(m_3(n)-m_1^2(n)\right)\right\} + \mathcal{O}\left(\frac{r}{n}\right) = \mathcal{O}\left(\frac{r^2}{n}\right),$$

since $\varepsilon \leq \frac{n-r}{n} \leq 1$; σ^2 cannot be $o\left(\frac{r^2}{n}\right)$ since $m_2(n) \geq 1$ and $m_3(n)-m_1^2(n) \geq 0$.

For each value of $g' = 0,1,\ldots,k-2,k$ there is a sum of finitely many terms in (3.2.3) contributing to μ_k (the last term in (3.2.3) corresponds to $g'=k$, $h'=0$). For odd values of g' such a sum is by lemma 2.3.4 and (4.1.4)

$$\mathcal{O}\left(r^{1-\frac{1}{2}g'-\frac{1}{2}}.n^{h'+g'-1}\right),$$

whereas for even values of g' it is

$$\mathcal{O}\left(r^{1-\frac{1}{2}g'}.n^{h'+g'-1}\right).$$

Furthermore the summation-indices in (3.2.3) satisfy the following inequalities:

$$1 - 2g - 2g' \geq 2h \geq 2h' - 2\left[\frac{g}{2}\right]$$

hence $$1 \geq 2h' + g + 2g' .$$

We start by considering sums in (3.2.3) corresponding to odd values of g'. According to the above they are

$$\mathcal{O}\left(r^{2h'+g+\frac{3}{2}g'-\frac{1}{2}}.n^{-h'-g-g'}\right).$$

As $\frac{r^2}{n} \to \infty$ and $h' \leq \left[\frac{k-g'}{2}\right] \leq \frac{k-g'}{2}$, they are

$$\mathcal{O}\left(r^{k+g+\frac{1}{2}g'-\frac{1}{2}} . n^{-\frac{1}{2}k-g-\frac{1}{2}g'} \right) ,$$

or, as $g' \geq 1$ and $g \geq 0$

$$\mathcal{O}\left(r^{k} . n^{-\frac{1}{2}k-\frac{1}{2}} \right) = \mathcal{O}\left(n^{-\frac{1}{2}} . \sigma^{k} \right) = o(\sigma^{k}) .$$

The sums corresponding to even values of g' are

$$\mathcal{O}\left(r^{2h'+g+\frac{3}{2}g'} . n^{-h'-g-g'} \right) .$$

Among these, sums having $h' \leq \frac{k-g'-1}{2}$ are

$$\mathcal{O}\left(r^{k+g+\frac{1}{2}g'-1} . n^{-\frac{1}{2}k-g-\frac{1}{2}g'+\frac{1}{2}} \right) ,$$

or, as $g' \geq 0$ and $g \geq 0$

$$\mathcal{O}\left(r^{k-1} . n^{-\frac{1}{2}k+\frac{1}{2}} \right) = \mathcal{O}\left(r^{-1} . n^{+\frac{1}{2}} . \sigma^{k} \right) = o(\sigma^{k}) .$$

The remaining sums for g' even have $h' = \left[\frac{k-g'}{2}\right] = \frac{k-g'}{2}$ which implies that k is even. However, the summation over h in (3.2.3) is then

$$\frac{k-g'}{2} - \left[\frac{g}{2}\right] \leq h \leq \left[\frac{k-g-g'}{2}\right] ,$$

which is empty if g is odd, and implies $h = \frac{k-g-g'}{2}$ if g is even. We note that for these sums the contribution of the remainder term in (4.1.3) is

$$\mathcal{O}\left(r^{2h'+g+\frac{3}{2}g'-1} . n^{-h-g-g'} \right) = \mathcal{O}\left(r^{k-1} . n^{-\frac{1}{2}k} \right) = \mathcal{O}\left(r^{-1} . \sigma^{k} \right) = o(\sigma^{k}) ,$$

and that, consequently, we may restrict our attention to the contribution of the leading term in (4.1.3).

So far we have shown that for odd values of k all (finitely many) terms in (3.2.3) are $o(\sigma^k)$ or

(4.1.6) $$\lim \mu_k \sigma^{-k} = 0 \qquad \text{for k odd.}$$

For even values of k only those terms remain to be considered having g' even ; g even ; $h' = \frac{k-g'}{2}$ and $h = \frac{k-g-g'}{2}$.

As $k-g' = \sum_{i=1}^{h'} k_i'$, $k_i' \geq 2$, it follows that $k_i' = 2$, $i = 1,2,\ldots,\frac{k-g'}{2}$ in (3.2.3); also $k-g-g' = \sum_{i=1}^{h} k_i$, $k_i \geq 2$, yields $k_i = 2$, $i = 1,2,\ldots,\frac{k-g-g'}{2}$.

Hence the configuration $\sum_{i=1}^{h'} C_{k_i',l_i'}^{(\alpha_i')}$ in (3.2.3), with $\sum_{i=1}^{h'} k_i' = k-g'$, $\sum_{i=1}^{h'} l_i' = l'-2g'$, can be written

$$(\tfrac{3}{2}k+\tfrac{1}{2}g'-l')C_{2,2} + (l'-k-g')C_{2,3} \quad ,$$

whereas the configuration $\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)}$, with $\sum_{i=1}^{h} k_i = k-g-g'$, $\sum_{i=1}^{h} l_i = l-2g-2g'$, is found to be

$$(\tfrac{3}{2}k+\tfrac{1}{2}g+\tfrac{1}{2}g'-l)C_{2,2} + (l-k-g-g')C_{2,3} \quad .$$

As by (4.1.1)

$$\sum \left\{ m^{(1)}\ldots m^{(k)} \mid g'C_{1,2} + (\tfrac{3}{2}k+\tfrac{1}{2}g'-l')C_{2,2} + (l'-k-g')C_{2,3}\right\} =$$
$$= n^{\frac{1}{2}k+\frac{1}{2}g'}.m_1(n)^{g'}.m_2(n)^{\frac{3}{2}k+\frac{1}{2}g'-l'}.m_3(n)^{l'-k-g'} \quad ,$$

we obtain from (3.2.3) and (4.1.3) (terms with $l > \frac{3}{2}k+\frac{1}{2}g+\frac{1}{2}g'$ vanish)

$$(4.1.7) \quad \mu_k = \sum_{g'=0}^{k}{}^{*} \sum_{l'=k+g'}^{\frac{3}{2}k+\frac{1}{2}g'} \binom{k}{g'} n^{\frac{1}{2}k+\frac{1}{2}g'}.m_1(n)^{g'}.m_2(n)^{\frac{3}{2}k+\frac{1}{2}g'-l'}.m_3(n)^{l'-k-g'}.$$
$$.\sum_{g=0}^{k-g'}{}^{*} \sum_{l=k+g+g'}^{\frac{3}{2}k+\frac{1}{2}g+\frac{1}{2}g'} \mathcal{N}\left((\tfrac{3}{2}k+\tfrac{1}{2}g+\tfrac{1}{2}g'-l)C_{2,2} + (l-k-g-g')C_{2,3} + gC_{1,2}\right).$$
$$.\mathcal{A}\left((\tfrac{3}{2}k+\tfrac{1}{2}g+\tfrac{1}{2}g'-l)C_{2,2}+(l-k-g-g')C_{2,3}+gC_{1,2} \,;\, (\tfrac{3}{2}k+\tfrac{1}{2}g'-l')C_{2,2}+(l'-k-g')C_{2,3}\right).$$
$$.\frac{g'!}{(\frac{1}{2}g')!}\left(\frac{r}{n}\right)^{l}\left(-2.\frac{n-r}{rn}\right)^{\frac{1}{2}g'} + \sigma(\sigma^k) \quad ,$$

where $\sum^{*}$ denotes a summation over even values only.

By theorem 2.1.1 and (2.1.7)

$$\mathcal{N}\left((\tfrac{3}{2}k+\tfrac{1}{2}g+\tfrac{1}{2}g'-l)C_{2,2} + (l-k-g-g')C_{2,3} + gC_{1,2}\right) =$$

$$= \frac{(k-g')!\ 2^{l-k-g-g'}}{g!(\tfrac{3}{2}k+\tfrac{1}{2}g+\tfrac{1}{2}g'-l)!(l-k-g-g')!} \quad ,$$

and by lemma (2.3.3) (terms with $l' > l - \frac{1}{2}g$ vanish)

$$\sum_{l'=k+g'}^{\frac{3}{2}k+\frac{1}{2}g'} \mathcal{A}\left((\tfrac{3}{2}k+\tfrac{1}{2}g+\tfrac{1}{2}g'-l)C_{2,2} + (l-k-g-g')C_{2,3} + gC_{1,2} ;\right.$$

$$\left.(\tfrac{3}{2}k+\tfrac{1}{2}g'-l')C_{2,2} + (l'-k-g')C_{2,3}\right).\ m_2(n)^{\frac{3}{2}k+\frac{1}{2}g'-l'} . m_3(n)^{l'-k-g'} =$$

$$= (-1)^{\frac{1}{2}g} \frac{g!}{(\frac{1}{2}g)!} m_2(n)^{\frac{3}{2}k+\frac{1}{2}g+\frac{1}{2}g'-l}\left(m_3(n)-m_2(n)\right)^{l-k-g-g'}\left(2m_3(n)-m_2(n)\right)^{\frac{1}{2}g} .$$

Substitution in (4.1.7) gives

$$(4.1.8)\quad \mu_k = \sum_{g'=0}^{k}{}^{*} \sum_{g=0}^{k-g'}{}^{*} \sum_{l=k+g+g'}^{\frac{3}{2}k+\frac{1}{2}g+\frac{1}{2}g'} \frac{k!}{(\frac{1}{2}g')!(\frac{1}{2}g)!(\frac{3}{2}k+\frac{1}{2}g+\frac{1}{2}g'-l)!(l-k-g-g')!} .$$

$$.n^{\frac{1}{2}k+\frac{1}{2}g'} m_1(n)^{g'} m_2(n)^{\frac{3}{2}k+\frac{1}{2}g+\frac{1}{2}g'-l}\left(2m_3(n)-2m_2(n)\right)^{l-k-g-g'}\left(m_2(n)-2m_3(n)\right)^{\frac{1}{2}g} .$$

$$. \left(\frac{r}{n}\right)^{l} \left(-2.\frac{n-r}{rn}\right)^{\frac{1}{2}g'} + \mathcal{O}(\sigma^k) .$$

Summing over l and setting $a = \frac{1}{2}g'$, $b = \frac{1}{2}g$ we have

$$\mu_k = \sum_{a=0}^{\frac{1}{2}k} \sum_{b=0}^{\frac{1}{2}k-a} \frac{k!}{a!b!(\frac{1}{2}k-a-b)!} n^{\frac{1}{2}k}\left(-2m_1^2(n).\frac{n-r}{n}\right)^{a}\left(m_2(n)-2m_3(n)\right)^{b}\left(\frac{r}{n}\right)^{k+a+2b} .$$

$$. \left\{ m_2(n) + 2\frac{r}{n}\left(m_3(n)-m_2(n)\right)\right\}^{\frac{1}{2}k-a-b} + \mathcal{O}(\sigma^k) =$$

$$= \sum_{a=0}^{\frac{1}{2}k} \frac{k!}{a!(\frac{1}{2}k-a)!}\, n^{\frac{1}{2}k}\left(-2m_1^2(n).\frac{n-r}{n}\right)^a.\left(\frac{r}{n}\right)^{k+a} .$$

$$.\left\{ m_2(n)\left(\frac{n-r}{n}\right)^2 + 2m_3(n)\frac{r(n-r)}{n^2}\right\}^{\frac{1}{2}k-a} + o(\sigma^k) =$$

$$= \frac{k!}{(\frac{1}{2}k)!}\, n^{\frac{1}{2}k}.\left(\frac{r}{n}\right)^k \left\{ 2\left(m_3(n)-m_1^2(n)\right)\frac{r(n-r)}{n^2} + m_2(n)\left(\frac{n-r}{n}\right)^2\right\}^{\frac{1}{2}k} + o(\sigma^k) =$$

$$= \frac{k!}{(\frac{1}{2}k)!}\, 2^{-\frac{1}{2}k}\left[2\frac{r^2(n-r)}{n^3}\left\{(n-r)m_2(n) + 2r\left(m_3(n)-m_1^2(n)\right)\right\}\right]^{\frac{1}{2}k} + o(\sigma^k) =$$

$$= \frac{k!}{(\frac{1}{2}k)!}\, 2^{-\frac{1}{2}k}\, \sigma^k + o(\sigma^k) ,$$

or

(4.1.9) $$\lim \mu_k\sigma^{-k} = \frac{k!}{(\frac{1}{2}k)!}\, 2^{-\frac{1}{2}k} \qquad \text{for } k \text{ even.}$$

Thus, the moments of $(\underline{x}-E\underline{x})\sigma^{-1}$ tend to the moments of the standard normal distribution. As these moments determine the distribution uniquely (cf. M.G. KENDALL and A. STUART (1958), 111) the result of theorem 4.1.1 is proved (loc.cit.,115).

The proof of theorem 4.1.2 will be seen to be closely analogous to the foregoing proof. As the random variable $\underline{y}$ is not affected by interchanging the sample sizes r_1 and r_2 we may assume without loss of generality that $r_1 \leq r_2$. In the appendix it is shown that in this case

(4.1.10)

$$A = \sum_{\nu=0}^{g'}(-1)^\nu\binom{g'}{\nu}\left\{E\left(\underline{y}^{(1)}\,|\,C_{1,2}\right)\right\}^\nu E\left(\underline{y}^{(1)}\cdots\underline{y}^{(k-\nu)}\,\Big|\,\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)} + (g'+g-\nu)C_{1,2}\right) =$$

$$= \mathcal{B}\left(\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)}\right)\left[\frac{g'!}{(\frac{1}{2}g')!}\left(\frac{2r_1r_2}{n^2}\right)^{g+g'} \prod_{i=1}^{h}\left\{\left(\frac{r_1}{n}\right)^{l_i-\tau_i}\left(\frac{r_2}{n}\right)^{\tau_i} + \left(\frac{r_1}{n}\right)^{\tau_i}\left(\frac{r_2}{n}\right)^{l_i-\tau_i}\right\}.\right.$$

$$\left. .\left\{-\frac{1}{2}\left(\frac{1}{r_1}+\frac{1}{r_2}-\frac{4}{n}\right)\right\}^{\frac{1}{2}g'} + \mathcal{O}\left(r_1^{\,h+g+\frac{1}{2}g'-1}.r_2^{\,1-g-g'-h}.n^{-1}\right)\right] =$$

$$= \mathcal{B}\left(\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)}\right). \mathcal{O}\left(r_1^{h+g+\frac{1}{2}g'} . r_2^{l-g-g'-h} . n^{-1}\right) \quad \text{if } g' \text{ is even,}$$

and

$$(4.1.11) \quad A = \mathcal{B}\left(\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)}\right). \mathcal{O}\left(r_1^{h+g+\frac{1}{2}g'-\frac{1}{2}} . r_2^{l-g-g'-h} . n^{-1}\right) \quad \text{if } g' \text{ is odd.}$$

From (3.5.7) we find

$$(4.1.12) \quad \sigma^2(\underline{y}) =$$

$$= \frac{4r_1r_2}{n^3}\left[\left\{n(n-r_1-r_2)+2r_1r_2\right\}m_2(n) + \left\{n(r_1+r_2)-4r_1r_2\right\}\left\{m_3(n)-m_1^2(n)\right\}\right] +$$

$$+ \mathcal{O}\left(\frac{r_1r_2}{n^2}\right) = \mathcal{O}\left(\frac{r_1r_2}{n}\right) \quad ;$$

from $\frac{r_1}{n} \leq 1-\varepsilon$ and $\frac{r_2}{n} \leq 1-\varepsilon$ it follows that either $n(n-r_1-r_2)$ or $2r_1r_2$ is not $\sigma(n^2)$, hence $\left\{n(n-r_1-r_2)+2r_1r_2\right\}$ is not $\sigma(n^2)$. Since $m_2(n) \geq 1$ and both terms inside the square brackets are non-negative, $\sigma^2(\underline{y})$ cannot be $\left(\frac{r_1r_2}{n}\right)$.

As we did in the proof of theorem 4.1.1, we start from (3.2.3) and evaluate the order of magnitude of the contribution to μ_k for each value of g'. For odd values of g' this contribution is by lemma 2.3.4 and (4.1.11)

$$\mathcal{O}\left(r_1^{h+g+\frac{1}{2}g'-\frac{1}{2}} . r_2^{l-g-g'-h} . n^{h'+g'-l}\right),$$

whereas for even values of g' it is

$$\mathcal{O}\left(r_1^{h+g+\frac{1}{2}g'} . r_2^{l-g-g'-h} . n^{h'+g'-l}\right) .$$

Making use of the same inequalities for the summation-indices as we did in the foregoing proof, we start with sums in (3.2.3) corresponding to odd values of g'. As $h \geq h' - \left[\frac{1}{2}g\right] \geq h' - \frac{1}{2}g$ and $\frac{r_1}{r_2} \leq 1$ these sums are

$$\mathcal{O}\left(r_1^{h'+\frac{1}{2}g+\frac{1}{2}g'-\frac{1}{2}} . r_2^{l-h'-\frac{1}{2}g-g'} . n^{h'+g'-l}\right) \quad ;$$

from $l \geq 2h'+g+2g'$ and $\frac{r_2}{n} < 1$ we find that they are

$$\mathcal{O}\left(r_1^{h'+\frac{1}{2}g+\frac{1}{2}g'-\frac{1}{2}} . r_2^{h'+\frac{1}{2}g+g'} . n^{-h'-g-g'} \right) ,$$

or, as $h' \leq \left[\frac{k-g'}{2}\right] \leq \frac{k-g'}{2}$ and $\frac{r_1 r_2}{n} \to \infty$,

$$\mathcal{O}\left(r_1^{\frac{1}{2}k+\frac{1}{2}g-\frac{1}{2}} . r_2^{\frac{1}{2}k+\frac{1}{2}g+\frac{1}{2}g'} . n^{-\frac{1}{2}k-g-\frac{1}{2}g'} \right) ;$$

since $g' \geq 1$ and $g \geq 0$, they are

$$\mathcal{O}\left(r_1^{\frac{1}{2}k-\frac{1}{2}} . r_2^{\frac{1}{2}k+\frac{1}{2}} . n^{-\frac{1}{2}k-\frac{1}{2}} \right) = \mathcal{O}\left(r_1^{-\frac{1}{2}} . r_2^{\frac{1}{2}} . n^{-\frac{1}{2}} . \sigma^k \right) = o(\sigma^k) .$$

For even values of g' and $h' \leq \frac{k-g'-1}{2}$ we find that the sums are

$$\mathcal{O}\left(r_1^{\frac{1}{2}k+\frac{1}{2}g-\frac{1}{2}} . r_2^{\frac{1}{2}k+\frac{1}{2}g+\frac{1}{2}g'-\frac{1}{2}} . n^{-\frac{1}{2}k-g-\frac{1}{2}g'+\frac{1}{2}} \right) ,$$

or, as $g' \geq 0$ and $g \geq 0$,

$$\mathcal{O}\left(r_1^{\frac{1}{2}k-\frac{1}{2}} . r_2^{\frac{1}{2}k-\frac{1}{2}} . n^{-\frac{1}{2}k+\frac{1}{2}} \right) = \mathcal{O}\left(r_1^{-\frac{1}{2}} . r_2^{-\frac{1}{2}} . n^{\frac{1}{2}} . \sigma^k \right) = o(\sigma^k) .$$

The remaining sums for even values of g' have $h' = \frac{k-g'}{2}$ which implies that k is even. As was shown in the proof of theorem 4.1.1 this also means that g is even and $h = \frac{k-g-g'}{2}$. For these sums the contribution of the remainder term in (4.1.10) is easily seen to be

$$\mathcal{O}\left(r_1^{-1} . \sigma^k \right) = o(\sigma^k) ;$$

as a result we may confine our attention to the contribution of the first term of (4.1.10).

Thus we have shown that for odd values of k

$$\lim \mu_k \sigma^{-k} = 0 , \qquad (k \text{ odd}) ; \tag{4.1.13}$$

for even values of k we may follow the proof of theorem 4.1.1 from (4.1.6) onwards if we replace the leading term of (4.1.3) by the first term of (4.1.10).

Since $\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)} = (\frac{3}{2}k+\frac{1}{2}g+\frac{1}{2}g'-1)C_{2,2} + (1-k-g-g')C_{2,3}$ is bichromatic and has

$$\prod_{i=1}^{h}\left\{\left(\frac{r_1}{n}\right)^{l_i-\tau_i}\left(\frac{r_2}{n}\right)^{\tau_i} + \left(\frac{r_1}{n}\right)^{\tau_i}\left(\frac{r_2}{n}\right)^{l_i-\tau_i}\right\} =$$

$$= \left(\frac{2r_1r_2}{n^2}\right)^{\frac{3}{2}k+\frac{1}{2}g+\frac{1}{2}g'-1} \cdot \left(\frac{r_1^2r_2+r_1r_2^2}{n^3}\right)^{1-k-g-g'} ,$$

this means replacing

$$\left(\frac{r}{n}\right)^{l} \left(-2\ \frac{n-r}{rn}\right)^{\frac{1}{2}g'}$$

by

$$\left(\frac{2r_1r_2}{n^2}\right)^{\frac{3}{2}k+\frac{3}{2}g+\frac{3}{2}g'-1} \left(\frac{r_1^2r_2+r_1r_2^2}{n^3}\right)^{1-k-g-g'} \left\{-\frac{1}{2}\left(\frac{1}{r_1}+\frac{1}{r_2}-\frac{4}{n}\right)\right\}^{\frac{1}{2}g'}$$

in (4.1.8). Following the proof of theorem (4.1.1) to the end we then obtain

$$\mu_k = \frac{k!}{(\frac{1}{2}k)!}\, 2^{-\frac{1}{2}k} \left[\frac{4r_1r_2}{n^3}\left[\left\{n(n-r_1-r_2) + 2r_1r_2\right\}m_2(n) + \right.\right.$$

$$\left.\left. + \left\{n(r_1+r_2) - 4r_1r_2\right\}\left\{m_3(n) - m_1^2(n)\right\}\right]\right]^{\frac{1}{2}k} + \sigma(\sigma^k) =$$

$$= \frac{k!}{(\frac{1}{2}k)!}\, 2^{-\frac{1}{2}k}\, \sigma^k + \sigma(\sigma^k) ,$$

or

(4.1.14) $$\lim \mu_k \sigma^{-k} = \frac{k!}{(\frac{1}{2}k)!}\, 2^{-\frac{1}{2}k} \qquad (k \text{ even}) .$$

This completes the proof of theorem 4.1.2.

4.2. Tendency towards the normal distribution (free sampling)

Theorem 4.2.1.

i If n tends to infinity and p_1 varies in such a way that

$$np_1^2 \longrightarrow \infty \quad \text{and} \quad p_1 \leqq 1 - \varepsilon$$

for some $\varepsilon > 0$, and if for all i

$$m_{i+} \leqq c \quad ,$$

where c is a constant independent of i and n, then in the case of free sampling 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 $\sigma^2(\underline{x}_B)$ are given by (3.4.3) and (3.4.4) with p replaced by p_1.

ii If in part i p_1 is replaced by p_2, and $\underline{x}_B$ by $\underline{x}_W$, the corresponding result for $\underline{x}_W$ is obtained.

Theorem 4.2.2.

If n tends to infinity and p_1 and p_2 vary in such a way that

$$np_1p_2 \longrightarrow \infty \quad \text{and} \quad p_1 \leqq 1 - \varepsilon \ , \quad p_2 \leqq 1 - \varepsilon$$

for some $\varepsilon > 0$, and if for all i

$$m_{i+} \leqq c \quad ,$$

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

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

tends to the standard normal one. $E\underline{y}$ and $\sigma^2(\underline{y})$ are given by (3.6.2) and (3.6.3).

We omit the proofs of these theorems, as they are closely analogous (and simpler) then those of the corresponding theorems of section 4.1.

4.3. Tendency towards the compound Poisson-distribution

Theorem 4.3.1.

i If, as r_1 and n tend to infinity,

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

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

and for all α and $k \geq 2$

(4.3.1) $$\lim \left(\frac{r_1}{n}\right)^{k+1} \sum \left\{ m^{(1)} \dots m^{(k)} \mid C_{k,k+1}^{(\alpha)} \right\} = 0 \ ,$$

and if for all i and j

$$m_{ij} \leq c_1 \ ,$$

where c_1 is a constant independent of i, j and n, then in the non free sampling case the distribution of $\frac{1}{2}\underline{x}_B$ tends to a compound Poisson-distribution with moment-generating function

$$\sum_{k=0}^{\infty} \frac{z^k}{k!} \lim E\left(\frac{1}{2}\underline{x}_B\right)^k = \exp\left\{ \lambda \sum_{h=1}^{\infty} m_h^* \frac{z^h}{h!} \right\} \ .$$

Assumption (4.3.1) is satisfied e.g. 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 does not depend on n.

ii The corresponding result for the free sampling case is obtained if $\frac{r_1}{n}$ is replaced by p_1 in part i.

iii By replacing $\underline{x}_B$ and r_1 or p_1 by $\underline{x}_W$ and r_2 or p_2, the corresponding results for $\underline{x}_W$ are obtained.

Theorem 4.3.2.

i If, as r_1, r_2 and n tend to infinity,

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

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

and for all α, $k \geq 2$ and $\nu=1,2$

$$\text{(4.3.2)} \qquad \lim \left(\frac{r_\nu}{n}\right)^{k+1} \sum \left\{ m^{(1)} \ldots m^{(k)} \mid C_{k,k+1}^{(\alpha)} \right\} = 0 \quad ,$$

and if for all i and j

$$m_{ij} \leq c_1 \quad ,$$

where c_1 is a constant independent of i, j and n, then in the non free sampling case the distribution of $\frac{1}{2}\underline{y}$ tends to the compound Poisson-distribution with moment-generating function

$$\sum_{k=0}^{\infty} \frac{z^k}{k!} \lim E\left(\tfrac{1}{2}\underline{y}\right)^k = \exp\left\{ \lambda \sum_{h=1}^{\infty} m_h^* \frac{z^h}{h!} \right\} .$$

Assumption (4.3.2) is satisfied e.g. if for all i

$$m_{i+} \leq c_2 \quad ,$$

where c_2 does not depend on i and

$$\lim \left(\frac{r_\nu}{n}\right) c_2 = 0 \quad , \qquad \nu=1,2 \quad ,$$

thus e.g. when c_2 does not depend on n.

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

We note that as $m_{++} \geq n$ by (1.1.3) and (1.1.4) (cf. 4.1.1), the condition $\lim\left(\frac{r_1}{n}\right)^2 m_{++} = 2\lambda$, $0 < \lambda < \infty$, of theorem 4.3.1 implies that $\lim \frac{r_1}{n} = 0$. Likewise, in theorem 4.3.2 the condition $\lim \frac{r_1 r_2}{n^2} m_{++} = \lambda$, $0 < \lambda < \infty$, as r_1, r_2 and n tend to infinity, ensures that $\lim \frac{r_1}{n} = \lim \frac{r_2}{n} = 0$.

To prove theorem 4.3.1 we shall first establish four lemmata.

Lemma 4.3.1

If the assumptions of theorem 4.3.1, part i are satisfied, then for $l > 2$, all α, and all $k(\geq l-1)$,

$$\lim \frac{r_1^l}{n^l} \sum \left\{ m^{(1)} \ldots m^{(k)} \mid C_{k,l}^{(\alpha)} \right\} = 0 \quad .$$

Proof

If $l = k+1$ the lemma is true by assumption. If $l < k+1$, a graph having configuration $C_{k,l}^{(\alpha)}$ has $k-l+1$ independent circuits. One can choose in this case $k-l+1$ joins, such that if they are taken away, a tree remains (cf. D. KOENIG (1936), 53). Consider now

$$m^{(1)} \ldots m^{(k)} \quad ,$$

where the corresponding graph has configuration $C_{k,l}^{(\alpha)}$. Let the $k-l+1$ joins that can be taken away correspond to $m^{(l)},\ldots,m^{(k)}$. Then

$$0 \leq m^{(1)} \ldots m^{(k)} \leq c_1^{k-l+1}\, m^{(1)} \ldots m^{(l-1)} \quad .$$

Therefore, summing over the l summation subscripts and multiplying by $\left(\frac{r_1}{n}\right)^l$ gives

$$0 \leq \frac{r_1^l}{n^l} \sum \left\{ m^{(1)} \ldots m^{(k)} \mid C_{k,l}^{(\alpha)} \right\} \leq c_1^{k-l+1} \frac{r_1^l}{n^l} \sum \left\{ m^{(1)} \ldots m^{(l-1)} \mid C_{l-1,l}^{(\beta)} \right\},$$

where the right hand side tends to zero by (4.3.1). This proves lemma 4.3.1.

Lemma 4.3.2

If the assumptions of theorem 4.3.1, part i are satisfied, then for all configurations $\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)}$, with $\sum_{i=1}^{h} k_i = k$, $\sum_{i=1}^{h} l_i = l$, for which at least one $l_i > 2$ $(i=1,2,\ldots,h)$

$$\lim \frac{r_1^l}{n^l} \sum \left\{ m^{(1)} \ldots m^{(k)} \mid \sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)} \right\} = 0 \quad .$$

Proof

$$(4.3.3)\qquad \frac{r_1^l}{n^l}\sum\left\{m^{(1)}\ldots m^{(k)}\mid \sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)}\right\} =$$

$$= \prod_{i=1}^{h}\frac{r_1^{l_i}}{n^{l_i}}\sum\left\{m^{(k_0+\ldots+k_{i-1}+1)}\ldots m^{(k_1+\ldots+k_i)}\mid C_{k_i,l_i}^{(\alpha_i)}\right\}$$

with $k_0=0$.

For every i such that $l_i=2$ (4.3.3) contains a factor

$$(4.3.4)\qquad \frac{r_1^2}{n^2}\sum_{ij} m_{ij}^{k_i} = \frac{r_1^2}{n^2} m_{++}\cdot\frac{\sum_{ij} m_{ij}^{k_i}}{m_{++}} \approx 2\lambda m_{k_i}^{*} \qquad ;$$

for every i such that $l_i>2$ (at least one such i exists) (4.3.3) contains a factor which for $n\to\infty$ tends to zero by lemma 4.3.1. This proves lemma 4.3.2.

Lemma 4.3.3

For k=1,2,... , and l=2,4,...,2k

$$\lim\left(\frac{r_1}{n}\right)^l\sum\left\{m^{(1)}\ldots m^{(k)}\mid\sum_{i=1}^{s} g_i C_{i,2}\right\} = 2^{\frac{l}{2}}\lambda^{\frac{l}{2}}\prod_{i=1}^{s}(m_i^{*})^{g_i} \qquad ,$$

where $2\sum_{i=1}^{s} g_i = l$, and $\sum_{i=1}^{s} i g_i = k$.

Proof

Evident by (4.3.3) and (4.3.4).

Lemma 4.3.4

If the assumptions of theorem 4.3.1, part i hold, then for every configuration $\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)}$, with $\sum_{i=1}^{h} k_i = k$, $\sum_{i=1}^{h} l_i = l$,

$$\lim\left(\frac{r_1}{n}\right)^l\sum\left\{m^{(1)}\ldots m^{(k)}\mid\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)}\right\} = \lim\left(\frac{r_1}{n}\right)^l\sum^{*}\left\{m^{(1)}\ldots m^{(k)}\mid\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)}\right\} .$$

Proof

The difference

$$\left(\frac{r_1}{n}\right)^l \left[\sum \left\{ m^{(1)} \ldots m^{(k)} \mid \sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)} \right\} - \sum^{*} \left\{ m^{(1)} \ldots m^{(k)} \mid \sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)} \right\} \right]$$

consists of sums of the type

$$\left(\frac{r_1}{n}\right)^l \sum \left\{ m^{(1)} \ldots m^{(k)} \mid \sum_{i=1}^{h'} C_{k_i',l_i'}^{(\alpha_i')} \right\} ,$$

where $\sum_{i=1}^{h'} l_i' = l' < l$. As $\frac{r_1}{n}$ tends to zero, each of these contributions tends to zero by lemmata 4.3.2 and 4.3.3. This proves lemma 4.3.4.

Now we have by (3.1.6), writing $\underline{x}$ for $\underline{x}_B$, and r for r_1,

$$\sum_{k=1}^{\infty} \frac{z^k}{k!} \lim E\left(\tfrac{1}{2}\underline{x}\right)^k = \sum_{k=1}^{\infty} 2^{-k} \frac{z^k}{k!} \cdot \sum_{l=2}^{2k} \sum_{h=1}^{\left[\frac{l}{2}\right]} \sum{}'' \left\lfloor \sum_{i=1}^{h} k_i = k,\ \sum_{i=1}^{h} l_i = l \right\rfloor$$

$$\mathcal{N}\left(\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)}\right) \lim E\left(\underline{x}^{(1)} \ldots \underline{x}^{(k)} \mid \sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)}\right) \sum^{*} \left\{ m^{(1)} \ldots m^{(k)} \mid \sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)} \right\} =$$

by (3.3.2)

$$= \sum_{k=1}^{\infty} 2^{-k} \frac{z^k}{k!} \cdot \sum_{l=2}^{2k} \sum_{h=1}^{\left[\frac{l}{2}\right]} \sum{}'' \left\lfloor \sum_{i=1}^{h} k_i = k,\ \sum_{i=1}^{h} l_i = l \right\rfloor \mathcal{N}\left(\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)}\right) .$$

$$\lim \frac{r^{!l}}{n^{!l}} \sum^{*} \left\{ m^{(1)} \ldots m^{(k)} \mid \sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)} \right\} =$$

by lemma 4.3.4

$$= \sum_{k=1}^{\infty} 2^{-k} \frac{z^k}{k!} \cdot \sum_{l=2}^{2k} \sum_{h=1}^{\left[\frac{l}{2}\right]} \sum{}'' \left\lfloor \sum_{i=1}^{h} k_i = k,\ \sum_{i=1}^{h} l_i = l \right\rfloor \mathcal{N}\left(\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)}\right) .$$

$$\lim \left(\frac{r}{n}\right)^l \sum \left\{ m^{(1)} \ldots m^{(k)} \mid \sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)} \right\} =$$

by lemmata 4.3.2 and 4.3.3

$$= \sum_{k=1}^{\infty} \frac{z^k}{k!} 2^{-k} \sum_{\substack{l=2 \\ l \text{ even}}}^{2k} \sum_{s=1}^{k} \sum \left\lfloor \sum_{i=1}^{s} g_i = \frac{l}{2},\ \sum_{i=1}^{s} i g_i = k \right\rfloor \mathcal{N}\left(\sum_{i=1}^{s} g_i C_{i,2}\right) 2^{\frac{l}{2}} \lambda^{\frac{l}{2}} \prod_{i=1}^{s} (m_i^*)^{g_i} =$$

by Theorem 2.1.1 and (2.1.9) and by changing $\frac{1}{2}$ into 1 :

$$= \sum_{k=1}^{\infty} z^k \sum_{l=1}^{k} \sum_{s=1}^{k} \sum_{\left[\sum_{i=1}^{s} g_i = l,\ \sum_{i=1}^{s} i g_i = k\right]} \lambda^l \prod_{i=1}^{s} \frac{1}{g_i!} \left\{\frac{m_i^*}{i!}\right\}^{g_i} =$$

$$= \sum_{l=1}^{\infty} \sum_{\left[\sum_{i=1}^{\infty} g_i = l\right]} \lambda^l \prod_{i=1}^{\infty} \frac{1}{g_i!} \left(\frac{m_i^* z^i}{i!}\right)^{g_i} =$$

$$= \sum_{l=1}^{\infty} \frac{1}{l!} \left(\lambda \sum_{i=1}^{\infty} \frac{m_i^* z^i}{i!}\right)^l = \exp\left\{\lambda \sum_{i=1}^{\infty} \frac{m_i^* z^i}{i!}\right\} - 1 .$$

Defining

$$E\left(\frac{1}{2}\underline{x}\right)^0 = 1 \qquad \text{for all } n,$$

the main result of theorem 4.3.1, part i is proved.

If $m_{i+} \leq c_2$, c_2 independent of i and $\lim \frac{r_1 c_2}{n} = 0$, then by (2.3.33)

$$\sum\left\{m^{(1)} \ldots m^{(k)} \mid C_{k,k+1}^{(\alpha)}\right\} \leq c_2^{k-1} \sum_{ij} m_{ij} .$$

Thus for $k \geq 2$

$$\frac{r_1^{k+1}}{n^{k+1}} \sum\left\{m^{(1)} \ldots m^{(k)} \mid C_{k,k+1}^{(\alpha)}\right\} \leq \left(\frac{r_1 c_2}{n}\right)^{k-1} \frac{r_1^2}{n^2} m_{++} \longrightarrow 0 .$$

If c_2 is also independent of n, $\lim \frac{r_1 c_2}{n} = 0$ since $\frac{r_1}{n}$ tends to zero.

We have proved part i of theorem 4.3.1. Part ii can be proved in the same way. In fact replacing $\frac{r_1}{n}$ by p_1 in the proof of part i, transforms it into a proof for part ii. Part iii follows by symmetry.

The proof of theorem 4.3.2 follows very closely the one of theorem 4.3.1, and is therefore omitted. We only point to the fact that

$$E\left\{y^{(1)} \ldots y^{(k)} \mid \sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)}\right\} \approx \mathcal{B}\left(\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)}\right) \prod_{i=1}^{h} \left\{\left(\frac{r_1}{n}\right)^{l_i-\tau_i} \left(\frac{r_2}{n}\right)^{\tau_i} + \left(\frac{r_1}{n}\right)^{\tau_i} \left(\frac{r_2}{n}\right)^{l_i-\tau_i}\right\} .$$

4.4. The degenerate case

Theorem 4.4.1.

If n tends to infinity and if

$$\frac{r_1^2}{n} \longrightarrow 0 ,$$

if moreover for all i

$$m_{i+} \leqq c ,$$

where c is a constant independent of i and n, then in the non free sampling case in the limit $\underline{x}_B = 0$ spr 0.

Theorem 4.4.2.

If n tends to infinity and if

$$\frac{r_1 r_2}{n} \longrightarrow 0 ,$$

if moreover for all i

$$m_{i+} \leqq c ,$$

where c is a constant independent of i and n, then in the non free sampling case in the limit $\underline{y} = 0$ spr 0.

Theorem 4.4.3.

If n tends to infinity and if

$$np_1^2 \longrightarrow 0 ,$$

if moreover for all i

$$m_{i+} \leqq c ,$$

where c is a constant independent of i and n, then in the free sampling case in the limit $\underline{x}_B = 0$ spr 0.

Theorem 4.4.4.

If n tends to infinity and if

$$np_1 p_2 \longrightarrow 0 ,$$

if moreover for all i

$$m_{i+} \leqq c ,$$

where c is a constant independent of i and n, then in the free sampling case in the limit $\underline{y} = 0$ spr 0.

We prove theorem 4.4.1. Using the abbreviations (4.1.1) it follows from (3.3.1) and (3.3.8) that

$$E\underline{x}_B \approx \frac{r_1(r_1-1)}{n} m_1(n) \quad \text{and} \quad \sigma^2(\underline{x}_B) \approx \frac{2r_1(r_1-1)}{n} m_2(n) \quad .$$

Since $m_1(n)$ and $m_2(n)$ remain bounded as n tends to infinity (cf. (4.1.1)) both $E\underline{x}_B$ and $\sigma^2(\underline{x}_B)$ tend to zero. The result of the theorem now follows from the BIENAYMÉ - CHEBYCHEV inequality. The proof of the other theorems proceeds in exactly the same way.

CHAPTER 5

5.1. A test for randomness

The theory of the preceding sections can be applied to a testing problem which often arises in relation to ecological and virological studies.

Consider e.g. an agricultural experiment carried out with the aim of studying the occurrence of a disease with a certain kind of crop. If the way in which the disease is transmitted is not known one may be able to obtain information about this question by studying the geographical distribution of the diseased plants among the healthy ones. If this distribution is a random one the conclusions drawn about the mechanism of the disease would be different from those for the case where the diseased plants tend to cluster. Let us consider the case where the plants are grown at the corners of a rectangular lattice. To test the hypothesis H_o that the r_1 diseased plants occur at random among the n plants one may proceed as follows. For every diseased plant one counts how many of its direct neighbours are diseased as well. Addition of these numbers gives the test statistic x_B. If H_o is true the value of the test statistic is an observation of a random variable $\underline{x}_B$, defined by (1.1.5), where

$$m_{ij} = \begin{cases} 1 & \text{if plants i and j are direct neighbours,} \\ 0 & \text{if not.} \end{cases}$$

Large values of the test statistic lead to rejection of H_o. Alternatively one can execute the test with a statistic y, corresponding to $\underline{y}$, defined by (1.1.5).
This method of testing randomness has been described by several authors, e.g. H. TODD (1940), P.A.P. MORAN (1948) and P.V. KRISHNA IYER.
A detailed example of an application is given by G.H. FREEMAN (1953).

Using the results obtained in the preceding sections the test procedures may be elaborated. If the inoculum of the disease is transmitted by e.g. insects a diseased plant influences not only its direct neighbours, but also - to a lesser extent - the other plants in the neighbourhood. In such cases the test might gain in power if the test statistic not only takes account of pairs of diseased plants, that are direct neighbours. In fact it is only natural to define the value of m_{ij} such that it is in some relation to the distance between plants i and j.

From the published examples of applications of tests of randomness it is apparent that the non free sampling case is the more important one. For not too small values of $E\underline{x}_B$ and $E\underline{y}$, the fact that under the hypothesis tested both random variables are approximately normally distributed can be used to determine approximate critical values.

In order to examine the consistency of the test based on $\underline{x}_B$ for the non free sampling case we define (cf. section 1.1)

$P(i)$ = probability that point i is black
$P(i,j)$ = probability that points i and j are both black,
etc.

Any hypothesis, i.e. a specification of a random mechanism that chooses r_1 points to be coloured black, determines a set of values for the probabilities $P(i)$, $P(i,j)$, ... In the non free sampling case H_0 implies

(5.1.1)
$$P(i) = \frac{r_1}{n}$$
$$P(i,j) = \frac{r_1(r_1-1)}{n(n-1)}, \text{ etc.,} \quad (i,j=1,2,\ldots,n,\ i\neq j).$$

We now prove

Theorem 5.1.1.

If r_1 and n tend to infinity in such a way that

$$\varepsilon \leq \frac{r_1}{n} \leq 1-\varepsilon$$

for some $\varepsilon > 0$, and if for all i

$$m_{i+} \leq c \quad ,$$

where c is a constant independent of i and n, then in the case of non free sampling the one-sided test based on the statistic $\underline{x}_B$ and a critical zone consisting of large values of x_B, is consistent for those alternative hypotheses satisfying

(5.1.2)
$$P(i) = \frac{r_1}{n} \quad , \qquad i = 1,2,\ldots,n \ ,$$

(5.1.3)
$$n^{-\frac{1}{2}} \sum_{ij} m_{ij}\{P(i|j) - P(i)\} \longrightarrow \infty \quad \text{for } n \longrightarrow \infty, \quad \text{and}$$

(5.1.4)
$$\sum_{ij} m_{ij}P(i,j) \sum_{kl\neq(i,j)} m_{kl}\{P(k,l|i,j) - P(k,l)\} = \mathcal{O}(n) \ .$$

Proof

Let $H' \neq H_o$ be a hypothesis satisfying (5.1.2), (5.1.3) and (5.1.4). Replace r_1 and $\underline{x}_B$ by r and $\underline{x}$ and denote $E(\underline{x} \mid H_o)$, $E(\underline{x} \mid H')$, $\sigma^2(\underline{x} \mid H_o)$ and $\sigma^2(\underline{x} \mid H')$ by μ, μ', σ^2 and σ'^2 respectively. The condition $m_{i+} \leq c$ implies (cf. section 4.1)

(5.1.5)
$$n \leq \sum_{ij} m_{ij}^2 \leq \sum_i m_{i+}^2 \leq c^2 n \quad ,$$
$$n \leq m_{++} \leq cn \quad ,$$

and (cf. 4.1.5)

(5.1.6)
$$\sigma = \mathcal{O}(n^{\frac{1}{2}}) \quad .$$

Now

(5.1.7)
$$\sigma'^2 = \sum_{(ijkl)\neq} m_{ij}m_{kl}\left\{P(i,j,k,l) - P(i,j)P(k,l)\right\} +$$
$$+ \; 4 \sum_{(ijk)\neq} m_{ij}m_{ik}\left\{P(i,j,k) - P(i,j)P(i,k)\right\} +$$
$$+ \; 2 \sum_{ij} m_{ij}^2 \, P(i,j) \left\{1 - P(i,j)\right\} = \mathcal{O}(n)$$

by (5.1.4) and (5.1.5). Also

(5.1.8)
$$\mu' - \mu = \sum_{ij} m_{ij}\left\{P(i,j) - \frac{r(r-1)}{n(n-1)}\right\} =$$
$$= \frac{r}{n} \sum_{ij} m_{ij}\left\{P(i|j) - \frac{r}{n}\right\} + \frac{r(n-r)}{n^2(n-1)} m_{++} =$$
$$= \frac{r}{n} \sum_{ij} m_{ij}\left\{P(i|j) - P(i)\right\} + \mathcal{O}(1) \quad ,$$

by (5.1.2) and (5.1.5). Hence for any fixed a,

(5.1.9)
$$\lim_{n\to\infty} \frac{\mu'-\mu-a\sigma}{\sigma'} = \infty \quad .$$

The probability of not rejecting H_o if H' is true is

$$P\left[\frac{\underline{x}-\mu}{\sigma} \leq a\right] = P\left[\frac{\underline{x}-\mu'}{\sigma'} \leq \frac{\mu-\mu'+a\sigma}{\sigma'}\right] \leq$$
$$\leq P\left[\left|\frac{\underline{x}-\mu'}{\sigma'}\right| \geq \frac{\mu'-\mu-a\sigma}{\sigma'}\right] \leq \left(\frac{\sigma'}{\mu'-\mu-a\sigma}\right)^2 \longrightarrow 0 \qquad \text{for } n \to \infty$$

by the BIENAYMÉ-CHEBYCHEV inequality and (5.1.9), which proves the theorem.

An analogous result may be proved for the test based on the statistic y and for the case of free sampling.

APPENDIX

In order to derive an asymptotic expression for

(A.1)
$$A = \sum_{\nu=0}^{g'} (-1)^{\nu} \binom{g'}{\nu} \left\{ E\left(\underline{z}^{(1)} \mid C_{1,2}\right)\right\}^{\nu} E\left(\underline{z}^{(1)} \ldots \underline{z}^{(k-\nu)} \mid \sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)} + (g'+g-\nu) C_{1,2}\right),$$

with $\sum_{i=1}^{h} k_i = k-g-g'$ and $\sum_{i=1}^{h} l_i = l-2g-2g'$, for the cases $\underline{z} = \underline{x}_B$ and $\underline{z} = \underline{y}$ and non free sampling, use is made of the expansion

$$\text{(A.2)} \quad \log \Gamma(x+a) \backsim (x+a-\tfrac{1}{2}) \log x - x + \tfrac{1}{2}\log 2\pi + \sum_{j=1}^{\infty} (-1)^{j+1} \frac{\Phi_{j+1}(a)}{j(j+1)x^j} ,$$

for $x \rightarrow \infty$. Here $\Phi_j(z)$ is the j-th BERNOULLI polynomial defined by

$$\frac{te^{tz}}{e^t-1} = \sum_{j=0}^{\infty} \Phi_j(z) \frac{t^j}{j!} \quad ;$$

$\Phi_j(z)$ is a polynomial of degree j in z, the coefficient of z^j being equal to 1. The first three are

$$\text{(A.3)} \qquad \begin{aligned} \Phi_0(z) &= 1 \\ \Phi_1(z) &= z - \tfrac{1}{2} \\ \Phi_2(z) &= z^2 - z + \tfrac{1}{6} \end{aligned}$$

(cf. E.T. WHITTAKER and G.N. WATSON (1915) chapter 13, where a slightly different version of these polynomials is discussed).

First we consider the case $\underline{z} = \underline{x}_B$. If r_1 and n tend to infinity in such a way that

$$\text{(A.4)} \qquad \frac{r_1(n-r_1)}{n} \longrightarrow \infty$$

we show that if g' is even

(A.5)
$$A = \frac{g'!}{(\frac{1}{2}g')!} \left(\frac{r_1}{n}\right)^{l} \left(\frac{-2(n-r_1)}{r_1 n}\right)^{\frac{1}{2}g'} + \mathcal{O}\left(\left(\frac{n-r_1}{r_1 n}\right)^{\frac{1}{2}g'-1} \cdot \frac{r_1^{l-2}}{n^l}\right) = \mathcal{O}\left(\left(\frac{n-r_1}{r_1 n}\right)^{\frac{1}{2}g'} \cdot \left(\frac{r_1}{n}\right)^{l}\right),$$

and if g' is odd

$$(A.6) \qquad A = \mathcal{O}\left(\left(\frac{n-r_1}{r_1 n}\right)^{\frac{1}{2}g'-\frac{1}{2}} \cdot \frac{r_1^{l-1}}{n^l}\right) \quad .$$

During the proof we shall write r instead of r_1.

By (3.3.2) and (A.2) we have in this case

$$A = \sum_{\nu=0}^{g'} (-1)^{\nu} \binom{g'}{\nu} \frac{\binom{r}{l-2\nu}\binom{r}{2}^{\nu}}{\binom{n}{l-2\nu}\binom{n}{2}^{\nu}} =$$

$$= \sum_{\nu=0}^{g'} (-1)^{\nu} \binom{g'}{\nu} \exp\Big[(\nu+1)\Big\{\log \Gamma(r+1) - \log \Gamma(n+1)\Big\} +$$

$$+ \nu\Big\{\log \Gamma(n-1) - \log \Gamma(r-1)\Big\} + \log \Gamma(n-l+2\nu+1) - \log \Gamma(r-l+2\nu+1)\Big] =$$

$$= \sum_{\nu=0}^{g'} (-1)^{\nu} \binom{g'}{\nu} \exp\Big[l \log \frac{r}{n} + \sum_{j=1}^{g'} \frac{(-1)^{j+1}}{j(j+1)} \Big\{(\nu+1)\phi_{j+1}(1) - \nu\phi_{j+1}(-1) +$$

$$- \phi_{j+1}(2\nu-l+1)\Big\}\left(\frac{1}{r^j} - \frac{1}{n^j}\right) + \mathcal{O}\left(\left(\frac{n-r}{rn}\right)^{\frac{1}{2}g'} . r^{-1}\right)\Big],$$

since $n^{-g'-1} \leq r^{-g'-1} \leq \left(\frac{n-r}{rn}\right)^{\frac{1}{2}g'} . r^{-1}$ by (A.4).

Defining

$$(A.7) \qquad (\nu+1)\phi_{j+1}(1) - \nu\phi_{j+1}(-1) - \phi_{j+1}(2\nu-l+1) = \sum_{s=0}^{j+1} a_{js}\nu^s ,$$

and

$$(A.8) \qquad a_s(r,n) = \sum_{j=\max(1,s-1)}^{g'} \frac{(-1)^{j+1}}{j(j+1)} a_{js}\left(\frac{1}{r^j} - \frac{1}{n^j}\right) ,$$

we have

$$A = \left(\frac{r}{n}\right)^l \sum_{\nu=0}^{g'} (-1)^{\nu}\binom{g'}{\nu} \exp\left[\sum_{s=0}^{g'+1} a_s(r,n)\,\nu^s\right] + \mathcal{O}\left(\left(\frac{n-r}{rn}\right)^{\frac{1}{2}g'} . \frac{r^{l-1}}{n^l}\right) .$$

We note that, because of (A.4),

$$(A.9) \qquad \frac{1}{r^j} - \frac{1}{n^j} = \mathcal{O}\left(\frac{n-r}{nr^j}\right) , \qquad \text{(it is not } o\left(\frac{n-r}{nr^j}\right)\text{)} ,$$

and hence

$$(A.10)\qquad \begin{aligned} a_o(r,n) &= \mathcal{O}\left(\frac{n-r}{nr}\right) ,\\ a_1(r,n) &= \mathcal{O}\left(\frac{n-r}{nr}\right) , \quad \text{and}\\ a_s(r,n) &= \mathcal{O}\left(\frac{n-r}{nr^{s-1}}\right) \quad \text{for } s \geq 2 .\end{aligned}$$

Expanding the exponentials in the last expression for A, we find

$$(A.11)\quad A = \left(\frac{r}{n}\right)^l \sum_{\nu=0}^{g'} (-1)^{\nu} \binom{g'}{\nu} \prod_{s=0}^{g'+1} \sum_{h_s=0}^{g'} \frac{1}{h_s!} \left(a_s(r,n)\nu^s\right)^{h_s} + \mathcal{O}\left(\left(\frac{n-r}{rn}\right)^{\frac{1}{2}g'} . \frac{r^{l-1}}{n^l}\right),$$

as $\left(a_s(r,n)\right)^{g'+1} = \mathcal{O}\left(\left(\frac{n-r}{rn}\right)^{g'+1}\right) = \mathcal{O}\left(\left(\frac{n-r}{rn}\right)^{\frac{1}{2}g'} . r^{-1}\right)$.

To every sequence of non-negative integers $h_o, h_1, \ldots, h_{g'+1}$, all less than $g'+1$, there corresponds a term T in (A.11)

$$T = \left(\frac{r}{n}\right)^l \sum_{\nu=0}^{g'} (-1)^{\nu} \binom{g'}{\nu} \nu^{\sum_{s=0}^{g'+1} sh_s} . \prod_{s=0}^{g'+1} \frac{1}{h_s!} \left(a_s(r,n)\right)^{h_s} =$$

$$= \left(\frac{r}{n}\right)^l \sum_{\nu=0}^{g'} (-1)^{\nu} \binom{g'}{\nu} \nu^{\sum_{s=0}^{g'+1} sh_s} . \mathcal{O}\left(\left(\frac{n-r}{n}\right)^{\sum_{s=0}^{g'+1} h_s} . r^{-(h_o+h_1+\sum_{s=2}^{g'+1} (s-1)h_s)}\right) ,$$

because of (A.10). Furthermore we note the identities

$$(A.12)\qquad \sum_{\nu=0}^{g'} (-1)^{\nu} \binom{g'}{\nu} \nu^p = \begin{cases} 0 & \text{if } p < g' \\ (-1)^{g'} . g'! & \text{if } p = g' \end{cases}$$

which are easily established by differentiating both members of

$$(1 - e^x)^{g'} = \sum_{\nu=0}^{g'} (-1)^{\nu} \binom{g'}{\nu} e^{\nu x}$$

p times with respect to x and setting $x = 0$. By (A.12) only those terms T having $\sum_{s=0}^{g'+1} sh_s \geq g'$ contribute to (A.11).

Suppose therefore that such a term T has $\sum_{s=0}^{g'+1} sh_s = p \geq g'$.

Then $2\sum_{s=2}^{g'+1} h_s \leq \sum_{s=2}^{g'+1} sh_s \leq p$, with equality if and only if p is even, $h_2 = \frac{1}{2}p$ and $h_s = 0$ for $s \neq 2$. Also

$$T = \mathcal{O}\left(\left(\frac{r}{n}\right)^l \cdot \left(\frac{n-r}{rn}\right)^{\frac{1}{2}p} \cdot \left(\frac{n-r}{n}\right)^{h_o+h_1+\sum_{s=2}^{g'+1} h_s - \frac{1}{2}p} \cdot r^{\frac{1}{2}p-h_o+\sum_{s=2}^{g'+1} h_s - \sum_{s=0}^{g'+1} sh_s}\right) =$$

$$= \mathcal{O}\left(\left(\frac{r}{n}\right)^l \cdot \left(\frac{n-r}{rn}\right)^{\frac{1}{2}p} \cdot \left(\frac{n}{r(n-r)}\right)^{\frac{1}{2}p-\sum_{s=2}^{g'+1} h_s} \cdot \left(\frac{n-r}{n}\right)^{h_o+h_1} \cdot r^{-h_o}\right) =$$

$$= \mathcal{O}\left(\left(\frac{r}{n}\right)^l \cdot \left(\frac{n-r}{rn}\right)^{\frac{1}{2}p} \cdot \left(\frac{n}{r(n-r)}\right)^{\frac{1}{2}p-\sum_{s=2}^{g'+1} h_s}\right).$$

If g' and p are both odd, then $p \geq 2\sum_{s=2}^{g'+1} h_s + 1$, and hence

$$T = \mathcal{O}\left(\left(\frac{r}{n}\right)^l \cdot \left(\frac{n-r}{rn}\right)^{\frac{1}{2}g'} \cdot \left(\frac{n}{r(n-r)}\right)^{\frac{1}{2}}\right) = \mathcal{O}\left(\left(\frac{n-r}{rn}\right)^{\frac{1}{2}g'-\frac{1}{2}} \cdot \frac{r^{l-1}}{n^l}\right) \quad ;$$

if g' is odd and p is even, then $p \geq g'+1$, and

$$T = \mathcal{O}\left(\left(\frac{r}{n}\right)^l \cdot \left(\frac{n-r}{rn}\right)^{\frac{1}{2}g'+\frac{1}{2}}\right) = \mathcal{O}\left(\left(\frac{n-r}{rn}\right)^{\frac{1}{2}g'-\frac{1}{2}} \cdot \frac{r^{l-1}}{n^l}\right) \quad ,$$

which, together with (A.11) proves (A.6).

If g' is even and p is odd, then $p \geq g'+1$ and $p \geq 2\sum_{s=2}^{g'+1} h_s + 1$, hence

$$T = \mathcal{O}\left(\left(\frac{r}{n}\right)^l \cdot \left(\frac{n-r}{rn}\right)^{\frac{1}{2}g'+\frac{1}{2}} \cdot \left(\frac{n}{r(n-r)}\right)^{\frac{1}{2}}\right) = \mathcal{O}\left(\left(\frac{n-r}{rn}\right)^{\frac{1}{2}g'-1} \cdot \frac{r^{l-2}}{n^l}\right) \quad ;$$

if g' and p are both even, but $p \geq g'+2$, then

$$T = \mathcal{O}\left(\left(\frac{r}{n}\right)^l \cdot \left(\frac{n-r}{rn}\right)^{\frac{1}{2}g'+1}\right) = \mathcal{O}\left(\left(\frac{n-r}{rn}\right)^{\frac{1}{2}g'-1} \cdot \frac{r^{l-2}}{n^l}\right) \quad ;$$

if $p = g'$, even, but $\frac{1}{2}p \geq \sum_{s=2}^{g'+1} h_s + 1$, then

$$T = \mathcal{O}\left(\left(\frac{r}{n}\right)^l \cdot \left(\frac{n-r}{rn}\right)^{\frac{1}{2}g'} \cdot \frac{n}{r(n-r)}\right) = \mathcal{O}\left(\left(\frac{n-r}{rn}\right)^{\frac{1}{2}g'-1} \cdot \frac{r^{l-2}}{n^l}\right) \quad .$$

The only remaining term for even values of g' has $p = g'$, $\sum_{s=2}^{g'+1} h_s = \frac{1}{2}p = \frac{1}{2}g'$, and hence $h_2 = \frac{1}{2}g'$, $h_s = 0$ for $s \neq 2$. Since the remainder term in (A.11) is also

$$\mathcal{O}\left(\left(\frac{n-r}{rn}\right)^{\frac{1}{2}g'-1} \cdot \frac{r^{l-2}}{n^l}\right) ,$$

we have, if g' is even,

$$A = \left(\frac{r}{n}\right)^l \sum_{\nu=0}^{g'} (-1)^\nu \binom{g'}{\nu} \frac{1}{(\frac{1}{2}g')!} \left(a_2(r,n)\nu^2\right)^{\frac{1}{2}g'} + \mathcal{O}\left(\left(\frac{n-r}{rn}\right)^{\frac{1}{2}g'-1} \cdot \frac{r^{l-2}}{n^l}\right) =$$

$$= \frac{g'!}{(\frac{1}{2}g')!} \left(\frac{r}{n}\right)^l \left(a_2(r,n)\right)^{\frac{1}{2}g'} + \mathcal{O}\left(\left(\frac{n-r}{rn}\right)^{\frac{1}{2}g'-1} \cdot \frac{r^{l-2}}{n^l}\right) \quad , \quad \text{by (A.12).}$$

From (A.3) and (A.7) - (A.9) we find

$$a_2(r,n) = -2.\frac{n-r}{rn} + \mathcal{O}\left(\frac{n-r}{r^2 n}\right) ,$$

which proves (A.5).

Next we turn to the case $\underline{z} = \underline{y}$. If $r_1 \leq r_2$ as r_1, r_2 and n tend to infinity we show that, if g' is even

(A.13)

$$A = \mathcal{B}\left(\sum_{i=1}^{h} c_{k_i,l_i}^{(\alpha_i)}\right) \cdot \left[\frac{g'!}{(\frac{1}{2}g')!}\left(\frac{2r_1r_2}{n^2}\right)^{g+g'} \prod_{i=1}^{h}\left\{\left(\frac{r_1}{n}\right)^{l_i-\tau_i}\left(\frac{r_2}{n}\right)^{\tau_i} + \left(\frac{r_1}{n}\right)^{\tau_i}\left(\frac{r_2}{n}\right)^{l_i-\tau_i}\right\} \cdot \right.$$

$$\left. \cdot \left\{-\frac{1}{2}\left(\frac{1}{r_1} + \frac{1}{r_2} - \frac{4}{n}\right)\right\}^{\frac{1}{2}g'} + \mathcal{O}\left(r_1^{h+g+\frac{1}{2}g'-1} . r_2^{l-g-g'-h} . n^{-l}\right)\right] =$$

$$= \mathcal{B}\left(\sum_{i=1}^{h} c_{k_i,l_i}^{(\alpha_i)}\right) \cdot \mathcal{O}\left(r_1^{h+g+\frac{1}{2}g'} . r_2^{l-g-g'-h} . n^{-l}\right) ,$$

and if g' is odd

$$\text{(A.14)} \qquad A = \mathcal{B}\left(\sum_{i=1}^{h} c_{k_i,l_i}^{(\alpha_i)}\right) \cdot \mathcal{O}\left(r_1^{h+g+\frac{1}{2}g'-\frac{1}{2}} . r_2^{l-g-g'-h} . n^{-l}\right) .$$

The fact that we suppose $r_1 \leq r_2$ does not imply any loss of generality.

By (3.5.8) we have in this case

$$A = \mathcal{B}\left(\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)}\right) . 2^{g'+g} \sum_{\nu=0}^{g'} (-1)^{\nu} \binom{g'}{\nu} \frac{r_1^{\nu} r_2^{\nu}}{n^{\nu}(n-1)^{\nu}} \cdot \sum_{\rho_1=0}^{1} \cdots \sum_{\rho_h=0}^{1}$$

$$\frac{r_1^{!\left(\sum_{i=1}^{h} \rho_i \tau_i + \sum_{i=1}^{h} (1-\rho_i)(l_i-\tau_i)+g+g'-\nu\right)} . r_2^{!\left(\sum_{i=1}^{h} (1-\rho_i)\tau_i + \sum_{i=1}^{h} \rho_i(l_i-\tau_i)+g+g'-\nu\right)}}{n^{!(l-2\nu)}} .$$

If the configuration $\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)}$ is non-bichromatic, i.e. $\mathcal{B}\left(\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)}\right) = 0$, the result of (A.13) and (A.14) is of course trivial. Suppose therefore that $\mathcal{B}\left(\sum_{i=1}^{h} C_{k_i,l_i}^{(\alpha_i)}\right) = 1$ and hence that

(A.15) $$1 \leq \tau_i \leq l_i - 1 \quad , \quad i = 1,2,\ldots,h \ .$$

Consider a term B of A corresponding to fixed values of $\rho_1, \rho_2, \ldots, \rho_h$. Introducing the abbreviations

$$M_1 = M_1(\rho_1,\rho_2,\ldots,\rho_h) = \sum_{i=1}^{h} \rho_i \tau_i + \sum_{i=1}^{h} (1-\rho_i)(l_i-\tau_i) \qquad \text{and}$$

$$M_2 = M_2(\rho_1,\rho_2,\ldots,\rho_h) = \sum_{i=1}^{h} (1-\rho_i)\tau_i + \sum_{i=1}^{h} \rho_i(l_i-\tau_i) \quad ,$$

and applying (A.2) we find

$$B = 2^{g'+g} \sum_{\nu=0}^{g'} (-1)^{\nu} \binom{g'}{\nu} r_1^{\nu} r_2^{\nu} \exp\Big[\log \Gamma(r_1+1) - \log \Gamma(r_1-M_1-g-g'+\nu+1) +$$
$$+ \log \Gamma(r_2+1) - \log \Gamma(r_2-M_2-g-g'+\nu+1) - \nu\left\{\log \Gamma(n+1) - \log \Gamma(n-1)\right\} +$$
$$- \log \Gamma(n+1) + \log \Gamma(n-l+2\nu+1)\Big] =$$
$$= \left(\frac{2r_1r_2}{n^2}\right)^{g+g'} \frac{r_1^{M_1} r_2^{M_2}}{n^{l-2g-2g'}} \sum_{\nu=0}^{g'} (-1)^{\nu} \binom{g'}{\nu} \exp\Big[\sum_{j=1}^{g'} r_1^{-j} \sum_{s=0}^{j+1} a_{js} \nu^{s} +$$
$$+ \sum_{j=1}^{g'} r_2^{-j} \sum_{s=0}^{j+1} b_{js} \nu^{s} + \sum_{j=1}^{g'} n^{-j} \sum_{s=0}^{j+1} c_{js} \nu^{s} + \mathcal{O}\left(r_1^{-g'-1}\right)\Big] ,$$

where a_{js} , b_{js} and c_{js} are defined by

$$\frac{(-1)^{j+1}}{j(j+1)}\left[\Phi_{j+1}(1)-\Phi_{j+1}(1-M_1-g-g'+\nu)\right]=\sum_{s=0}^{j+1}a_{js}\nu^s ,$$

(A.16)
$$\frac{(-1)^{j+1}}{j(j+1)}\left[\Phi_{j+1}(1)-\Phi_{j+1}(1-M_2-g-g'+\nu)\right]=\sum_{s=0}^{j+1}b_{js}\nu^s ,$$

$$\frac{(-1)^{j+1}}{j(j+1)}\left[-(\nu+1)\Phi_{j+1}(1)+\nu\Phi_{j+1}(-1)+\Phi_{j+1}(1-l+2\nu)\right]=\sum_{s=0}^{j+1}c_{js}\nu^s .$$

By (A.15) $M_1 \geq h$; as $M_1+M_2=\sum_{i=1}^{h} l_i = l-2g-2g'$, independent of $\rho_1,\rho_2,\ldots,\rho_h$ we have

$$r_1^{M_1}r_2^{M_2}=\mathcal{O}\left(r_1^h r_2^{l-2g-2g'-h}\right) ,$$

and

$$B=\left(\frac{2r_1r_2}{n^2}\right)^{g+g'}\frac{r_1^{M_1}\,r_2^{M_2}}{n^{l-2g-2g'}}\sum_{\nu=0}^{g'}(-1)^\nu\binom{g'}{\nu}\exp\left[\sum_{s=0}^{g'+1}d_s(r_1,r_2,n)\,\nu^s\right]+$$
$$+\mathcal{O}\left(r_1^{h+g-1}.r_2^{l-g-g'-h}.n^{-l}\right) ,$$

where

(A.17)
$$d_s(r_1,r_2,n)=\sum_{j=\max(1,s-1)}^{g'}\left(a_{js}r_1^{-j}+b_{js}r_2^{-j}+c_{js}n^{-j}\right) .$$

We note that

$$d_s(r_1,r_2,n)=\mathcal{O}\left(r_1^{-1}\right) \quad \text{for } s=0,1$$
$$=\mathcal{O}\left(r_1^{-s+1}\right) \quad \text{for } s\geq 2 .$$

Expanding the exponentials in the expression for B we obtain

(A.18)
$$B=\left(\frac{2r_1r_2}{n^2}\right)^{g+g'}\frac{r_1^{M_1}\,r_2^{M_2}}{n^{l-2g-2g'}}\sum_{\nu=0}^{g'}(-1)^\nu\binom{g'}{\nu}\prod_{s=0}^{g'+1}\sum_{h_s=0}^{g'}\frac{1}{h_s!}\left[d_s(r_1,r_2,n).\nu^s\right]^{h_s}+$$
$$+\mathcal{O}\left(r_1^{h+g-1}.r_2^{l-g-g'-h}.n^{-l}\right).$$

To every sequence of non-negative integers $h_1,h_2,\ldots,h_{g'+1}$, all less than $g'+1$, there corresponds a term T in (A.18), and

$$T=\left[\sum_{\nu=0}^{g'}(-1)^\nu\binom{g'}{\nu}.\nu^{\sum_{s=0}^{g'+1}sh_s}\right].\mathcal{O}\left(r_1^{h+g+g'-h_0-h_1-\sum_{s=2}^{g'+1}(s-1)h_s}.r_2^{l-g-g'-h}.n^{-l}\right).$$

By (A.12) we only have to consider terms T having $p = \sum_{s=0}^{g'+1} sh_s \geq g'$.
We recall that $2\sum_{s=2}^{g'+1} h_s \leq p$ with equality if and only if p is even, $h_2 = \frac{1}{2}p$ and $h_s = 0$ for $s \neq 2$. Using the same procedure as we did for $\underline{z} = \underline{x}_B$, we find that for odd values of g' all terms T are

$$\mathcal{O}\left(r_1^{h+g+\frac{1}{2}g'-\frac{1}{2}} . r_2^{l-g-g'-h} . n^{-l}\right);$$

since this also holds for the remainder term in (A.18) it holds for B too.
For even values of g', T is

$$\mathcal{O}\left(r_1^{h+g+\frac{1}{2}g'-1} . r_2^{l-g-g'-h} . n^{-l}\right),$$

unless $p = g'$, $h_2 = \frac{1}{2}g'$ and $h_s = 0$ for $s \neq 2$.

Summing terms B over all values of $\rho_1, \rho_2, \ldots, \rho_h$ we obtain (A.14) if g' is odd. If g' is even, we have for a bichromatic configuration

$$A = \frac{g'!}{(\frac{1}{2}g')!}\left(\frac{2r_1r_2}{n^2}\right)^{g+g'} \sum_{\rho_1=0}^{1} \cdots \sum_{\rho_h=0}^{1} \frac{r_1^{M_1} r_2^{M_2}}{n^{l-2g-2g'}} \left(d_2(r_1,r_2,n)\right)^{\frac{1}{2}g'} +$$

$$+ \mathcal{O}\left(r_1^{h+g+\frac{1}{2}g'-1} . r_2^{l-g-g'-h} . n^{-l}\right).$$

By (A.3), (A.16) and (A.17)

$$d_2(r_1,r_2,n) = -\frac{1}{2}\left(\frac{1}{r_1} + \frac{1}{r_2} - \frac{4}{n}\right) + \mathcal{O}\left(r_1^{-2}\right),$$

and as a result we may change $d_2(r_1,r_2,n)$ into $-\frac{1}{2}\left(\frac{1}{r_1} + \frac{1}{r_2} - \frac{4}{n}\right)$ for $g' \geq 2$ without affecting the order of the remainder term (for $g' = 0$ this is trivially true). Since this is independent of $\rho_1, \rho_2, \ldots, \rho_h$ we may now carry out the summation over these indices to obtain the result of (A.13).

REFERENCES CITED

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

A.R. BLOEMENA (1960) On the ISING model of ferromagnetism, Report S 278, Dept. of Mathematical Statistics of the Mathematical Centre, Amsterdam.

C. VAN EEDEN, A.R. BLOEMENA (1959) On probability distributions arising from points on a lattice, Report S 257, Dept. of Mathematical Statistics of the Mathematical Centre, Amsterdam.

G.H. FREEMAN (1953) Spread of diseases in a rectangular plantation with vacancies, Biometrika 40, 287.

M.G. KENDALL (1955) Rank correlation methods, 2nd Ed., Griffin, London.

M.G. KENDALL, A. STUART (1958) The advanced theory of statistics, Vol. 1, Griffin, London.

J.G. KIRKWOOD (1938) Order and disorder in binary solid solutions, Journ. Chem. Phys. 6, 70.

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, Journ. Indian Soc. Agric. Statist. 1, 173.

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.

P.V. KRISHNA IYER (1950[a]) The theory of probability distributions of points on a lattice, Ann. Math. Statist. 21, 198; 22, 310.

P.V. KRISHNA IYER (1950[b], 1951, 1952) Further contributions to the theory of probability distributions of points on a line, Journ. Indian Soc. Agric. Statist. 2, 141; 3, 80; 4, 50.

H.A. KUIPERS (1957) Over de verdeling van het aantal "runs" in reeksen van alternatieven, Report S 219 (Ov 7), Dept. of Mathematical Statistics of the Mathematical Centre, Amsterdam.

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

P.A.P. MORAN (1948) The interpretation of statistical maps, Journ. Roy. Stat. Soc. B 10, 243.

C.A.G. NASS (1960) Een Monte-Carlo-bepaling van overschrijdingskansen, in verband met een keuzetest, Statistica Neerl. 14, 19.

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

E.T. WHITTAKER, G.N. WATSON (1915) A course of modern analysis, 2nd Ed., Cambr. Univ. Press.