MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM

STATISTISCHE AFDELING

Leiding: Prof. Dr D. van Dantzig
Chef van de Statistische Consultatie: Prof. Dr J. Hemelrijk

Report S 212 (VP 12)

The asymptotic distribution for large m of Terpstra's statistic for the problem of m rankings

by

Ph. van Elteren

(Prepublication)

1956

1. Introduction

Consider m random vectors $\underline{x}^{(\alpha)}$ ($\alpha = 1,2,\ldots,m$) with n components $\underline{x}_1^{(\alpha)}, \underline{x}_2^{(\alpha)}, \ldots, \underline{x}_n^{(\alpha)}$ being the results of measurements on n objects. M. FRIEDMAN (1937) and T.J. TERPSTRA (1955) have constructed distributionfree tests for the hypothesis H_0, that these vectors are independent and that for each α the components of $\underline{x}^{(\alpha)}$ have the same distributionfunction. Terpstra considered the more general case that an arbitrary number of components of each $\underline{x}_i^{(\alpha)}$ is available. This number here is restricted to 1 except for section 5, where the case is treated that some observations are missing.

The alternatives for these tests are not precisely formulated by the authors but as their statistics can be considered as means of rankcorrelation-measures they will often lead to rejection of H_0 when the vectors are positively correlated pair by pair.

Let $\underline{T}_{\alpha,\beta}$ be Kendall's rankcorrelation statistic (cf. KENDALL (1948), Chapters 1 and 2) for the vectors $\underline{x}^{(\alpha)}$ and $\underline{x}^{(\beta)}$ given by

$$(1.1) \qquad \underline{T}_{\alpha,\beta} \overset{\text{def}}{=} \sum_{i<j} \underline{x}_{ij}^{(\alpha)} \underline{x}_{ij}^{(\beta)} \,, \quad ^{1)}$$

where

$$(1.2) \qquad x_{ij}^{(\alpha)} \overset{\text{def}}{=} \operatorname{sgn}(\underline{x}_i^{(\alpha)} - \underline{x}_j^{(\alpha)}) \quad \text{(cf. D. VAN DANTZIG and J. HEMELRIJK (1954))},$$

then Terpstra's statistic $\underline{T}$ is defined by

$$(1.3) \qquad \underline{T} \overset{\text{def}}{=} \sum_{\alpha<\beta} \underline{T}_{\alpha,\beta} = \sum_{\alpha<\beta} \sum_{i<j} \underline{x}_{ij}^{(\alpha)} \underline{x}_{ij}^{(\beta)} =$$

$$\tfrac{1}{2} \sum_{i<j} \Big(\sum_{\alpha} \underline{x}_{ij}^{(\alpha)}\Big)^2 - \tfrac{1}{2} \sum_{i<j} \sum_{\alpha} \big(\underline{x}_{ij}^{(\alpha)}\big)^2 .$$

Let $u_1^{(\alpha)}, u_2^{(\alpha)}, \ldots, u_{g_\alpha}^{(\alpha)}$ be the different values assumed by the components of vector $\underline{x}^{(\alpha)}$ in such order that $u_1^{(\alpha)} < u_2^{(\alpha)} < \ldots < u_{g_\alpha}^{(\alpha)}$, then the number of components of $\underline{x}^{(\alpha)}$ whose observed value is $u_h^{(\alpha)}$ [1)] (i.e. the size of the h^{th} tie) is denoted by $t_h^{(\alpha)}$.

---

1) In this paper α and β are supposed to run through the values $1,2,\ldots,m$; i,j,k and l through $1,2,\ldots,n$ and h through $1,2,\ldots,g_\alpha$, with the restrictions mentioned under the summation symbols $\sum$. The random character of variable is denoted by underlining its symbol; an arbitrary value assumed by a random variable is often denoted by the same symbol not underlined.

For the construction of a distributionfree test for H_0 based on the statistic $\underline{T}$, the distribution of $\underline{T}$ under the randomization hypothesis H_0' implied by H_0 is investigated. This hypothesis H_0' states that all possibilities to allot for each α $t_1^{(\alpha)}$ values $u_1^{(\alpha)}$, $t_2^{(\alpha)}$ values $u_2^{(\alpha)}, \ldots, t_{g_\alpha}^{(\alpha)}$ values $u_{g_\alpha}^{(\alpha)}$ to the components of vector $\underline{x}^{(\alpha)}$ have the same probability. It follows that the test applies also if the components of $\underline{x}^{(\alpha)}$ denote ranks allotted to n objects in order of some qualitative property.

The following notation will be used (cf. TERPSTRA (1955))

$$(1.4)\quad \begin{cases} G_2^{(\alpha)} \stackrel{\text{def}}{=} 1 - \binom{n}{2}^{-1} \sum_h \binom{t_h^{(\alpha)}}{2} & (n \geq 2) \\ G_2^{(\alpha)} \stackrel{\text{def}}{=} 0 & (n < 2) \end{cases} \quad \text{and}$$

$$(1.5)\quad \begin{cases} G_3^{(\alpha)} \stackrel{\text{def}}{=} 1 - \binom{n}{3}^{-1} \sum_h \binom{t_h^{(\alpha)}}{3} & (n \geq 3) \\ G_3^{(\alpha)} \stackrel{\text{def}}{=} 0 & (n < 3) \,. \end{cases}$$

If further

$$(1.6)\quad G_2 \stackrel{\text{def}}{=} m^{-1} \sum_\alpha G_2^{(\alpha)}$$

$$(1.7)\quad G_3 \stackrel{\text{def}}{=} m^{-1} \sum_\alpha G_3^{(\alpha)}$$

and

$$(1.8)\quad \underline{x}_{ij} \stackrel{\text{def}}{=} m^{-\frac{1}{2}} \sum_\alpha \underline{x}_{ij}^{(\alpha)}$$

(1.3) can be written as

$$(1.9)\quad \underline{T} = \frac{m}{2} \sum_{i<j} \underline{x}_{ij}^2 - \frac{1}{4} mn\,(n-1) G_2 \,.$$

As G_2 is a constant under H_0' the distribution of $\underline{T}$ is determined by that of $\sum_{i<j} \underline{x}_{ij}^2$, the sum of squares of the signtest statistics $\underline{x}_{ij}$ applied to the differences $\underline{x}_i^{(\alpha)} - \underline{x}_j^{(\alpha)}$ for $\alpha = 1,2,\ldots,m$.

---

2) If in this paper the word "asymptotic" is used, it always refers to large values of m; the distribution of random variables is always considered under hypothesis H_0' (except for section 5).

Terpstra's results contain the asymptotic distribution for $n \to \infty$ of $\underline{T}$. In this paper the asymptotic distribution for $m \to \infty$ is considered. For that purpose the asymptotic distribution of Friedman's statistic $\underline{S}$ is used. That statistic is defined by

$$(1.10)\quad \underline{S} \overset{\text{def}}{=} \frac{1}{4} \sum_i \Big(\sum_\alpha \sum_j \underline{x}_{ij}^{(\alpha)} \Big)^2 = \frac{m}{4} \sum_i \Big(\sum_j \underline{x}_{ij} \Big)^2 = \frac{m}{4} \sum_i \underline{x}_i^2$$

where

$$(1.11)\quad \underline{x}_i \overset{\text{def}}{=} \sum_j \underline{x}_{ij} .$$

If

$$(1.12)\quad \lim_{m \to \infty} \left\{ 3G_2 + (n-2)G_3 \right\} > 0$$

the asymptotic distribution of [2)]

$$(1.13)\quad \underline{X}_1 \overset{\text{def}}{=} \left[(n(n+1) \sum_\alpha \left\{ 1 - \binom{n+1}{3}^{-1} \sum_\alpha \binom{t_h^{(\alpha)}+1}{3} \right\} \right]^{-1} .12\underline{S} =$$

$$= \left[n \left\{ 3G_2 + (n-2)G_3 \right\} \right]^{-1} .3 \sum_i \underline{x}_i^2$$

is a χ^2-distribution with $n - 1$ degrees of freedom (cf. M. FRIEDMAN (1937), A. BENARD and PH. VAN ELTEREN (1953)).

Condition (1.12) is satisfied if the number of vectors $\underline{x}^{(\alpha)}$ with $g^{(\alpha)} \geqq 2$ is $O(m)$.

2. Asymptotic distribution of $\sum_{i<j} \underline{x}_{ij}^2$

According to the central limit theorem for random vectors (cf. J.V. USPENSKY (1937) p. 318) the variables $\underline{x}_{ij}$ asymptotically possess a multinormal distribution. The means and the covariance matrix of these variables under hypothesis H_0' are derived as follows (cf. TERPSTRA (1955)).

$$(2.1)\quad \mathcal{E}\, \underline{x}_{ij} = m^{-\frac{1}{2}} \sum_\alpha \mathcal{E}\, \underline{x}_{ij}^{(\alpha)} = 0,$$

$$(2.2)\quad \mathcal{E}\, \underline{x}_{ij}^2 = m^{-1} \sum_\alpha \left(\mathcal{E}\, \underline{x}_{ij}^{(\alpha)} \right)^2 = m^{-1} \sum_\alpha G_2^{(\alpha)} = G_2 \qquad (i \neq j),$$

$$(2.3)\quad \mathcal{E}\, \underline{x}_{ij} \underline{x}_{ik} = m^{-1} \sum_\alpha \mathcal{E}\, \underline{x}_{ij}^{(\alpha)} \underline{x}_{ik}^{(\alpha)} = \frac{1}{3} m^{-1} \sum_\alpha G_3^{(\alpha)} = \frac{1}{3} G_3 \quad (\neq (i,j,k)),$$

---

2) See p. 2.

$$(2.4) \qquad \mathcal{E}\,\underline{x}_{ij}\underline{x}_{kl} = 0 \qquad (\neq (i,j,k,l)) \ ,$$

and as $\underline{x}_{ij} = -\underline{x}_{ji}$

$$\mathcal{E}\,\underline{x}_{ji}\underline{x}_{ik} = \mathcal{E}\,\underline{x}_{ij}\underline{x}_{ki} = -\frac{1}{3}\,G_3 \text{ etc.}$$

The asymptotic distribution of $\underline{T}$ can, for each value of n, be determined from the identity

$$\sum_{i<j} \underline{x}_{ij}^2 = \sum_{\nu} \lambda_\nu \underline{z}_\nu^2 \qquad (\nu = 1,2,\ldots,\tbinom{n}{2}) \ ,$$

where the quantities $\underline{z}_\nu$ are asymptotically independent and N(0,1)-distributed random variables and $\lambda_1,\lambda_2,\ldots,\lambda_{\binom{n}{2}}$ are the latent roots of the known covariance matrix of the $\binom{n}{2}$ variables $\underline{x}_{ij}$. The computation of these latent roots can be avoided if the following simple relation is used

$$(2.5) \qquad \sum_{i<j} \underline{x}_{ij}^2 = n^{-1}\Big(\sum_i \underline{x}_i^2 + \sum_{i<j<k} \underline{x}_{ijk}^2\Big)$$

with $\underline{x}_i$ defined by (1.11) and $\underline{x}_{ijk}$ by

$$(2.6) \qquad \underline{x}_{ijk} \overset{\text{def}}{=} \underline{x}_{ij} + \underline{x}_{jk} + \underline{x}_{ki} \ .$$

The variables $\underline{x}_1,\underline{x}_2,\ldots,\underline{x}_n$, $\underline{x}_{123},\underline{x}_{124},\ldots,\underline{x}_{12n}$, $\underline{x}_{134},\ldots,\underline{x}_{n-2,n-1,n}$ possess asymptotically a multinormal distribution. Their means are zero and for the covariance of a variable $\underline{x}_i$ and a variable $\underline{x}_{ijk}$ is found

$$\mathcal{E}\,\underline{x}_i\underline{x}_{jkl} = 0 \qquad (j<k<l) \ .$$

Thus the variables $\underline{x}_i$ are asymptotically independent of the variables $\underline{x}_{ijk}$ and consequently $\sum_i \underline{x}_i^2$ is asymptotically independent of $\sum_{i<j<k} \underline{x}_{ijk}^2$.

As the asymptotic distribution of $\sum_i \underline{x}_i^2$ is given by (1.12) only the asymptotic distribution of $\sum_{i<j<k} \underline{x}_{ijk}^2$ has to be considered.

The variables $\underline{x}_{ijk}$ have asymptotically a multinormal distribution with covariance matrix given by

$$(2.7) \qquad \mathcal{E}\,\underline{x}_{ijk}^2 = 3\,\mathcal{E}\,\underline{x}_{ij}^2 - 6\,\mathcal{E}\,\underline{x}_{ij}\,\underline{x}_{ik} = 3\,G_2 - 2\,G_3 \qquad (\neq (i,j,k))$$

$$(2.8)\quad \mathcal{E}\,\underset{\sim}{x}_{ijk}\underset{\sim}{x}_{ijl} = \mathcal{E}\,\underset{\sim}{x}_{ij}^2 - 2\mathcal{E}\,\underset{\sim}{x}_{ij}\underset{\sim}{x}_{ik} + 2\,\mathcal{E}\,\underset{\sim}{x}_{ij}\underset{\sim}{x}_{kl} =$$

$$= \frac{1}{3}\,(3\,G_2 - 2\,G_3) \qquad (\neq (i,j,k,l))\ .$$

The other covariances are zero.

As this covariance matrix is singular, we use the following identity

$$(2.9)\qquad \underset{\sim}{x}_{ijk} = \underset{\sim}{x}_{ijn} + \underset{\sim}{x}_{ink} + \underset{\sim}{x}_{njk}\ .$$

It follows from (2.9) that all variables $\underset{\sim}{x}_{ijk}$ can be expressed in terms of the variables $\underset{\sim}{x}_{ijn}$, with $i < j \leqq n-1$, whose covariance matrix $A = (a_{(ij),(kl)})$ $(i < j \leqq n-1;\ k < l \leqq n-1)$ is defined by:

$$a_{(ij),(kl)} = \mathcal{E}\,\underset{\sim}{x}_{ijn}\underset{\sim}{x}_{kln} =$$

$$= 3\,G_2 - 2\,G_3 \qquad \text{for } i = k;\ j = l,$$

$$= \frac{1}{3}\,(3\,G_2 - 2\,G_3) \qquad \text{for } i = k;\ j \neq l \text{ or } j = l;\ i \neq k,$$

$$= -\frac{1}{3}(3\,G_2 - 2\,G_3) \qquad \text{for } i = l \text{ or } j = k,$$

$$= 0 \qquad \text{for } \neq (i,j,k,l).$$

Formula (2.9) gives the following identity:

$$\sum_{i<j<k} \underset{\sim}{x}_{ijk}^2 = (n-2) \sum_{i<j\leqq n-1} \underset{\sim}{x}_{ijn}^2 - 2 \sum_{i<j<k\leqq n-1} (\underset{\sim}{x}_{ijn}\underset{\sim}{x}_{ikn} +$$

$$+ \underset{\sim}{x}_{ikn}\underset{\sim}{x}_{jkn} - \underset{\sim}{x}_{jkn}\underset{\sim}{x}_{ijn})\ .$$

The right hand member is a quadratic form in $\underset{\sim}{x}_{ijn}$ with matrix $B = (b_{(ij)(kl)})$ $(i<j\leqq n-1,\ k<l\leqq n-1)$ given by:

$$b_{(ij)(kl)} = n-2 \text{ for } i = k;\ j = l$$
$$= -1 \quad \text{for } i = k,\ j \neq l \text{ or } j = l,\ i \neq k$$
$$= +1 \quad \text{for } i = l \text{ or } j = k$$
$$= 0 \quad \text{for } \neq (i,j,k,l).$$

The product AB is found to be a diagonal matrix with diagonal elements $\frac{1}{3}\,n(3\,G_2 - 2\,G_3)$. Hence, if $3\,G_2 - 2\,G_3 \neq 0$

$$A^{-1} = 3\,n^{-1}(3\,G_2 - 2\,G_3)^{-1}B$$

and thus

$$\underline{X}_2 \overset{\text{def}}{=} 3\, n^{-1}(3\, G_2 - 2\, G_3)^{-1} \sum_{i<j\leq n-1} \sum_{k<l\leq n-1} b_{(ij)(kl)} \underline{x}_{ijn} \underline{x}_{kln} =$$

(2.10)

$$= 3\, n^{-1}(3\, G_2 - 2\, G_3)^{-1} \sum_{i<j<k} \underline{x}^2_{ijk}$$

asymptotically has a χ^2-distribution with $\binom{n-1}{2}$ degrees of freedom.

The following theorem is now easily derived from (1.9), (1.13), (2.5) and (2.10).

Theorem I:

The asymptotic distribution for $m \to \infty$ under H_0' of Terpstra's statistic $\underline{T}$ is the convolution of the distributionsof two independent variates:

1. $\frac{m}{6}\left\{3\, G_2 + (n-2)G_3\right\} \underline{X}_1 - \frac{1}{4} mn(n-1)G_2$

and

2. $\frac{m}{6}(3\, G_2 - 2\, G_3)\underline{X}_2$

where $\underline{X}_1$ and $\underline{X}_2$ have χ^2-distributionswith $n-1$ and $\binom{n-1}{2}$ degrees of freedom respectively.

3. Remarks about the application of theorem I

From (1.4) and (1.5) is deduced

$$(3.1)\quad 3\, G_2^{(\alpha)} - 2\, G_3^{(\alpha)} = \frac{\left[n^3 + \sum_h \left\{2(t_h^{(\alpha)})^3 - 3n(t_h^{(\alpha)})^2\right\}\right]}{n(n-1)(n-2)}$$

If a tie of size t is divided into two ties of sizes u and v respectively ($t = u + v$), the value of $3\, G_2^{(\alpha)} - 2\, G_3^{(\alpha)}$ is increased by

$$(3.2)\quad \frac{2(u^3 + v^3 - t^3) - 3n(u^2 + v^2 - t^2)}{n(n-1)(n-2)} = \frac{6\, uv(n-t)}{n(n-1)(n-2)} .$$

Now it is seen from (3.1) that $3\, G_2^{(\alpha)} - 2\, G_3^{(\alpha)} = 0$ if $g_\alpha = 1$ and from (3.2) that this also holds if $g_\alpha = 2$. In all other cases $3\, G_2^{(\alpha)} - 2\, G_3^{(\alpha)}$ will be positive. It follows that $3\, G_2 - 2\, G_3$ tends to zero if and only if the number of vectors with $g_\alpha \geq 3$ is $O(m^{1-\varepsilon})$ $(0 < \varepsilon < 1)$ for large m, and then the asymptotic distribution will be a χ^2-distribution with $(n-1)$ degrees of freedom. It follows also that, if for all vectors $g_\alpha \leq 2$, Terpstra's and Friedman's tests are equivalent as then all $\underline{x}_{ijk}$ are zero and thus

$$\sum_{i<j} \underline{x}^2_{ij} = n^{-1} \sum_i \underline{x}^2_i .$$

Now the case that $3\,G_2 - 2\,G_3$ converges to a positive value is considered. Then Terpstra's statistic $\underline{T}$ can be written in the following form

$$(3.3)\quad \underline{T} = \frac{m}{6}(3\,G_2 - 2\,G_3)\underline{X} - \frac{1}{4}\,mn(n-1)G_2$$

where

$$(3.4)\quad \underline{X} = \frac{3\,G_2 + (n-2)G_3}{3\,G_2 - 2\,G_3}\,\underline{X}_1 + \underline{X}_2 = c\,\underline{X}_1 + \underline{X}_2 \quad \text{(say)},$$

with $\underline{X}_1$ and $\underline{X}_2$ defined as in theorem I.

The asymptotic densityfunction $f(x)$ of $\underline{X}$ can be expressed in the following way:

$$f(x) = c^{-1}\int_0^x f_{\frac{1}{2}(n-1)(n-2)}(x-z)\,f_{n-1}\left(\frac{z}{c}\right)dz$$

where $f_\nu(x)$ denotes the densityfunction of the χ^2-distribution with ν degrees of freedom.

If $\frac{1}{2}(n-1)$ is denoted by k and $\frac{1}{4}(n-1)(n-2)$ by l, then

$$(3.5)\quad f(x) = 2^{-(k+l)}c^{-1}\left\{\Gamma(k)\Gamma(l)\right\}^{-1}\int_0^x e^{-\frac{x-z}{2}}e^{-\frac{z}{2c}}(x-z)^{l-1}\left(\frac{z}{c}\right)^{k-1}dz$$

$$= 2^{-(k+l)}c^{-k}b^{-(k+l-1)}\left\{\Gamma(k)\Gamma(l)\right\}^{-1}e^{-\frac{x}{2c}}\int_0^{bx} e^{-\frac{1}{2}t}t^{l-1}(bx-t)^{k-1}dt\,,$$

where $b = 1 - c^{-1}$, $t = (x-z)(1-c^{-1})$.

For odd values of n, k is an integer number and thus

$$\int_0^{bx} e^{-\frac{1}{2}t}t^{l-1}(bx-t)^{k-1}dt = \sum_{j=0}^{k-1}(-1)^j\binom{k-1}{j}(bx)^{k-1-j}\int_0^{bx} e^{-\frac{1}{2}t}t^{l+j-1}dt.$$

It follows that:

$$(3.6)\quad f(x) = (2c)^{-k}b^{-l}\left\{\Gamma(l)\right\}^{-1}e^{-\frac{x}{2c}}\sum_{j=0}^{k-1}\left(-\frac{2}{b}\right)^j\frac{\Gamma(l+j)}{\Gamma(j+1)\Gamma(k-j)}x^{k-1-j}\cdot F_{2(l+j)}(bx)$$

where $F_\nu(x)$ is the distributionfunction of the χ^2-distribution with ν degrees of freedom.

If $F(x)$ is the distributionfunction of $\underline{X}$, (3.6) gives

$$(3.7)\quad F(x) = (2c)^{-k}b^{-l}\left\{\Gamma(l)\right\}^{-1}\sum_{j=0}^{k-1}\left(-\frac{2}{b}\right)^j\frac{\Gamma(l+j)}{\Gamma(j+1)\Gamma(k-j)}I_{k-1-j,2(l+j)}(x),$$

where $I_{r,s}(x) \overset{\text{def}}{=} \int_0^x e^{-\frac{t}{2c}}t^r F_s(bt)dt$ (r integer, $r \geq 0$, $s > 0$) and by induction

$$(3.8)\quad I_{r,s}(x) = 2c\sum_{i=0}^{r}c^i r!\,i\left\{2^r b^{\frac{s}{2}}\frac{\Gamma(r+\frac{s}{2}-i)}{\Gamma(\frac{s}{2})}F_{2r+s-2i}(x) - 2^i e^{-\frac{x}{2c}}x^{r-i}F_s(bx)\right\},$$

where $r^{!i}$ denotes the i-th factorial power of r.

Substituting (3.8) in (3.7) interchanging the order of summation and putting h = k - 1 - i the following expression is found

$$(3.9)\quad F(x) = \frac{1}{\Gamma(l)} \sum_{h=0}^{k-1} \sum_{j=0}^{h} \frac{(-1)^j}{c^h \Gamma(h-j+1)\Gamma(j+l)} \times$$

$$\times \left\{ \Gamma(h+l) F_{2(h+l)}(x) - \left(\frac{c}{c-1}\right)^{l+j} \Gamma(l+j) e^{-\frac{x}{2c}} \left(\frac{x}{2}\right)^{h-j} F_{2(l+j)}(x(1-c^{-1})) \right\}.$$

In this way for odd values of n, a finite expansion of F(x) in terms of χ^2-distributionfunctions is derived. For instance for n = 3 (k = 1; l = $\frac{1}{2}$)

$$(3.10)\quad F(x) = \frac{1}{\Gamma(\frac{1}{2})} \left\{ \Gamma(\tfrac{1}{2}) F_1(x) - \left(\frac{c}{c-1}\right)^{\frac{1}{2}} \Gamma(\tfrac{1}{2}) e^{-\frac{x}{2c}} F_1(x(1-c^{-1})) \right\}$$

$$= F_1(x) - \sqrt{\frac{c}{c-1}} \cdot e^{-\frac{x}{2c}} F_1(x(1-c^{-1}))$$

or for large x

$$F(x) \approx 1 - \sqrt{\frac{c}{c-1}} \cdot e^{-\frac{x}{2c}} .$$

For even values of n, k is equal to an integer $+\frac{1}{2}$, and formula (3.5) gives an infinite expansion, in terms with alternating signs. In that case, the expansion in positive terms, due to H. ROBBINS and E.J.G. PITMAN (1949) seems to be preferable. It gives:

$$(3.11)\quad F(x) = \sum_{j=0}^{\infty} K_j F_{\frac{1}{2}n(n-1)+2j}(x)$$

with

$$K_j \overset{\text{def}}{=} \frac{(n-3+2)(n-3+4)\ldots(n-3+2j)}{2 \cdot 4 \cdot \ldots \cdot 2j} \left(\frac{c-1}{c}\right)^j c^{-\frac{1}{2}(n-1)} .$$

This expansion is only useful for small values of n, as the series (3.11) converges too slowly for larger values. For large even and odd values of n, numerical convolution of the distributions of $c\underline{X}_1$ and $\underline{X}_2$ will be easier than the application of the formulas (3.9) or (3.11).

4. Comparison of the exact and the asymptotic distribution of $\underline{T}$

The asymptotic distribution of $\underline{T}$ defined in theorem I will in practice be used as an approximation to the exact distribution of $\underline{T}$ for relatively large values of m. For the exact distribution of $\underline{X}$ mean and variance can be derived by (3.3) from Terpstra's results for $\underline{T}$ (cf. TERPSTRA (1955). They are

$$\mathcal{E}\underline{X} = \frac{3}{2} n(n-1)G_2(3G_2 - 2G_3)^{-1}$$

and

$$\sigma^2\{\underline{X}\} = 36\, m^{-2}(3G_2 - 2G_3)^{-2}\sigma^2\{\underline{T}\} =$$

$$= n(n-1)(3G_2 - 2G_3)^{-2}\cdot\left[2(n-2)\left\{G_3^2 - m^{-2}\sum_{\alpha}(G_3^{(\alpha)})^2\right\} + \right.$$

$$\left. + 9\left\{G_2^2 - m^{-2}\sum_{\alpha}(G_2^{(\alpha)})^2\right\}\right].$$

And for the asymptotic distribution (cf. (3.4))

$$\mathcal{E}\underline{X} = c\,\mathcal{E}\underline{X}_1 + \mathcal{E}\underline{X}_2 = \frac{3}{2} n(n-1)G_2(3G_2 - 2G_3)^{-1}$$

and

$$\sigma^2\{\underline{X}\} = c^2\sigma^2\{\underline{X}_1\} + \sigma^2\{\underline{X}_2\} = n(n-1)(3G_2 - 2G_3)^{-2} \times$$

$$\times\left\{2(n-2)G_3^2 + 9\,G_2^2\right\}.$$

Exact and asymptotic distribution have the same mean and the exact variance of $\underline{X}$ is always smaller than the asymptotic variance. In proportion to the exact variance the difference is $O(m^{-1})$.

The author computed the exact distribution of $\underline{X}$ for $n = 3$, $m = 3,4,5,6$ and $g_\alpha = 3$ ($\alpha = 1,2,\ldots,m$) (hence $c = n + 1 = 4$) and compared them to the corresponding asymptotic distribution (cf. (3.10)).

In table I the frequencies $fq(X) \overset{def}{=} 6^{m-1}P[\underline{X} = X]$ and the tailprobabilities $P[\underline{X} \geq X]$ for the exact distribution and $P[\underline{X} \geq X] = 1 - F(X)$ for the asymptotic distribution are given and in chart I a grafical representation of $P[\underline{X} \geq X]$. For the values of X till about X = 20 the asymptotic distribution appearently underestimates the tailprobabilities, for values larger than about X = 30, it overestimates them. The long tail of the asymptotic distribution to the right may explain its larger variance (see above). Near its 5 percent point (round X = 25) the asymptotic distribution gives a relatively good approximation to its tailprobabilities even for such small values of m as considered here.

Table I

Distribution of $\underline{X} = 6\, m^{-1}\underline{T} + 9$ for $n = 3$ and:

	m = 3		m = 4		m = 5		m = 6		m = ∞
X	fq(X)	P[X≧X]	fq(X)	P[X≧X]	fq(X)	P[X≧X]	fq(X)	P[X≧X]	P[X ≧ X]
0	-	-	15	1	-	-	310	1	1
1,8	-	-	-	-	370	1	-	-	0,8756
2	-	-	-	-	-	-	1200	0,9601	0,8581
3	17	1	48	0,9306	-	-	-	-	0,7708
4	-	-	-	-	-	-	1680	0,8058	0,6876
6	-	-	60	0,7083	-	-	825	0,5898	0,5413
6,6	-	-	-	-	430	0,7145	-	-	0,5030
8	-	-	-	-	-	-	300	0,4837	0,4234
9	-	-	28	0,4306	-	-	-	-	0,3741
10	-	-	-	-	-	-	1080	0,4451	0,3303
11	12	0,5278	-	-	-	-	-	-	0,2917
11,4	-	-	-	-	240	0,3827	-	-	0,2775
12	-	-	6	0,3009	-	-	900	0,3062	0,2575
15	-	-	24	0,2731	-	-	-	-	0,1770
16	-	-	-	-	-	-	300	0,1905	0,1563
16,2	-	-	-	-	95	0,1975	-	-	0,1524
18	-	-	20	0,1620	-	-	470	0,1519	0,1217
19	6	0,1944	-	-	-	-	-	-	0,1074
20	-	-	-	-	-	-	120	0,0914	0,0948
21	-	-	-	-	100	0,1242	-	-	0,0836
22	-	-	-	-	-	-	120	0,0760	0,0738
24	-	-	6	0,0694	-	-	66	0,0606	0,0575
25,8	-	-	-	-	30	0,0471	-	-	0,0459
26	-	-	-	-	-	-	120	0,0521	0,0448
27	1	0,0278	8	0,0417	-	-	-	-	0,0395
28	-	-	-	-	-	-	180	0,0367	0,0349
30,6	-	-	-	-	20	0,0239	-	-	0,0252
34	-	-	-	-	-	-	42	0,0135	0,0165
35,4	-	-	-	-	10	0,0085	-	-	0,0138
36	-	-	1	0,0046	-	-	20	0,0081	0,0128
38	-	-	-	-	-	-	30	0,0055	0,0100
44	-	-	-	-	-	-	12	0,0017	0,0047
45	-	-	-	-	1	0,0008	-	-	0,0042
54	-	-	-	-	-	-	1	0,0001	0,0014

$fq(X) \overset{def}{=} 6^{m-1}\, P[\underline{X} = X]$.

Table I

Distribution of $\underline{X} = 6\, m^{-1}\underline{T} + 9$ for n = 3 and:

	m = 3		m = 4		m = 5		m = 6		m = ∞
X	fq(X)	$P[\underline{X} \geq X]$	fq(X)	$P[\underline{X} \geq X]$	fq(X)	$P[\underline{X} \geq X]$	fq(X)	$P[\underline{X} \geq X]$	$P[\underline{X} \geq X]$
0	-	-	15	1	-	-	310	1	1
1,8	-	-	-	-	370	1	-	-	0,8756
2	-	-	-	-	-	-	1200	0,9601	0,8581
3	17	1	48	0,9306	-	-	-	-	0,7708
4	-	-	-	-	-	-	1680	0,8058	0,6876
6	-	-	60	0,7083	-	-	825	0,5898	0,5413
6,6	-	-	-	-	430	0,7145	-	-	0,5030
8	-	-	-	-	-	-	300	0,4837	0,4234
9	-	-	28	0,4306	-	-	-	-	0,3741
10	-	-	-	-	-	-	1080	0,4451	0,3303
11	12	0,5278	-	-	-	-	-	-	0,2917
11,4	-	-	-	-	240	0,3827	-	-	0,2775
12	-	-	6	0,3009	-	-	900	0,3062	0,2575
15	-	-	24	0,2731	-	-	-	-	0,1770
16	-	-	-	-	-	-	300	0,1905	0,1563
16,2	-	-	-	-	95	0,1975	-	-	0,1524
18	-	-	20	0,1620	-	-	470	0,1519	0,1217
19	6	0,1944	-	-	-	-	-	-	0,1074
20	-	-	-	-	-	-	120	0,0914	0,0948
21	-	-	-	-	100	0,1242	-	-	0,0836
22	-	-	-	-	-	-	120	0,0760	0,0738
24	-	-	6	0,0694	-	-	66	0,0606	0,0575
25,8	-	-	-	-	30	0,0471	-	-	0,0459
26	-	-	-	-	-	-	120	0,0521	0,0448
27	1	0,0278	8	0,0417	-	-	-	-	0,0395
28	-	-	-	-	-	-	180	0,0367	0,0349
30,6	-	-	-	-	20	0,0239	-	-	0,0252
34	-	-	-	-	-	-	42	0,0135	0,0165
35,4	-	-	-	-	10	0,0085	-	-	0,0138
36	-	-	1	0,0046	-	-	20	0,0081	0,0128
38	-	-	-	-	-	-	30	0,0055	0,0100
44	-	-	-	-	-	-	12	0,0017	0,0047
45	-	-	-	-	1	0,0008	-	-	0,0042
54	-	-	-	-	-	-	1	0,0001	0,0014

$fq(X) \stackrel{def}{=} 6^{m-1}\, P[\underline{X} = X]$.

Chart I

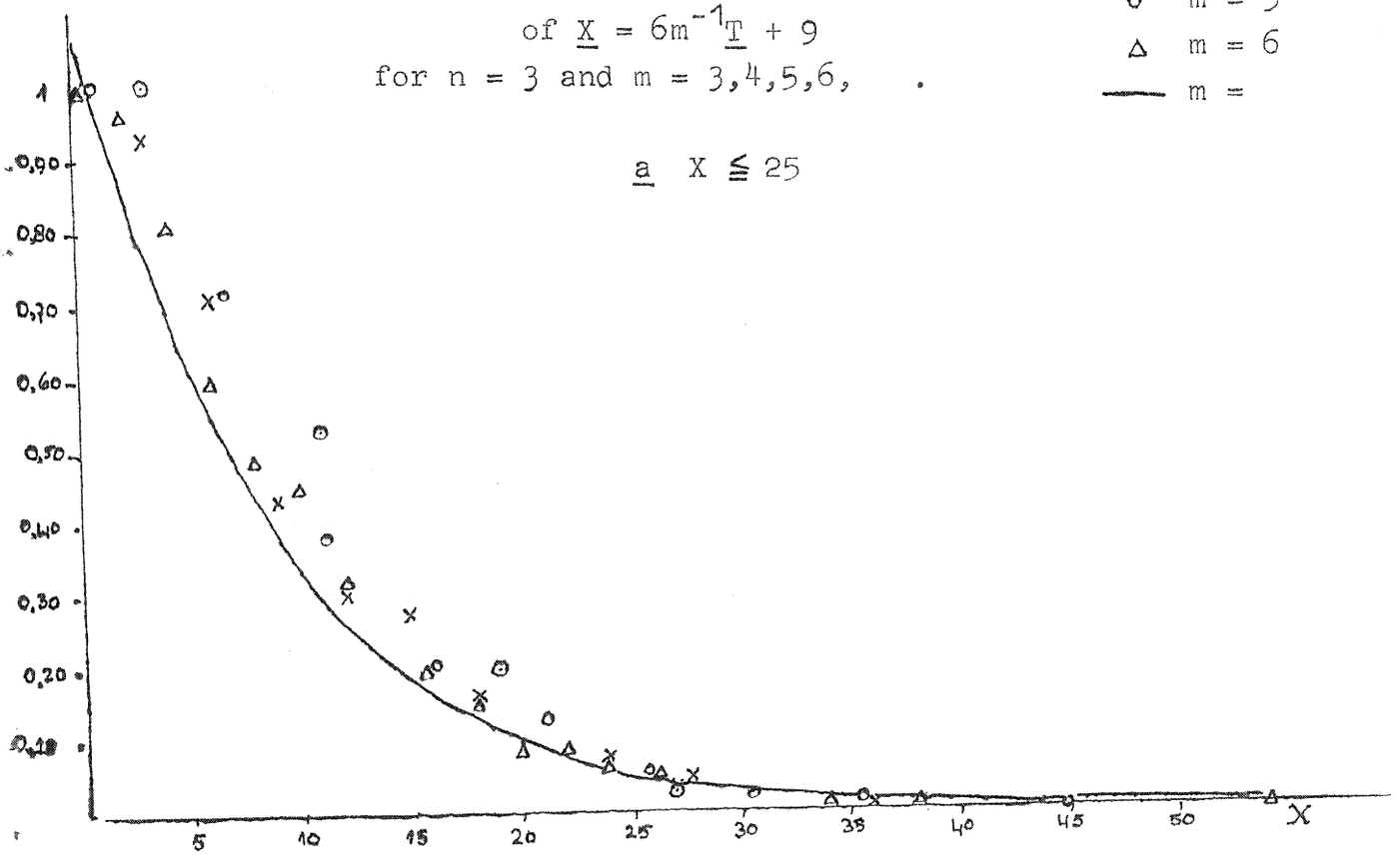
Tail-Probability-function $P\left[\underline{X} \geqq X\right]$

of $\underline{X} = 6m^{-1}\underline{T} + 9$

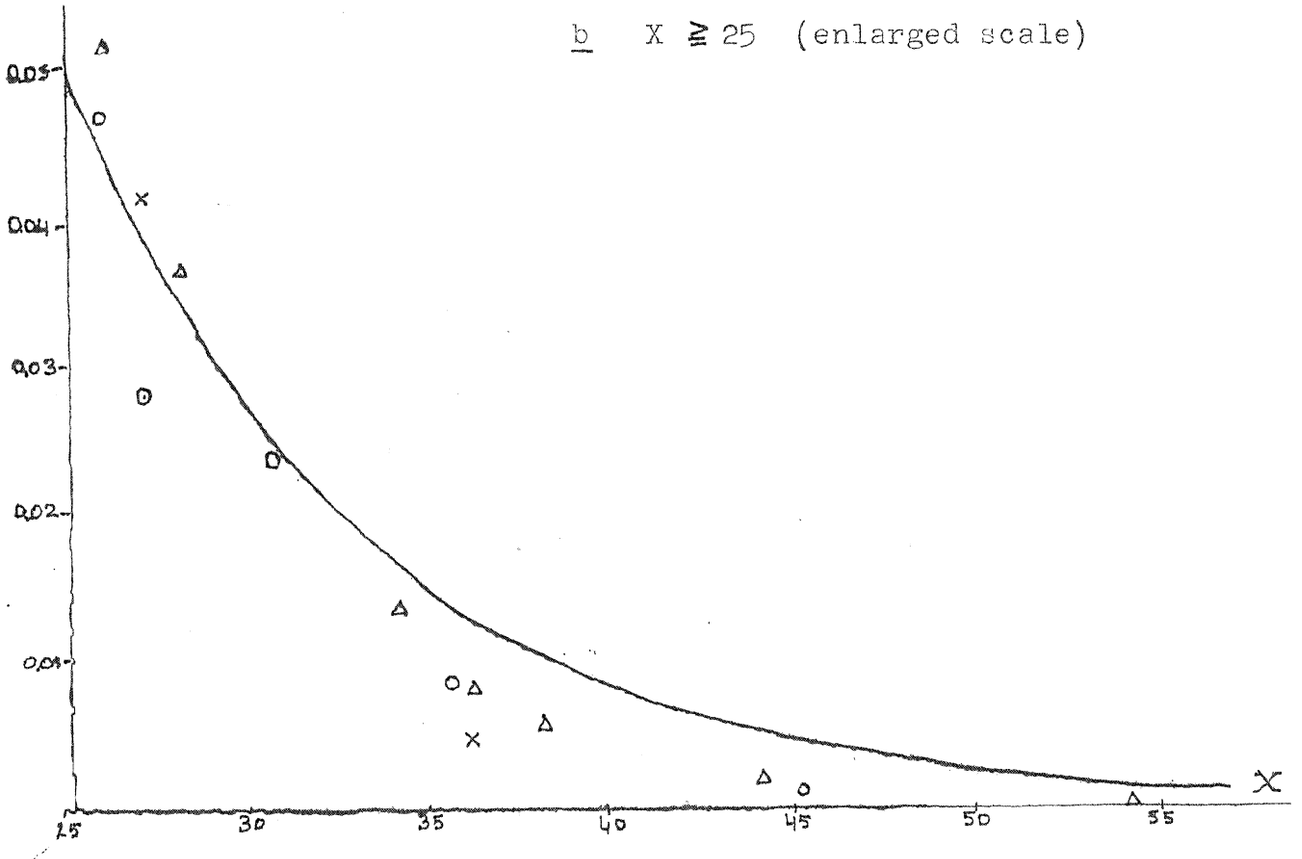
for n = 3 and m = 3,4,5,6, .

⊙ m = 3
× m = 4
o m = 5
△ m = 6
—— m =

a $X \leqq 25$



b $X \geqq 25$ (enlarged scale)



5. Missing observations

If only $n^{(\alpha)}$ components of $x^{(\alpha)}$ have been observed ($\alpha = 1,2,\ldots,m$; $n^{(\alpha)} < n$), the following modifications of hypothesis H_0' can be considered.

H_0'': All $\prod_\alpha \{n^{(\alpha)}!\}$ possibilities to allot for each α the observed values to the observed components of $\underline{x}^{(\alpha)}$ have the same probability.

H_0''': All $\prod_\alpha \{n!/(n - n^{(\alpha)})!\}$ possibilities to take for each α $n^{(\alpha)}$ components of vector $\underline{x}^{(\alpha)}$ and allotting to them the observed values have the same probability.

Hypothesis H_0'' is appropriate if the components to be observed have been chosen according to a particular design (balanced in complete blocks e.g.) or if some observations fail and one has reasons to assume that this happens with different probabilities for different components of the same vector $\underline{x}^{(\alpha)}$. Application of the generalization of the method of m rankings treated in Terpstra (1955) leads to a test for H_0'' valid for large n. We give here two tests of hypothesis H_0''', valid for large m, which are less complicated than the corresponding tests for H_0''. These tests can be used if one has omitted observations at random in order to reduce the size of the experiment or if the probability of failure of an observation is the same for all components of $\underline{x}^{(\alpha)}$

The statistics of these tests are $\underline{T}$ and $\underline{S}$, defined by (1.9) and (1.10) respectively, if the definitions of $x_{ij}^{(\alpha)}$ (cf. (1.2)), $G_2^{(\alpha)}$ (cf. (1.4)) and $G_3^{(\alpha)}$ (cf. (1.5)) are modified in the following way:

$$(5.1)\quad \begin{cases} \underline{x}_{ij}^{(\alpha)} \underline{\text{def}} \operatorname{sgn}(\underline{x}_i^{(\alpha)} - \underline{x}_j^{(\alpha)}) \text{ if both } \underline{x}_i^{(\alpha)} \text{ and } \underline{x}_j^{(\alpha)} \text{ are observed,} \\ \underline{x}_{ij}^{(\alpha)} \underline{\text{def}} \ 0 \text{ if } \underline{x}_i^{(\alpha)} \text{ or } \underline{x}_j^{(\alpha)} \text{ is not observed,} \end{cases}$$

$$(5.2)\quad G_2^{(\alpha)} \underline{\text{def}} \frac{n^{(\alpha)}(n^{(\alpha)} - 1) - \sum_h t_h^{(\alpha)}(t_h^{(\alpha)} - 1)}{n(n-1)},$$

$$(5.3)\quad G_3^{(\alpha)} \underline{\text{def}} \frac{n^{(\alpha)}(n^{(\alpha)} - 1)(n^{(\alpha)} - 2) - \sum_h t_h^{(\alpha)}(t_h^{(\alpha)} - 1)(t_h^{(\alpha)} - 2)}{n(n-1)(n-2)}.$$

For the tie-sizes $t_h^{(\alpha)}$ the same definition holds as is given in section 1 (under (1.3)) but now their total for vector α equals $n^{(\alpha)}$ instead of n.

The modified statistics $\underline{S}$ and $\underline{T}$ respectively have under H_0''' for large m the asymptotic distributions given above in (1.13) and theorem I respectively. The first remark in section 3 changes in so far the right hand member of (3.2) can not be zero except for the cases $n^{(\alpha)} = n$ or $n^{(\alpha)} = 0$. It follows that $3\,G_2^{(\alpha)} - 2\,G_3^{(\alpha)}$ is positive for $0 < n^{(\alpha)} < n$ if not all observed components of $\underline{x}^{(\alpha)}$ are equal.

Finally I want to thank Prof. Dr D. van Dantzig for his helpful suggestions which gave the paper its final form, Constance van Eeden who read the paper throuroughly and J.Th. Runnenburg, A. Benard and J. Fabius who suggested many improvements.