VERDELINGSVRYE TOETSINGSMETODES VIR DIE PROBLEEM VAN TWEE OF MEER STEEKPROEWE

deur

HERMANUS HOFMEYR LEMMER.

Voorgelê ter vervulling van 'n deel van die vereistes vir die graad

D.Sc.(WISKUNDIGE STATISTIEK)

in die Fakulteit WIS- EN NATUURKUNDE.

UNIVERSITEIT VAN PRETORIA.

PRETORIA.

MEI 1964.

PROMOTOR: PROF. DR. D. J. STOKER.

AAN MY OUERS

AAN ANNETTE.

VOORWOORD.

Met die voltooiing van hierdie proefskrif is dit vir my 'n voorreg om my waardering te kan uitspreek teenoor almal wat 'n aandeel gehad het in my akademiese opleiding.

My opregte en innige dank aan my ouers wat my in staat gestel het om myself op akademiese gebied toe te rus en my steeds aangemoedig het deur hulle opregte belangstelling in my vordering. Ek dra hierdie werk in dankbare erkentlikheid aan hulle op, asook aan my eggenote, Annette, wat vir my steeds 'n bron was van inspirasie.

Aan u, Prof. D. J. STOKER, my opregte dank vir die geleentheid wat u my gebied het om onder u leiding te studeer en vir die groot en verantwoordelike taak wat u as promotor vervul het. Ek betuig aan u my waardering dat u altyd bereid was om my met raad en daad by te staan en dat u gewilliglik baie van u tyd afgestaan het in belang van hierdie navorsing.

Baie dankie, mnre. J. G. de Jager en N. G. N. Swart, vir u hulp met die numeriese berekeninge soos saamgevat in tabel 5.3.

Ten slotte 'n woord van dank aan die W.N.N.R. en die Universiteit van Pretoria wat elk vir een jaar 'n beurs aan my toegeken het en my sodoende in staat gestel het om voltyds te studeer.

INHOUD.

HOOFSTUK I.

Bladsy.

ALGEMENE INLEIDING.

HOOFSTUK II.

VERDELINGSVRYE TOETSINGSGROOTHEDE VIR DIE PROBLEEM VAN k STEEKPROEWE ($k \geq 2$). LIMIETVERDELINGS ONDER H_o.

HOOFSTUK III.

'n ALGEMENE TOETSINGSGROOTHEID. LIMIETVERDELING ONDER 'n ALGEMENE HIPOTESE H.

INHOUD (VERVOLG). Bladsy.

HOOFSTUK IV.

DIE PROBLEEM VAN m-RANGSKIKKINGS.

HOOFSTUK V.

DIE ANALISE VAN VARIANSIE.

INHOUD (VERVOLG). Bladsy.

HOOFSTUK I.

ALGEMENE INLEIDING.

1.1. INLEIDING.

In hierdie hoofstuk word enkele fundamentele begrippe uit die algemene toetsingsteorie, wat in hierdie proefskrif gebruik word, kortliks behandel. Meer besonderhede kan onder andere gevind word in: CRAMéR (1946), HOEFFDING (1951), NOETHER (1949) en LOèVE (1955). Hoofstukke I en II van STOKER (1955) bevat 'n aansienlike hoeveelheid resultate wat in hierdie studie gebruik word.

1.2. DEFINISIES.

1.2.1. Konvergensie in waarskynlikheid en konvergensie met waarskynlikheid een.

Laat $\underline{x}_1, \underline{x}_2, \ldots$[1] 'n ry variante wees, dan konvergeer $\underline{x}_N$ in waarskynlikheid na 'n konstante c indien, vir enige $\varepsilon > 0$,

$$(1.1) \quad \lim_{N \to \infty} P[|\underline{x}_N - c| > \varepsilon] = 0.$$

Die ry $\underline{x}_1, \underline{x}_2, \ldots$ konvergeer met waarskynlikheid een na c indien vir $\varepsilon > 0$ en $\delta > 0$ 'n positiewe natuurlike getal N bestaan sodat

$$(1.2) \quad P[\bigcap_{n \geq N} (|\underline{x}_n - c| < \varepsilon)] > 1 - \delta.$$

1.2.2. Die simbole o, O, o_p en O_p (CRAMéR (1946) p. 122).

Laat f(N) enige willekeurige funksie van N wees.

Indien f(N)/N eindig bly as N na sy limietwaarde, naamlik 0 of ∞, streef, skryf ons $f(N) = O(N)$ en

---

1) 'n Variant word van 'n getal onderskei deur onderstreping van die simbool wat die variant voorstel. Indien geen verwarring moontlik is nie, word die onderstreping soms weggelaat.

interpreteer dit: f(N) is hoogstens van orde N.

Indien f(N)/N na nul nader as N na sy limietwaarde streef, skryf ons $f(N) = o(N)$ en interpreteer dit: f(N) is van kleiner orde as N.

Hieruit volg:

$f(N) = O(1)$ beteken dat f(N) eindig of nul is as N na sy limietwaarde nader.

$f(N) = o(1)$ beteken dat f(N) na nul nader as N na sy limietwaarde nader.

Indien f(x,N)/N in waarskynlikheid na nul nader as $N \to \infty$, d.w.s. as $\lim_{N\to\infty} P[\,|f(x,N)/N| > \varepsilon] = 0$, dan skryf ons $f(x,N) = o_p(N)$ en interpreteer dit: f(x,N) is in waarskynlikheid van kleiner orde as N.

Indien daar 'n ry van gelykmatig begrensde getalle $\{K_N\}$ bestaan en $|f(x,N)/N - K_N|$ in waarskynlikheid na 0 nader as $N \to \infty$, d.w.s. as $\lim_{N\to\infty} P[\,|f(x,N)/N - K_N| > \varepsilon] = 0$, dan skryf ons $f(x,N) = O_p(N)$ en interpreteer dit: f(x,N) is in waarskynlikheid van orde N.

1.3. ALGEMENE BEGRIPPE.

Stel E_N is 'n N-dimensionale Euclidiese ruimte en $\underline{Z}$ 'n stogastiese vektor waarvan die waardes Z in E_N lê, d.w.s. $Z = (z_1, z_2, \ldots, z_N) \in E_N$.

Onder 'n hipotese verstaan ons 'n uitspraak of bewering oor die waarskynlikheidsverdelings van die $\underline{z}_i$, $i=1,2,\ldots,N$. Die toetsing van 'n hipotese H_o bestaan daarin dat daar deur middel van een of ander kriterium bepaal word of die gegewens lei tot verwerping, al dan nie, van die hipotese wat getoets word. Iedere punt $Z \in E_N$ lei dus tot 'n uitspraak „H_o word verwerp" of „H_o word nie verwerp nie". Die versameling punte Z wat lei tot die uitspraak „H_o word verwerp", word die kritieke gebied van die toets genoem.

Stel T_N is 'n meetbare funksie op die steekproefruimte E_N wat alleen waardes $T_N(Z)$ in die geslote interval (0,1) kan aanneem. As ons deur waarneming 'n punt $Z \varepsilon E_N$ vind, bepaal ons deur middel van 'n ewekansige meganisme met 'n waarskynlikheid $T_N(Z)$ of ons die hipotese sal verwerp. Indien daar geen verwarring moontlik is nie, sal ons die toets wat deur die waarskynlikheidsfunksie $T_N(Z)$ bepaal word, die toets $T_N \overset{def}{=} T_N(Z)$ noem. (Sien STOKER (1955)).

Die waarskynlikheid dat die toets T_N tot verwerping van die hipotese H_o lei wanneer die hipotese H_a geld (waar is), is dan $E(\underline{T}_N | H_a)$ en word die onderskeidingsvermoë van die toets T_N genoem.

Die onbetroubaarheid of maat van die toets T_N (om H_o te toets), word gedefinieer deur $E(\underline{T}_N | H_o)$.

Indien vir alle toetse T_N met onbetroubaarheid gelyk aan α, één toets T_N' gevind kan word sodanig dat

$E(\underline{T}_N' | H_a) \geq E(\underline{T}_N | H_a)$ vir alle toetse T_N, dan word die toets T_N' die mees onderskeidende toets met onbetroubaarheid α genoem om H_o teen H_a te toets.

'n Ry van toetse $\{T_N\}$ word asimptoties onderskeidend genoem om H_o te toets teen 'n klas van alternatiewe indien: i) die maat van die toetse gelyk is aan $\alpha < 1$, onafhanklik van N, en

ii) die onderskeidingsvermoë na één nader as $N \to \infty$ vir elke alternatief in die klas van alternatiewe.

1.4. ASIMPTOTIESE RELATIEWE DOELTREFFENDHEID. NIE-SENTRALE χ^2-VERDELINGS.

Twee toetse word met mekaar vergelyk deur die „asimptotiese relatiewe doeltreffendheid" van die een toets met betrekking tot die ander toets te bereken. Dié begrip word soos volg gedefinieer:

Beskou die rye van toetse $\{T_N\}$ en $\{T'_N\}$ van dieselfde maat α gebaseer op toetsingsgroothede t_N en t'_N respektiewelik. Die asimptotiese relatiewe doeltreffendheid (hierna genoem a.r.d.) van die toets T_N met betrekking tot die toets T'_N word gegee deur $\lim_{N' \to \infty} N'/N$ waar N die aantal waarnemings is wat nodig is om die toets T_N plaaslik (in die omgewing van die nulhipotese) dieselfde onderskeidingsvermoë te gee as die toets T'_N gebaseer op N' waarnemings. Die definisie word in 'n gewysigde vorm gebruik.

Indien $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_k$ onderling onafhanklike normaalverdeelde variante is met gemiddeldes $E(\underline{x}_i) = \mu_i$ en variansies: $\mathrm{var}(\underline{x}_i) = 1$, $i=1,2,\ldots,k$, dan besit

$T = \sum_{i=1}^{k} \underline{x}_i^2$ 'n nie-sentrale χ^2-verdeling met k grade van vryheid en noem ons

$$(1.3) \quad \lambda^2 = \sum_{i=1}^{k} \mu_i^2 \qquad 2)$$

die nie-sentraliteitsparameter van die verdeling (sien SCHEFFÉ (1959) p. 412 waar λ as nie-sentraliteitsparameter gebruik word in plaas van λ^2).

Indien verder $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_k$ k variante is wat gesamentlik normaal verdeel is met $E(\underline{x}_i) = \mu_i$ en variansie-kovariansiematriks (voortaan „momentematriks" genoem, waarmee bedoel word: tweede orde sentrale momente-matriks) $\Lambda = \{\sigma_{ii'}\}$ waar $\sigma_{ii'} = \mathrm{kov}(x_i, x_{i'})$ met $i,i'=1,2,\ldots,k$ sodanig dat $|\Lambda| > 0$, dan besit $\vec{x}'\Lambda^{-1}\vec{x}$ 'n nie-sentrale χ^2-verdeling met k grade van vryheid (hierna aangedui deur g.v.v.) en nie-sentraliteitsparameter

$$(1.4) \quad \lambda^2 = \vec{\mu}'\Lambda^{-1}\vec{\mu}$$

waar $\vec{x}' = \{x_1, x_2, \ldots, x_k\}$, $\vec{\mu}' = \{\mu_1, \mu_2, \ldots, \mu_k\}$ met $\vec{x}$ en $\vec{\mu}$

---

2) Meer algemeen, indien die x_i's gemiddeldes μ_i en variansies σ_i^2 besit, dan besit $T = \sum_{i=1}^{k} (x_i/\sigma_i)^2$ 'n nie-sentrale χ^2-verdeling met k grade van vryheid en nie-sentraliteitsparameter $\lambda^2 = \sum_{i=1}^{k} (\mu_i/\sigma_i)^2$.

die ooreenkomstige kolomvektore en Λ^{-1} die resiproke matriks van Λ. (Sien KENDALL en STUART (1961) vol. II p. 229).

Indien die kofaktore van Λ aangedui word deur $|\Lambda_{jj'}|$ vir $j,j'=1,2,\ldots,k$, dan kan bostaande resultaat soos volg uitgedruk word:

$$T = \sum_{j=1}^{k} \sum_{j'=1}^{k} \frac{|\Lambda_{jj'}|}{|\Lambda|} x_j x_{j'}$$ besit 'n nie-sentrale

χ^2-verdeling met k g.v.v. en nie-sentraliteitsparameter

$$(1.5) \quad \lambda^2 = \sum_{j=1}^{k} \sum_{j'=1}^{k} \frac{|\Lambda_{jj'}|}{|\Lambda|} E(x_j)E(x_{j'}). \quad 3)$$

Volgens SCHEFFé (1959) p. 413 geld verder dat

$E(T) = k + \lambda^2$.

Die a.r.d. van T_N met betrekking tot T_N' kan nou soos volg uitgedruk word (sien KENDALL en STUART (1961) vol. II p. 274-275):

Indien t_N en t_N' asimptoties, vir $N\to\infty$, onder dieselfde alternatiewe hipotese plaaslik (in die omgewing van die nulhipotese) nie-sentrale χ^2-verdelings met dieselfde aantal grade van vryheid besit met nie-sentraliteitsparameters λ en λ' respektiewelik, word die a.r.d. van die toets T_N met betrekking tot T_N' gegee deur λ/λ'.

1.5. KWADRATIESE VORME IN NORMAALVERDEELDE VARIANTE.

1.5.1. Sentrale χ^2-verdelings.

Stel $\underline{x}_1,\underline{x}_2,\ldots,\underline{x}_k$ is gesamentlik normaalverdeelde variante met verwagtingswaardes $E(\underline{x}_i) = 0$ en momentematriks Λ met karakteristieke wortels (getalle) $\varkappa_1,\varkappa_2,\ldots,\varkappa_k$.

Lemma 1.1. Indien ν van die karakteristieke wortels van $\Lambda \neq 0$ is, kan daar 'n ortogonale matriks $C = \{c_{ij}\}$ gevind

---

3) Tensy anders vermeld, deurloop i,i',j,j',h en m in hierdie hoofstuk die waardes $1,2,\ldots,k$ en ζ,ζ' die waardes $1,2,\ldots,k-1$.

word sodanig dat

(1.6) $T = \sum_i \sum_j \sum_m c_{im} c_{jm} \gamma_m \underline{x}_i \underline{x}_j$

'n χ^2-verdeling met ν g.v.v. besit waar

(1.7) $\gamma_i = \begin{cases} \frac{1}{\varkappa_i} & \text{vir } \varkappa_i \neq 0 \\ 1 & \text{vir } \varkappa_i = 0. \end{cases}$

Bewys: Omdat ν van die karakteristieke wortels van $\Lambda \neq 0$ is, is Λ van rang ν (CRAMéR (1946) p. 113).

Daar bestaan dus (omdat Λ simmetries is) 'n nie-singuliere matriks P sodanig dat

(1.8) $P'\Lambda P = \{\delta_i \delta_{ij}\}$ met $\delta_i = \begin{cases} 1 \text{ vir } i=1,2,\ldots,\nu \\ 0 \text{ vir } i=\nu+1,\ldots,k \end{cases}$

en δ_{ij} Kronecker se delta, want (vergelyk SCHEFFé (1959) p. 399 lemma 11'):

Neem P = CL waar C ortogonaal is sodanig dat $C'\Lambda C = \{\varkappa_i \delta_{ij}\}$ waarby veronderstel word dat $\varkappa_1, \varkappa_2, \ldots, \varkappa_\nu$ almal $\neq 0$ is, terwyl $\varkappa_{\nu+1}, \varkappa_{\nu+2}, \ldots, \varkappa_k$ almal 0 is.

Kies dan L as die diagonaalmatriks met i^{de} diagonaalelement $= \begin{cases} \varkappa_i^{-\frac{1}{2}} & \text{vir } \varkappa_i \neq 0 \\ 1 & \text{vir } \varkappa_i = 0 \end{cases}$, $i=1,2,\ldots,k.$

Die karakteristieke funksie van die verdeling van die x 'e word gegee deur

(1.9) $\emptyset_{x_1,\ldots x_k}(\vec{t}) = e^{-\frac{1}{2}\vec{t}\,'\Lambda\vec{t}}$

waar $\vec{t}\,' = \{t_1, t_2, \ldots, t_k\}$ met $\vec{t}$ die ooreenkomstige kolom-vektor.

Stel $\vec{t} = P\vec{u}$ en transformeer $\vec{x}$ na $\vec{y}$ deur die kontra-grediënte transformasie $\vec{y} = P'\vec{x}$, dan is (sien CRAMéR (1946) p. 111):

(1.10) $$\begin{aligned} \emptyset_{y_1,\ldots,y_k}(\vec{u}) &= E[\text{eksp}(i\sum_j u_j y_j)] \\ &= E[\text{eksp}(i\sum_j t_j x_j)] \\ &= \emptyset_{x_1,\ldots,x_k}(\vec{t}) \\ &= \text{eksp}(-\tfrac{1}{2}\vec{t}\,'\Lambda\vec{t}) \end{aligned}$$

$$= \text{eksp}(-\tfrac{1}{2}\vec{u}'P'\Lambda P\vec{u})$$

$$= \text{eksp}(-\tfrac{1}{2}\sum_{i=1}^{\nu} u_i^2).$$

$\sum_{i=1}^{k} y_i^2$ besit dus 'n χ^2-verdeling met ν g.v.v. (CRAMéR (1946) p. 314).

Nou is

(1.11) $LL' = \{\gamma_i \delta_{ij}\}$ met γ_i in (1.7) gedefinieer.

Dus

(1.12) $\sum_{i=1}^{k} y_i^2 = \vec{y}'\vec{y}$

$= \vec{x}'CLL'C'\vec{x}$

$= \vec{x}'\{\sum_m c_{im}c_{jm}\gamma_m\}\vec{x}$ waar $\{\sum_m c_{im}c_{jm}\gamma_m\}$ 'n simmetriese matriks is

$= \sum_i \sum_j \sum_m c_{im}c_{jm}\gamma_m x_i x_j.$

Gevolglik het

(1.6) $T = \sum_i \sum_j \sum_m c_{im}c_{jm}\gamma_m x_i x_j$

'n χ^2-verdeling met ν g.v.v.

Opmerkings.

1). Indien alle $\varkappa_i = 1$ of 0, d.i. indien die x-waardes reeds 'n momentematriks met karakteristieke wortels 0 of 1 het, dan is alle $\gamma_m = 1$, sodat

(1.13) $T = \sum_i \sum_j \sum_m c_{im}c_{jm}x_i x_j$ uit (1.6)

$= \sum_i \sum_j \delta_{ij}x_i x_j$ omdat C ortogonaal is

$= \sum_i x_i^2.$

2). Vir 'n simmetriese matriks A sal $A = A^2$ wees as en slegs as alle karakteristieke wortels van A 0 of 1 is (SCHEFFé (1959) p.403).

3). Die karakteristieke getalle van die simmetriese matriks $A = \{a_{ij}\}$ word gegee deur die oplossing van

(1.14) $|A-\varkappa I| = 0$ - sien CRAMéR (1946) p. 113 - waar I die eenheidsmatriks is.

Dié vergelyking kan geskryf word in die vorm (sien SCHUTTE en VAN ROOY (1961) p. 210):

(1.15) $\varkappa^k - a_1\varkappa^{k-1} + a_2\varkappa^{k-2} - \ldots +(-1)^k a_k = 0$

waarby

$$(1.16)\quad \begin{cases} a_1 = \sum\limits_{i=1}^{k} a_{ii}, \\ a_2 = \sum\sum\limits_{i<j} \begin{vmatrix} a_{ii}, & a_{ij} \\ a_{ji}, & a_{jj} \end{vmatrix}, \\ a_3 = \sum\sum\sum\limits_{i<j<h} \begin{vmatrix} a_{ii}, & a_{ij}, & a_{ih} \\ a_{ji}, & a_{jj}, & a_{jh} \\ a_{hi}, & a_{hj}, & a_{hh} \end{vmatrix} \\ \vdots \\ a_k = |\{a_{ij}\}| = |A|. \end{cases}$$

As nou $a_k = |A| = 0$, beteken dit dat tenminste een karakteristieke wortel = 0.

Van enige simmetriese matriks A kan die karakteristieke wortels dus gevind word deur oplossing van vergelyking (1.15).

4). Volgens TURNBULL (1947) p. 66 volg indien die k wortels van (1.15) aangedui word deur $\varkappa_1, \varkappa_2, \ldots, \varkappa_k$, dat

$$(1.17)\quad \begin{cases} a_1 = \sum\limits_{i} \varkappa_i \\ a_2 = \sum\sum\limits_{i<j} \varkappa_i\varkappa_j \\ a_3 = \sum\sum\sum\limits_{i<j<h} \varkappa_i\varkappa_j\varkappa_h \\ \vdots \\ a_k = \varkappa_1\varkappa_2\ldots\varkappa_k \quad . \end{cases}$$

As dus $\varkappa_1 = \varkappa_2 = \ldots = \varkappa_{k-1} = 1$ en $\varkappa_k = 0$ in welke geval (1.15) herlei tot

(1.18) $\varkappa(\varkappa-1)^{k-1} = 0$,

volg:

$$(1.19)\quad \begin{cases} a_1 = k-1 \\ a_2 = \binom{k-1}{2} \\ a_3 = \binom{k-1}{3} \\ \vdots \\ a_{k-1} = \binom{k-1}{k-1} = 1 \\ a_k = 0 \quad . \end{cases}$$

5). As $A = I - \vec{\nu}\vec{\nu}'$ met $\vec{\nu}' = \{\sqrt{\nu_1}, \sqrt{\nu_2}, \ldots, \sqrt{\nu_k}\}$ en $\vec{\nu}$ die ooreenkomstige kolomvektor, dan is

(1.20) $a_{ii} = 1-\nu_i$ en

(1.21) $a_{ij} = -\sqrt{\nu_i \nu_j}$ vir $i \neq j$.

Met behulp van (1.16) volg, indien

(1.22) $\sum\limits_{i=1}^{k} \nu_i = 1$ met alle $\nu_i > 0$,

dat

$$\begin{aligned} a_1 &= \sum_i (1-\nu_i) \\ &= (k-1), \\ a_2 &= \sum_{i<j}\sum \begin{vmatrix} 1-\nu_i, & -\sqrt{\nu_i\nu_j} \\ -\sqrt{\nu_i\nu_j}, & 1-\nu_j \end{vmatrix} \\ &= \sum_{i<j}\sum (1-\nu_i-\nu_j) \\ &= \binom{k-1}{2}, \text{ ens.}, \end{aligned}$$

wat in ooreenstemming is met die resultate in vergelyking (1.19).

Lemma 1.2. Indien

(1.22) $\sum\limits_{i} \nu_i = 1$ met alle $\nu_i > 0$,

dan is $A = I - \vec{\nu}\,\vec{\nu}'$ van rang $(k-1)$.

Bewys: Vergelyking (1.14) word in hierdie geval:

(1.23) $|I - \vec{\nu}\vec{\nu}' - \varkappa I| = |(1-\varkappa)I - \vec{\nu}\vec{\nu}'|$.

Vir $\varkappa = 1$ word bostaande:

$$|-\vec{\nu}\vec{\nu}'| = \begin{vmatrix} -\nu_1, & -\sqrt{\nu_1\nu_2}, & \ldots, & -\sqrt{\nu_1\nu_k} \\ -\sqrt{\nu_1\nu_2}, & -\nu_2, & \ldots, & -\sqrt{\nu_2\nu_k} \\ \cdot & \cdot & \cdot & \cdot \\ -\sqrt{\nu_1\nu_k}, & -\sqrt{\nu_2\nu_k}, & \ldots, & -\nu_k \end{vmatrix}$$

$$= (-1)^k(\nu_1\nu_2\ldots\nu_k)^k \begin{vmatrix} 1, 1, \ldots, 1 \\ 1, 1, \ldots, 1 \\ \cdot\ \cdot\ \cdot\ \cdot \\ 1, 1, \ldots, 1 \end{vmatrix}$$

wat 'n determinant van rang 1 is, sodat $\varkappa = 1$ 'n $(k-1)$-voudige wortel van (1.14) is (sien SCHUTTE en VAN ROOY (1961) p. 220 en HADLEY, G (1961) „Linear Algebra" pp. 237, 242-245).

Vir $\varkappa = 0$ herlei (1.23) na:

$$(1.24)\quad |I-\vec{\nu}\vec{\nu}'| = \begin{vmatrix} 1-\nu_1, & -\sqrt{\nu_1\nu_2}, & \dots, & -\sqrt{\nu_1\nu_k} \\ -\sqrt{\nu_1\nu_2}, & 1-\nu_2, & \dots, & -\sqrt{\nu_2\nu_k} \\ \dots & \dots & \dots & \dots \\ -\sqrt{\nu_1\nu_k}, & -\sqrt{\nu_2\nu_k}, & \dots, & 1-\nu_k \end{vmatrix}$$

$= 0$, want

$$(1.25)\quad \sqrt{\nu_j}(1-\nu_j) - \sum_{i(\neq j)} \sqrt{\nu_i}\sqrt{\nu_i\nu_j} = \sqrt{\nu_j}(1-\nu_j) - \sqrt{\nu_j} \sum_{i(\neq j)} \nu_i$$
$$= \sqrt{\nu_j}(1-\sum_i \nu_i)$$
$$= 0 \text{ vir alle } j.$$

Uit (1.24) volg dat $\varkappa = 0$ ook 'n wortel van (1.23) is.

Die rang van A is gelyk aan die aantal karakteristieke wortels wat $\neq 0$ is (CRAMÉR (1946) p. 113), d.w.s. gelyk aan (k-1).

Opmerking.

Uit A ontstaan, deur weglating van die k^{de} ry en kolom, 'n nie-singuliere matriks van rang (k-1), naamlik

$$A_1 = \begin{Bmatrix} 1-\nu_1, & -\sqrt{\nu_1\nu_2}, & \dots, & -\sqrt{\nu_1\nu_{k-1}} \\ -\sqrt{\nu_1\nu_2}, & 1-\nu_2, & \dots, & -\sqrt{\nu_2\nu_{k-1}} \\ \dots & \dots & \dots & \dots \\ -\sqrt{\nu_1\nu_{k-1}}, & -\sqrt{\nu_2\nu_{k-1}}, & \dots, & 1-\nu_{k-1} \end{Bmatrix}$$

wat die momentematriks is van $x_1, x_2, \dots, x_{k-1}$.

Lemma 1.3. Indien

$$(1.22)\quad \sum_i \nu_i = 1 \text{ met alle } \nu_i > 0,$$

word die resiproke matriks van A_1 gegee deur

$$A_1^{-1} = \frac{1}{\nu_k}\begin{Bmatrix} \nu_1+\nu_k, & \sqrt{\nu_1\nu_2}, & \dots, & \sqrt{\nu_1\nu_{k-1}} \\ \sqrt{\nu_1\nu_2}, & \nu_2+\nu_k, & \dots, & \sqrt{\nu_2\nu_{k-1}} \\ \dots & \dots & \dots & \dots \\ \sqrt{\nu_1\nu_{k-1}}, & \sqrt{\nu_2\nu_{k-1}}, & \dots, & \nu_{k-1}+\nu_k \end{Bmatrix}$$

Bewys: Dit is voldoende om aan te toon dat $A_1 A_1^{-1} = I$, die eenheidsmatriks.

Laat $A_1 A_1^{-1} = \{\alpha_{\zeta\zeta'}\}$, 'n (k-1) x (k-1) matriks waarby $\zeta, \zeta' = 1, 2, \dots, (k-1)$, dan is:

$$\alpha_{11} = \nu_k^{-1}[(1-\nu_1)(\nu_1+\nu_k)-\nu_1\nu_2-\nu_1\nu_3-\ldots-\nu_1\nu_{k-1}]$$
$$= 1 \text{ m.b.v. } (1.22),$$
$$\alpha_{22} = \nu_k^{-1}[-\nu_1\nu_2+(1-\nu_2)(\nu_2+\nu_k)-\nu_2\nu_3-\ldots-\nu_2\nu_{k-1}]$$
$$= 1, \text{ ens., sodat}$$
$$\alpha_{\zeta\zeta} = 1 \text{ vir } \zeta = 1,2,\ldots,(k-1).$$

Verder is

$$\alpha_{12} = \nu_k^{-1}[(1-\nu_1)\sqrt{\nu_1\nu_2} -(\nu_2+\nu_k)\sqrt{\nu_1\nu_2}-\nu_3\sqrt{\nu_1\nu_2}-\ldots-\nu_{k-1}\sqrt{\nu_1\nu_2}]$$
$$= \nu_k^{-1}\sqrt{\nu_1\nu_2}[1-\nu_1-\nu_2-\nu_k-\nu_3-\ldots-\nu_{k-1}]$$
$$= 0 \text{ m.b.v. } (1.22),$$
$$\alpha_{13} = \nu_k^{-1}[(1-\nu_1)\sqrt{\nu_1\nu_3}-\nu_2\sqrt{\nu_1\nu_3}-(\nu_3+\nu_k)\sqrt{\nu_1\nu_3}-\ldots-\nu_{k-1}\sqrt{\nu_1\nu_3}]$$
$$= 0, \text{ ens., sodat}$$
$$\alpha_{\zeta\zeta'} = 0 \text{ vir } \zeta'\neq\zeta.$$

Dus $\{\alpha_{\zeta\zeta'}\} = I$, waarmee die lemma bewys is.

1.5.2. Nie-sentrale χ^2-verdelings.

Stel $\underline{x}_1,\underline{x}_2,\ldots,\underline{x}_k$ is gesamentlik normaalverdeelde variante met verwagtingswaardes $E(\underline{x}_i) = \mu_i$ en momentematriks Λ.

STELLING 1.1.

Indien

(1.22) $\sum_i \nu_i = 1$ met alle $\nu_i > 0$,

(1.26) $\Lambda = I-\vec{\nu}\vec{\nu}'$,

(1.27) $\sum_i \sqrt{\nu_i}\ \mu_i = 0$,

dan besit $T = \sum_{i=1}^{k} x_i^2$ 'n nie-sentrale χ^2-verdeling met (k-1) g.v.v. en nie-sentraliteitsparameter

(1.28) $$\lambda^2 = \sum_{i=1}^{k} [E(\underline{x}_i)]^2 = \sum_{i=1}^{k} \mu_i^2.$$

Bewys: Laat $\vec{x}'_{(1)} = \{x_1,x_2,\ldots,x_{k-1}\}$ met ooreenkomstige kolomvektor $\vec{x}_{(1)}$, $\vec{\mu}'_{(1)} = \{\mu_1,\mu_2,\ldots,\mu_{k-1}\}$ met ooreenkomstige kolomvektor $\vec{\mu}_{(1)}$.

Die momentematriks A_1 is nie-singulier en van rang $(k-1)$ - sien lemma 1.2 (opmerking).

Volgens KENDALL en STUART (1961) vol. II p. 229 (sien ook §1.4) besit $\vec{x}'_{(1)}A_1^{-1}\vec{x}_{(1)}$ 'n nie-sentrale χ^2-verdeling met $(k-1)$ g.v.v. en nie-sentraliteitsparameter

$$\lambda^2 = \vec{\mu}'_{(1)}A_1^{-1}\vec{\mu}_{(1)}.$$

Omdat die momentematriks van $x_1, x_2, \ldots, x_k$ gegee word deur $\Lambda = I-\vec{\nu}\vec{\nu}'$, volg dat $\operatorname{var}(\underline{x}_i) = 1-\nu_i$ vir $i=1,2,\ldots,k$ en $\operatorname{kov}(\underline{x}_i,\underline{x}_j) = -\sqrt{\nu_i\nu_j}$ vir $i \neq j$.

Gevolglik is

(1.29) $\sum_i \sqrt{\nu_i}\operatorname{kov}(\underline{x}_i,\underline{x}_j) = 0$ vir alle j (sien (1.25)),

sodat $E[\sum_i \sqrt{\nu_i}(x_i-\mu_i)]^2 = 0$ m.b.v. WILKS (1962) p. 81.

Gevolglik is $\sum_i \sqrt{\nu_i}(x_i-\mu_i) = 0$ met waarskynlikheid een.

Uit (1.27) is dus

(1.30) $\sum_i \sqrt{\nu_i}\, x_i = 0$ met waarskynlikheid een.

Nou kan $\vec{x}'_{(1)}A_1^{-1}\vec{x}_{(1)}$ met waarskynlikheid een in 'n ander vorm geskryf word, naamlik:

$$\vec{x}'_{(1)}A_1^{-1}\vec{x}_{(1)}$$

$$= \{x_1, x_2, \ldots, x_{k-1}\}\frac{1}{\nu_k}\begin{Bmatrix} \nu_1+\nu_k, & \sqrt{\nu_1\nu_2}, & \ldots, & \sqrt{\nu_1\nu_{k-1}} \\ \sqrt{\nu_1\nu_2}, & \nu_2+\nu_k, & \ldots, & \sqrt{\nu_2\nu_{k-1}} \\ \cdot & \cdot & \cdot & \cdot \\ \sqrt{\nu_1\nu_{k-1}}, & \sqrt{\nu_2\nu_{k-1}}, & \ldots, & \nu_{k-1}+\nu_k \end{Bmatrix}\begin{Bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ x_{k-1} \end{Bmatrix}$$

$$= \nu_k^{-1}[\nu_k\sum_\zeta x_\zeta^2 + (\sum_\zeta \sqrt{\nu_\zeta}\, x_\zeta)^2]$$

$$= \nu_k^{-1}[\nu_k\sum_\zeta x_\zeta^2 + (-\sqrt{\nu_k}\, x_k)^2] \quad \text{met waarskynlikheid een uit (1.30)}$$

$$= \sum_{i=1}^{k} x_i^2.$$

Dus besit $\vec{x}'_{(1)}A_1^{-1}\vec{x}_{(1)}$ en $\sum_{i=1}^{k} x_i^2$ dieselfde verdeling.

Op analoë wyse is

$$\vec{\mu}'_{(1)}A_1^{-1}\vec{\mu}_{(1)} = \sum_{i=1}^{k}[E(x_i)]^2 = \sum_{i=1}^{k}\mu_i^2.$$

Hiermee is die stelling bewys.

Opmerkings.

1). Indien $x_1, x_2, \ldots, x_k$ asimptoties gesamentlik 'n normaalverdeling besit met asimptotiese verwagtingswaardes $E(\underline{x}_i) = \mu_i$ en asimptotiese momentematriks Λ, waarby

(1.27) $\sum\limits_i \sqrt{\nu_i}\, \mu_i = 0$,

dan besit $T = \sum\limits_{i=1}^{k} x_i^2$ asimptoties 'n χ^2-verdeling met (k-1) g.v.v. en nie-sentraliteitsparameter $\lambda^2 = \sum\limits_i \mu_i^2$.

2). As alle $\mu_i = 0$, besit T 'n sentrale χ^2-verdeling met (k-1) g.v.v.

1.6. DIE PROBLEEM VAN k ONAFHANKLIKE STEEKPROEWE.

In die praktyk beskik ons soms oor 'n aantal, sê k, onafhanklike steekproewe, afkomstig uit verskillende populasies met onbekende verdelingsfunksies F_i, i=1,2,...,k. Die vraag is dan of die verdelingsfunksies almal identies is, dit wil sê of die k steekproewe almal uit dieselfde of identiese populasies afkomstig is.

In die parametriese geval is dit gebruiklik om vir elke steekproef sekere groothede (parameters) te bereken (bv. gemiddelde, standaard-afwyking, ens.) en dan deur statistiese inferensie afleidings te maak aangaande die onderskeie populasie-parameters. 'n Ander metode word in hierdie studie gevolg, nl. die sogenaamde nie-parametriese (verdelingsvrye) metode.

Die steekproefwaardes word gesamentlik in stygende volgorde volgens grootte (of volgens 'n bepaalde eienskap) gerangskik en rangnommers toegeken. Die wyse waarop elke bepaalde steekproef se waardes versprei lê in vergelyking met die res van die waardes, gee dan 'n aanduiding in hoeverre sommige populasies moontlik verskuif lê met betrekking tot die ander, of 'n groter verspreiding toon, ens.

Verskeie toetsingsgroothede, gebaseer op rangnommers, is vir hierdie probleem gedefinieer in die geval waar k = 2. In hierdie studie word verskeie van die toetsingsgroothede vir die probleem van twee steekproewe uitgebrei na dié van k steekproewe en word ook nuwe toetsingsgroothede voorgestel.

In hoofstuk II word 'n algemene klas van toetsingsgroothede vir die probleem van k steekproewe behandel en hulle asimptotiese verdelings bepaal onder die nulhipotese H_o. Dieselfde toetsingsgroothede se asimptotiese verdelings word in hoofstuk III bepaal onder 'n algemene hipotese H. Verder word die asimptotiese relatiewe doeltreffendheid en asimptotiese onderskeidendheid van die toetse ondersoek.

Hoofstuk IV bevat toetse vir, en uitbreidings van, die probleem van m-rangskikkings.

In hoofstuk V word 'n tweefaktor variansie-analise behandel.

HOOFSTUK II.

VERDELINGSVRYE TOETSINGSGROOTHEDE VIR DIE PROBLEEM VAN k STEEKPROEWE ($k \geq 2$). LIMIETVERDELINGS ONDER H_o [1)].

2.1. INLEIDING.

Gegee $N = \sum_{i=1}^{k} n_i$ onderling onafhanklike steekproefwaardes $\underline{x}_{ij}$; $j=1,2,\ldots,n_i$; $i=1,2,\ldots,k$ waar $\underline{x}_{ij}$ vir vaste i afkomstig is uit 'n populasie met verdelingsfunksie F_i, $i=1,2,\ldots,k$. (F_i nie noodwendig kontinu nie).

Die probleem wat hier beskou word, is om die nulhipotese H_o: $F_1 \equiv F_2 \equiv \ldots \equiv F_k \equiv F$, sê, naamlik dat alle verdelingsfunksies F_i, $i=1,2,\ldots,k$ identies is, te toets teen die alternatiewe hipotese H_a dat nie almal identies is nie.

In hierdie hoofstuk word 'n algemene toets vir bogenoemde probleem behandel waarin voorsiening gemaak word vir die geval waar gelyke waardes onder die waarnemings kan voorkom deurdat die F_i's nie noodwendig kontinu veronderstel word nie. Daar word bewys dat die voorgestelde toetsingsgrootheid onder sekere voorwaardes asimptoties 'n χ^2-verdeling besit.

Verskeie spesiale gevalle word bespreek, waaronder dié van KRUSKAL (1952), die uitbreidings van die toetse van TERRY (1952) en VAN DER WAERDEN (1957) vir twee onafhanklike steekproewe na k onafhanklike steekproewe, die toetse van MOOD (1954), KLOTZ (1962) en andere.

2.2. DIE TOETSINGSGROOTHEID T_k.

2.2.1. Definisies.

Stel die funksie $\Psi_N(\delta)$ is, vir alle waardes van N,

---

1). Hierdie is 'n uitbreiding van die artikel van LEHMER en STOKER(1961) na die geval waar gelyke waarnemings kan voorkom.

2). H_o impliseer H_{oo}, nl. dat alle permutasies van die waardes gelykkansig is (STOKER(1955) pp. 13-14). In hoofstuk II word deurgaans onder H_{oo} gewerk.

kontinu en monotoon stygend [3] in δ $(0\leq\delta\leq 1)$ en dat daar 'n kontinue en strengstygende funksie $\Psi(\delta)$ in δ vir $0<\delta<1$ bestaan sodanig dat vir alle δ met $0<\delta<1$ geld $\lim_{N\to\infty}\Psi_N(\delta) = \Psi(\delta)$ en wel, vir iedere $\varepsilon>0$, gelykmatig op $\varepsilon\leq\delta\leq 1-\varepsilon$. Dui die rangnommer van x_{ij} in die gesamentlike rangskikking van die N waarnemings aan deur r_{ij} indien alle x_{ij} verskillend is (sien §2.2.2 in die geval van gelyke waarnemings).

Stel [4]

$$(2.1)\quad \underline{t}_i = \sum_{j=1}^{n_i} \Psi_N(\delta_{ij}) \quad \text{waar}$$

$$(2.2)\quad \delta_{ij} = r_{ij}/(N+1).$$

Definieer soortgelyk

$$(2.3)\quad \delta_h = h/(N+1) \qquad \text{(let op die verskil tussen } \delta_h \text{ en } \delta_{ij}\text{).}$$

Laat

$$(2.4)\quad \hat{\underline{t}}_i = (N-n_i)^{\frac{1}{2}}[\underline{t}_i - E(\underline{t}_i|H_o)][N\mathrm{var}(\underline{t}_i|H_o)]^{-\frac{1}{2}}.$$

Dan definieer ons as toetsingsgrootheid:

$$(2.5)\quad \underline{T}_k = \sum_{i=1}^{k} \hat{\underline{t}}_i^2 .$$

Opmerking.

In §2.3 word aangetoon dat die asimptotiese verdeling van T_k vir $N\to\infty$ onder sekere voorwaardes 'n χ^2-verdeling met (k-1) g.v.v. is. Die faktor $[(N-n_i)/N]^{\frac{1}{2}}$ in die definisie van $\hat{t}_i$ het tot gevolg dat $E(T_k) = (k-1)$. Sien lemma 2.4.

---

3). Alle resultate bly geldig indien „stygend" deur „dalend" vervang word. Sien bv. lemma 2.7.

4). In hierdie hoofstuk word deurgaans, tensy anders vermeld, veronderstel dat die indekse i,i' die waardes 1,2,..,k deurloop; j,j' die waardes $1,2,\ldots,n_i$ en h,h' die waardes $1,2,\ldots,N$.

2.2.2. Gelyke waarnemings (Knope).

Indien daar onder die x_{ij}'s gelyke waarnemings voorkom as gevolg van diskontinuïteite in die verdelingsfunksies F_i en/of as gevolg van metings tot 'n bepaalde akkuraatheid, kan daar aan hierdie gelyke waarnemings op verskillende maniere rangnommers toegeken word, bv.:

a). Lotingsprinsipe.

Dui die gerangskikte stel x_{ij}-waardes aan deur $z_1, z_2, \ldots, z_N$ waar $z_1 \leq z_2 \leq \ldots \leq z_N$. Gestel daar kom 'n knoop (ry van gelykes) van lengte g voor, naamlik

$z_{\nu+1} = z_{\nu+2} = \ldots = z_{\nu+g}$ en geen ander z_h vir $h \not\equiv (\nu+1, \nu+2, \ldots, \nu+g)$ is gelyk aan die z-waarde van die knoop nie.

Die lotingsprinsipe bestaan daarin dat aan elkeen van die g gelyke waarnemings deur 'n ewekansige lotingsmeganisme een van die range $\nu+1, \nu+2, \ldots, \nu+g$ toegeken word. Hierdeur bly die N! permutasies van die waarnemings onder H_o gelykkansig en speel die gelykheid van die waarnemings geen verdere rol nie. (Sien byvoorbeeld vergelyking (2.20)).

b). Metode van gemiddelde range. ('n Meer volledige bespreking word gevind in STOKER (1955)).

Aan elkeen van die gelyke waarnemings $z_{\nu+1}, z_{\nu+2}, \ldots, z_{\nu+g}$ word dieselfde gemiddelde rang toegeken, naamlik

$$(2.6) \quad g^{-1} \sum_{\mu=1}^{g} (\nu+\mu) = \nu + \tfrac{1}{2}(g+1).$$

Hierdeur bly die som van die range van die N waarnemings onveranderd ongeag of daar gelykes onder die waarnemings voorkom al dan nie. Hierdie prosedure word op elke knoop toegepas. Word aan iedere z_h 'n verskillende indeks toegeken, dan bly die N! so verkreë permutasies van die waarnemings gelykkansig onder H_o.

Op ooreenkomstige wyse as hierbo vervang ons $\Psi_N[h/(N+1)]$ in die knoop deur a_h waar

(2.7) $a_h = g^{-1} \sum_{\mu=1}^{g} \Psi_N[(\nu+\mu)/(N+1)]$ vir alle waardes van h waarvoor $\nu+1 \leq h \leq \nu+g$.

Is alle knope van lengte een (d.w.s. alle waardes verskillend), dan is $a_h = \Psi_N[h/(N+1)] = \Psi_N(\delta_h)$.

Gestel dat daar onder die N z_h's λ knope van lengtes $g_1, g_2, \ldots, g_\lambda$ voorkom. Stel

(2.8) $G_\alpha = g_1 + g_2 + \ldots + g_\alpha$

waar $\alpha \in \{1,2,\ldots,\lambda\}$ en laat

(2.9) $G_o = 0$.

Dan is

(2.10) $G_\lambda = \sum_{\alpha=1}^{\lambda} g_\alpha = N$ 5)

waar alle ongelyke waardes as knope van lengte een beskou word. a_h word in hierdie geval:

(2.11) $a_h = g_\alpha^{-1} \sum_{\mu=1}^{g_\alpha} \Psi_N[(G_{\alpha-1}+\mu)/(N+1)]$

wanneer $G_{\alpha-1}+1 \leq h \leq G_\alpha$.

Definieer ook

(2.12) $a_{ij} = a_h$ indien $x_{ij} = z_h$.

Vir ongelyke waardes (knope van lengte een) is

$a_{ij} = \Psi_N(\delta_{ij})$.

Die definisie (2.1) van t_i word nou soos volg gewysig om knope in te sluit:

(2.13) $t_i = \sum_{j=1}^{n_i} a_{ij}$.

Let op dat a_h met een indeks en a_{ij} met twee indekse verskillend is in betekenis.

c). Die sogenaamde „non-random" metòde bespreek deur CROUSE (1960). Hierop word in hierdie studie nie ingegaan nie.

---

5). α deurloop die waardes $1,2,\ldots,\lambda$ tensy anders vermeld.

2.2.3. 'n Aantal Lemmas.

Lemma 2.1. Die verwagtingswaarde van $\underline{t}_i$ onder H_o word, vir vaste i, gegee deur:

(2.14) $E(\underline{t}_i|H_o) = n_i N^{-1} \sum_h a_h.$

Bewys (sien STOKER (1955)):

$$E(\underline{t}_i|H_o) = E[\sum_j a_{ij}|H_o]$$

$$= \binom{N}{n_i}^{-1} \sum{}'[\sum_j a_{ij}]$$

waar $\sum'$ die sommasie oor alle moontlike kombinasies van n_i groothede uit N aandui, wat almal ewekansig is onder H_o

$$= \binom{N}{n_i}^{-1}\binom{N-1}{n_i-1} \sum_h a_h$$

$$= n_i N^{-1} \sum_h a_h.$$

Opmerking.

Uit (2.11) volg:

$$\sum_h a_h = \sum_\alpha [\Psi_N(\frac{G_{\alpha-1}+1}{N+1})+\Psi_N(\frac{G_{\alpha-1}+2}{N+1})+\ldots+\Psi_N(\frac{G_{\alpha-1}+g_\alpha}{N+1})]$$

$$= \sum_h \Psi_N[h/(N+1)] \quad \text{want } G_\lambda = N \text{ uit (2.10)}$$

$$= \sum_h \Psi_N(\delta_h).$$

Gevolglik is

(2.15) $E(\underline{t}_i|H_o) = n_i N^{-1} \sum_h \Psi_N(\delta_h).$

Die verwagtingswaarde van t_i onder H_o is dus onafhanklik van die teenwoordigheid van gelyke waarnemings. Dit is egter nie die geval by die variansie nie.

Stel

(2.16) $\sigma_a^2 = N^{-1} \sum_h (a_h - \bar{\Psi}_N)^2$ waar

(2.17) $\bar{\Psi}_N = N^{-1} \sum_h a_h = N^{-1} \sum_h \Psi_N(\delta_h).$

Deurgaans word onder die metode van gemiddelde range gewerk tensy anders vermeld en word aangeneem dat nie alle waardes gelyk is nie (d.w.s. nie net een knoop wat alle waardes bevat nie), sodat $\sigma_a^2 > 0$ vir eindige N.

Lemma 2.2. Die variansie van $\underline{t}_i$ onder H_o word gegee deur: 6)

(2.18) $\mathrm{var}(\underline{t}_i|H_o) = n_i(N-n_i)(N-1)^{-1}\sigma_a^2$.

Bewys (sien STOKER(1955)):

$$\mathrm{var}(t_i|H_o) = \binom{N}{n_i}^{-1} \Sigma'[\sum_j (a_{ij} - \bar{\Psi}_N)]^2$$

waarby Σ' dieselfde beteken as in lemma 2.1

$$= \binom{N}{n_i}^{-1} \Sigma'[\sum_j (a_{ij}-\bar{\Psi}_N)^2 + \sum_{j\neq j'}\sum (a_{ij}-\bar{\Psi}_N)(a_{ij'}-\bar{\Psi}_N)]$$

$$= \binom{N}{n_i}^{-1}\binom{N-1}{n_i-1} \sum_h (a_h-\bar{\Psi}_N)^2 +$$

$$+ \binom{N}{n_i}^{-1}\binom{N-2}{n_i-2}\sum_{h\neq h'}\sum (a_h-\bar{\Psi}_N)(a_{h'}-\bar{\Psi}_N)$$

$$= n_i(N-n_i)(N-1)^{-1}\sigma_a^2$$ m.b.v. USPENSKY(1937) p.176.

Opmerkings.

1). k=2 lewer: $\mathrm{var}(t_i|H_o) = mn(N-1)^{-1}N^{-1}\sum_h (a_h-\bar{\Psi}_N)^2$

met $m=n_1$ en $n=N-n_1$ wat ooreenstem met 'n resultaat afgelei deur STOKER(1955) p. 28.

2). Laat

(2.19) $\Psi_{Nh} = \Psi_N(\delta_h)$.

Definieer nou

(2.20) $\sigma_c^2 = N^{-1}\sum_h (\Psi_{Nh} - \bar{\Psi}_N)^2$;

dan kry ons soortgelyk in die geval van die lotingsprinsipe by gelyke waarnemings

(2.21) $\mathrm{var}(t_i|H_o) = n_i(N-n_i)(N-1)^{-1}\sigma_c^2$.

Lemma 2.3. Die kovariansie tussen t_i en $t_{i'}$ $(i'\neq i)$ onder H_o word gegee deur:

(2.22) $\mathrm{kov}(\underline{t}_i,\underline{t}_{i'}|H_o) = -n_i n_{i'}(N-1)^{-1}\sigma_a^2$.

Bewys: $\mathrm{kov}(\underline{t}_i,\underline{t}_{i'}|H_o) = E[(\underline{t}_i-E\underline{t}_i)(\underline{t}_{i'} - E\underline{t}_{i'})]$ $(i'\neq i)$

$$= E[(\sum_j a_{ij}-n_i\bar{\Psi}_N)(\sum_{b=1}^{n_{i'}} a_{i'b} - n_{i'}\bar{\Psi}_N)]$$

---

6). Aangesien σ_a^2 'n variant is wanneer knope voorkom, moet $\mathrm{var}(t_i|H_o)$ geïnterpreteer word as $\mathrm{var}(t_i|H_{oo};Z)$ waar $Z \equiv (z_1,z_2,..,z_N)$. Deurgaans in hierdie hoofstuk word dan onder voorwaarde van die gegewe stel waarnemings gewerk. Die verkorte notasie $\mathrm{var}(t_i)$ word ook soms gebruik.

$$= \sum_{j=1}^{n_i} \sum_{b=1}^{n_{i'}} E[(a_{ij}-\bar{\Psi}_N)(a_{i'b}-\bar{\Psi}_N)].$$

Let op dat x_{ij} en $x_{i'b}$ nie dieselfde waarneming kan aandui nie omdat $i' \neq i$ vir alle j,b met $1 \leq j \leq n_i$, $1 \leq b \leq n_{i'}$.

Dui ons die som oor alle moontlike kombinasies van n_i en $n_{i'}$ verskillende groothede uit N aan deur Σ'', dan volg verder:

$$\text{kov}(\underline{t}_i, \underline{t}_{i'} | H_o) = \sum_{j=1}^{n_i} \sum_{b=1}^{n_{i'}} \binom{N}{n_i}^{-1} \binom{N-n_i}{n_{i'}}^{-1} \Sigma'' (a_{ij}-\bar{\Psi}_N)(a_{i'b}-\bar{\Psi}_N)$$

$$= \sum_{h \neq h'} \Sigma \binom{N}{n_i}^{-1} \binom{N-n_i}{n_{i'}}^{-1} \binom{N-2}{n_i-1} \binom{N-1-n_i}{n_{i'}-1} (a_h-\bar{\Psi}_N)(a_{h'}-\bar{\Psi}_N)$$

$$= n_i n_{i'} (N-1)^{-1} N^{-1} \sum_h (a_h-\bar{\Psi}_N) \sum_{h'(\neq h)} (a_{h'}-\bar{\Psi}_N)$$

$$= -n_i n_{i'} (N-1)^{-1} \sigma_a^2 \quad \text{omdat} \quad \sum_{h'} (a_{h'}-\bar{\Psi}_N) = 0.$$

Opmerking.

In die geval van die lotingsprinsipe kry ons:

(2.23) $\text{kov}(\underline{t}_i, \underline{t}_{i'} | H_o) = -n_i n_{i'} (N-1)^{-1} \sigma_c^2.$

Lemma 2.4. Die verwagtingswaarde van T_k onder H_o word gegee deur: $E(\underline{T}_k | H_o) = (k-1)$.

Bewys: $E(\underline{T}_k | H_o) = E[\sum_i (N-n_i) N^{-1} (\underline{t}_i - E\underline{t}_i)^2 (\text{var}(\underline{t}_i))^{-1}]$

$$= \sum_i (N-n_i) N^{-1} E(\underline{t}_i - E\underline{t}_i)^2 [\text{var}(\underline{t}_i)]^{-1}$$

$$= \sum_i (N-n_i) N^{-1}$$

$$= (k-1).$$

2.3. LIMIETVERDELING VAN T_k (ONDER H_o).

By die bepaling van die limietverdeling van T_k maak ons gebruik van 'n stelling soos aangegee deur STOKER (1955), wat 'n samevatting is van resultate afgelei deur NOETHER (1949), HOEFFDING (1951) en andere, naamlik:

STELLING 2.1.

Beskou 'n ry van rye $B_N = (b_{N1}, b_{N2}, \ldots, b_{NN})$, $N=1,2,\ldots$ en stel

(2.24) $\underline{S}_N = \sum_h b_{Nh} \underline{d}_h$ waar $(\underline{d}_1, \underline{d}_2, \ldots, \underline{d}_N)$ 'n stogastiese vektor is, gedefinieer oor die N! gelykkansige permutasies van die elemente van 'n tweede ry $A_N = (a_{N1}, a_{N2}, \ldots, a_{NN})$.

Indien

(2.25) $\bar{a}_N = N^{-1} \sum_h a_{Nh}$ en

(2.26) $\bar{b}_N = N^{-1} \sum_h b_{Nh}$,

dan is $E(\underline{S}_N) = N\bar{b}_N\bar{a}_N$ en

$$\mathrm{var}(\underline{S}_N) = (N-1)^{-1} \sum_h (b_{Nh}-\bar{b}_N)^2 \sum_h (a_{Nh}-\bar{a}_N)^2.$$

Verder is

(2.27) $\tilde{\underline{S}}_N = [\underline{S}_N - E(\underline{S}_N)][\mathrm{var}(\underline{S}_N)]^{-\frac{1}{2}}$

asimptoties normaal (0,1) verdeel wanneer $N \to \infty$ indien die rye A_N en B_N voldoen aan die voorwaarde

(2.28) $$\lim_{N\to\infty} \frac{N^{\frac{1}{2}(r-2)} \sum_h (a_{Nh}-\bar{a}_N)^r \sum_h (b_{Nh}-\bar{b}_N)^r}{[\sum_h (a_{Nh}-\bar{a}_N)^2]^{\frac{1}{2}r} [\sum_h (b_{Nh}-\bar{b}_N)^2]^{\frac{1}{2}r}} = 0 \text{ vir } r=3,4,\ldots$$

'n Voldoende voorwaarde ten einde dat (2.28) geld, is

(2.29) $$\lim_{N\to\infty} \frac{N \operatorname*{maks}_h (a_{Nh}-\bar{a}_N)^2 \operatorname*{maks}_h (b_{Nh}-\bar{b}_N)^2}{\sum_h (a_{Nh}-\bar{a}_N)^2 \sum_h (b_{Nh}-\bar{b}_N)^2} = 0.$$

Word daar aan voorwaarde (2.28) of (2.29) in waarskynlikheid voldoen (byvoorbeeld in die geval van knope), dan geld die stelling, naamlik dat die voorwaardelike verdelingsfunksie van $\tilde{\underline{S}}_N$ (d.i. die verdelingsfunksie van $\tilde{\underline{S}}_N$ vir 'n gegewe stel knope) na die standaard normaal verdelingsfunksie nader - sien STOKER (1955) pp. 33-34.

Stel nou:

(2.30) $a_{Nh} = a_h$ vir $h=1,2,\ldots,N$ en $N=1,2,3,\ldots$

(2.31) $$\begin{aligned} b_{Nh} &= c_1 \text{ vir } h=1,2,\ldots,n_1 \\ &= c_2 \text{ vir } h=n_1+1, n_1+2, \ldots, n_1+n_2 \\ &\ldots\ldots\ldots\ldots\ldots\ldots \\ &= c_k \text{ vir } h=\sum_{m=1}^{k-1} n_m+1, \ldots, N=\sum_i n_i , \end{aligned}$$

waar $c_1, c_2, \ldots$ en c_k willekeurige eindige konstantes is.

Stel verder

(2.32) $\nu_{Ni} = n_i/N$ en

(2.33) $\nu_i = \lim_{N\to\infty} \nu_{Ni}$ vir $i=1,2,\ldots,k$,

dan is $\sum_i \nu_i = 1$.

Stel ook

(2.34) $0\leq\nu_i<1$ vir $i=1,2,\ldots,k$.

STELLING 2.2.

Indien

(2.35) $$\frac{\max_h (a_h-\bar{a}_N)^2}{N^{-1}\sum_h (a_h-\bar{a}_N)^2} = o_p(N) \text{ vir } N\to\infty \quad 7),$$

(2.36) minstens twee van die c_i's in (2.31) verskillend is, sê $c_p\neq c_q$, $p\neq q$,

(2.37) $\nu_p, \nu_q > 0$,

dan geld voorwaarde (2.29) in waarskynlikheid.

Bewys: Vir eindige c_i is $\max_h (b_{Nh}-\bar{b}_N)^2$ altyd eindig (vir alle N), en wel

(2.38) $\max_h (b_{Nh}-\bar{b}_N)^2 \leq \max c_i^2$, onafhanklik van N. Verder volg:

$$\begin{aligned} N^{-1}\sum_h (b_{Nh}-\bar{b}_N)^2 &= N^{-1}\sum_i n_i(c_i-\bar{b}_N)^2 \\ &= \sum_i n_i N^{-1}(c_i-\sum_{i'} n_{i'}N^{-1}c_{i'})^2 \\ &\xrightarrow[\infty]{N} \sum_i \nu_i(c_i-\sum_{i'}\nu_{i'}c_{i'})^2 \\ &= e>0 \text{ as vir minstens een i geld } \nu_i>0 \end{aligned}$$

en $\sum_{i'}\nu_{i'}c_{i'}\neq c_i$, d.i. as $\sum_{i'(\neq i)} \nu_{i'}c_{i'}\neq c_i-\nu_i c_i = c_i \sum_{i'(\neq i)} \nu_{i'}$ want $\sum_{i'}\nu_{i'}=1$, d.i. as $c_i\neq \sum_{i'(\neq i)} \nu_{i'}c_{i'} / \sum_{i'(\neq i)} \nu_{i'}$

mits $\sum_{i'(\neq i)} \nu_{i'}>0$.

Hieraan word voldoen as tenminste nog een $\nu_{i'}>0$ vir $i'\neq i$ en indien dan vir daardie i' geld $c_{i'}\neq c_i$. Origens

---

7). Die limietoorgange geld deurgaans vir $N\to\infty$.

bly die konstantes c_i willekeurig.

Nou volg:

$$\frac{\underset{h}{\text{maks}}(b_{Nh}-\bar{b}_N)^2}{\sum\limits_h (b_{Nh}-\bar{b}_N)^2} \leq \frac{\text{maks } c_i^2}{N.N^{-1}\sum\limits_h (b_{Nh}-\bar{b}_N)^2} = O(N^{-1}).$$

$$\frac{\underset{h}{\text{maks}}(a_{Nh}-\bar{a}_N)^2 \underset{h}{\text{maks}}(b_{Nh}-\bar{b}_N)^2}{N^{-1}\sum\limits_h (a_{Nh}-\bar{a}_N)^2 \sum\limits_h (b_{Nh}-\bar{b}_N)^2} = \frac{\underset{h}{\text{maks}}(a_h-\bar{\Psi}_N)^2}{N^{-1}\sum\limits_h (a_h-\bar{\Psi}_N)^2} .O(N^{-1})$$

$$= o_p(N).O(N^{-1}) \text{ uit } (2.35)$$

$$= o_p(1) \text{ sodat aan voorwaarde}$$

(2.29) in waarskynlikheid voldoen word.

Opmerking.

Voorwaarde (2.35) kan in 'n ander vorm geskryf word, en wel soos volg:

Die uitdrukking σ_a^2 is, vir vaste N en gegewe funksie Ψ_N, alleen 'n funksie van die knope $g_1, g_2, \ldots, g_\lambda$ en ons dui die afhanklikheid aan deur te skryf:

(2.39) $\sigma^2_{N,a}(g_1, g_2, \ldots, g_\lambda) = N^{-1}\sum\limits_h (a_h-\bar{\Psi}_N)^2.$

Dui die gemeenskaplike verdelingsfunksie van die k verdelings aan deur $F(x)$ (onder H_o is $F_1 \equiv F_2 \equiv \cdots \equiv F_k \equiv F$), dan neem ons as voorwaarde

(2.40) (Iedere diskontinuïteit van $F(x)) \leq \theta < 1$.

Indien $\lim\limits_{N\to\infty} \bar{\Psi}_N$, waarby $\bar{\Psi}_N = N^{-1}\sum\limits_h \Psi_{Nh}$, bestaan en gelyk is aan a (eindig), lê a in die bereik van $\Psi(\delta_h)$, en dan bestaan daar volgens stelling 2.4 van STOKER (1955) 'n positiewe konstante c^2 sodanig dat

(2.41) $\lim\limits_{N\to\infty} \text{inf. } \sigma^2_{N,a}(g_1, g_2, \ldots, g_\lambda) \geq c^2$ spr 0 [8)]

en wel is c^2 gelykmatig van 0 geskei op alle klasse van

---

8). „spr α" beteken: „behoudens 'n waarskynlikheid hoogstens gelyk aan α". Sien VAN DANTZIG (1947-1950) p. 205. Spr 0 beteken dus: met waarskynlikheid een.

verdelings $F(x)$ waarvoor die grootste diskontinuïteit gelykmatig van 1 geskei bly.

Voorwaarde (2.35) herlei dus onder (2.40) na:

$$\underset{h}{\text{maks}}(a_h-\bar{\Psi}_N)^2 \leq \underset{h}{\text{maks}}[\Psi_N(\delta_h)-\bar{\Psi}_N]^2 = o(N), \quad \text{d.i.}$$

$$(2.42) \quad \underset{h}{\text{maks}}[\Psi_N(\delta_h)-\bar{\Psi}_N]^2 = o(N).$$

Stelling 2.2 geld dus onder voorwaardes (2.36), (2.37), (2.40) en (2.42).

STELLING 2.3.

Onder die voorwaardes (2.36), (2.37), (2.40) en (2.42) is $\sum_i c_i \underline{t}_i$ asimptoties normaal verdeel as $N\to\infty$.

Bewys: Uit stelling 2.1 en 2.2(opmerking) volg dat

$$\tilde{\underline{S}}_N = [\underline{S}_N-E\underline{S}_N][\text{var}(\underline{S}_N)]^{-\frac{1}{2}} \quad \text{asimptoties normaal } (0,1)$$

verdeel is as $N\to\infty$ waarby

$$(2.43) \quad \underline{S}_N = \sum_h b_{Nh}\underline{d}_h$$
$$= \sum_i \sum_j c_i \underline{a}_{ij}$$
$$= \sum_i c_i \sum_j \underline{a}_{ij}$$
$$= \sum_i c_i \underline{t}_i$$

waar t_i in vergelyking (2.13) gedefinieer is.

Opmerkings.

1). Deurdat die c_i willekeurig is behalwe dat minstens twee verskillend moet wees, volg dat 'n willekeurige lineêre som van die t_i's asimptoties normaal verdeel is as $N\to\infty$.

2). Kies $c_i=1$ en alle ander $c_{i'}=0$ $(i'\neq i)$. Dan is $\underline{S}_N = \underline{t}_i$ en indien verder $0<\nu_i<1$ vir $n_i\to\infty$ en $N\to\infty$, in welke geval aan voorwaardes (2.36) en (2.37) voldoen is, volg uit stelling 2.3 dat $\underline{t}_i$ (vir vaste i) asimptoties normaal verdeel is onder voorwaardes (2.40) en (2.42). Sien ook opmerking 3 hieronder.

3). Indien

(2.40) (Iedere diskontinuïteit van $F(x)) \leq \theta < 1$,

(2.42) $\underset{h}{\text{maks}}(\Psi_{Nh}-\bar{\Psi}_N)^2 = o(N)$ vir $N\to\infty$ en

(2.44) $\nu_i>0$ vir $i=1,2,\ldots,k$,

dan kan ook m.b.v. STOKER (1955) stelling 2.2 opmerking 1 bewys word dat t_i asimptoties normaal verdeel is as $N\to\infty$.

4). Indien

(2.45) $\sigma_a^{-(2+\delta)}N^{-1}\sum_h|a_h-\bar{\Psi}_N|^{2+\delta} = O_p(1)$ waar $\delta>0$, en

(2.46) $n_i\to\infty$ en $(N-n_i)\to\infty$ as $N\to\infty$,

dan volg m.b.v. STOKER (1955) stelling 2.3 dat t_i asimptoties normaal verdeel is as $N\to\infty$.

5). Onder voorwaarde (2.40) is die voorwaarde

(2.47) $\sum_h|\Psi_{Nh}-\bar{\Psi}_N|^{2+\delta} = O(N)$ vir $\delta>0$

voldoende vir die geldigheid van (2.45).

6). Volgens HOEFFDING (1951) is die voorwaardes

(2.35) $$\frac{\underset{h}{\text{maks}}(a_h-\bar{\Psi}_N)^2}{N^{-1}\sum_h(a_h-\bar{\Psi}_N)^2} = o_p(N)$$ en

(2.48) $$\frac{\sum_h|a_h-\bar{\Psi}_N|^{2+\delta}}{[\sum_h(a_h-\bar{\Psi}_N)^2]^{1+\frac{1}{2}\delta}} = o_p(1). \quad , \ \delta>0,$$

ekwivalent, waaruit volg dat die voorwaarde

(2.45) $\sigma_a^{-(2+\delta)}N^{-1}\sum_h|a_h-\bar{\Psi}_N|^{2+\delta} = O_p(1)$ vir $\delta>0$

voldoende is vir die geldigheid van

(2.35) $$\frac{\underset{h}{\text{maks}}(a_h-\bar{\Psi}_N)^2}{N^{-1}\sum_h(a_h-\bar{\Psi}_N)^2} = o_p(N) .$$

7). Met behulp van die opmerking na stelling 2.2 en opmerking 6 hierbo volg, onder voorwaarde (2.40), dat

(2.47) $\sum_h|\Psi_{Nh}-\bar{\Psi}_N|^{2+\delta} = O(N)$, $\delta>0$,

voldoende is vir die geldigheid van

(2.42) $\underset{h}{\text{maks}}(\Psi_{Nh}-\bar{\Psi}_N)^2 = o(N)$.

STELLING 2.4.

Indien

(2.40) (Iedere diskontinuïteit van $F(x)$) $\leq \Theta < 1$ en

(2.44) $\nu_i > 0$ vir $i=1,2,\ldots,k$,

dan is die asimptotiese variansie-kovariansiematriks M_1 van die k variante $\hat{t}_i$, $i=1,2,\ldots,k$, van rang $(k-1)$, waar $\hat{t}_i$ gedefinieer is in (2.4).

Bewys: Onder voorwaarde (2.40) geld (2.41) sodat uit (2.4) volg (vergelyk (2.18)):

$$\begin{aligned} \mathrm{var}(\hat{t}_i) &= (N-n_i)N^{-1}\mathrm{var}[(t_i - Et_i)(\mathrm{var}[t_i])^{-\frac{1}{2}}] \\ &= 1-n_iN^{-1} \quad \text{sodat} \end{aligned}$$

$$\lim_{N\to\infty} \mathrm{var}(\hat{t}_i) = 1-\nu_i .$$

Met behulp van lemmas 2.2 en 2.3 volg dat

$$\begin{aligned} \mathrm{kov}(\hat{t}_i,\hat{t}_{i'}) &= (N-1)N^{-1}n_i^{-\frac{1}{2}}n_{i'}^{-\frac{1}{2}}\sigma_a^{-2}\mathrm{kov}(t_i,t_{i'}) \qquad i'\neq i \\ &= -(n_in_{i'})^{\frac{1}{2}}N^{-1} \quad \text{sodat} \end{aligned}$$

$$\lim_{N\to\infty} \mathrm{kov}(\hat{t}_i,\hat{t}_{i'}) = -\sqrt{\nu_i\nu_{i'}} .$$

Die variante $\hat{t}_i$ besit dus asimptoties die momentematriks

$$M_1 = \begin{pmatrix} 1-\nu_1 & ,-\sqrt{\nu_1\nu_2}, & \ldots\ldots, & -\sqrt{\nu_1\nu_k} \\ -\sqrt{\nu_1\nu_2}, & 1-\nu_2 & ,\ldots\ldots, & -\sqrt{\nu_2\nu_k} \\ \cdot & \cdot & \cdot & \cdot \\ -\sqrt{\nu_1\nu_k}, & -\sqrt{\nu_2\nu_k}, & \ldots\ldots, & 1-\nu_k \end{pmatrix}$$

$$= I-\vec{\nu}\vec{\nu}\,'$$

waar I die eenheidsmatriks is, $\vec{\nu}\,'$ die ryvektor $(\sqrt{\nu_1},\sqrt{\nu_2},\ldots,\sqrt{\nu_k})$ en $\vec{\nu}$ die ooreenkomstige kolomvektor.

Omdat $\sum_i \nu_i = 1$ en alle $\nu_i > 0$, volg uit lemma 1.2 dat M_1 van rang $(k-1)$ is.

Opmerking.

Dit kan maklik bewys word dat die variansie-kovariansiematriks M_2 van die variante t_i ook van rang $(k-1)$ is, waar $\mathrm{var}(t_i)$ in vergelyking (2.18) en $\mathrm{kov}(t_i,t_{i'})$ in vergelyking (2.22) gegee is.

Lemma 2.5. Indien vir die variante $\hat{\underline{t}}_i$, $i=1,2,\ldots,k$ geld

(2.49) $\lim\limits_{N\to\infty} \mathrm{var}(\hat{\underline{t}}_i) > 0$,

(2.50) $\lim\limits_{N\to\infty} E|\hat{\underline{t}}_i|^{2+\delta} < \infty$ vir enige $\delta>0$ en

(2.51) $\hat{\underline{S}} \overset{\mathrm{def}}{=} \sum\limits_i e_i\hat{\underline{t}}_i$ (e_i onafhanklik van N) is asimptoties normaal verdeel vir alle reële eindige koëffisiënte e_i, $i=1,2,\ldots,k$,

dan is die gesamentlike verdeling van $\hat{\underline{t}}_1,\hat{\underline{t}}_2,\ldots,\hat{\underline{t}}_k$ asimptoties normaal as $N\to\infty$.

Bewys: Sien die bewys van lemma 5 in LEMMER en STOKER (1961) of CROUSE (1960).

Lemma 2.6. Die voorwaardes

(2.40) (Iedere diskontinuïteit van $F(x)$) $\leq \theta < 1$,

(2.42) $\underset{h}{\mathrm{maks}}(\Psi_{Nh}-\bar{\Psi}_N)^2 = o(N)$ en

(2.44) $\nu_i>0$ vir $i=1,2,\ldots,k$

is voldoende vir die geldigheid van die voorwaardes

(2.49) $\lim\limits_{N\to\infty} \mathrm{var}(\hat{\underline{t}}_i) > 0$ en

(2.50) $\lim\limits_{N\to\infty} E|\hat{\underline{t}}_i|^{2+\delta} < \infty$ vir $\delta>0$.

Bewys: In stelling 2.4 is bewys dat

$$\lim_{N\to\infty} \mathrm{var}(\hat{\underline{t}}_i) = 1-\nu_i$$
$$> 0 \text{ m.b.v. } (2.44).$$

Hiermee word aan (2.49) voldoen.

Onder die voorwaarde

(2.42') $N^{-1}\sum\limits_h(\Psi_{Nh}-\bar{\Psi}_N)^4 = O(N)$

word aan voorwaarde (2.50) voldoen – sien bylaag A of LEMMER en STOKER(1961) lemma 6.

Opmerkings.

1). In al die toepassings van hierdie lemma (sien $§^e$ 2.5.2 en 2.5.3) word aan voorwaarde (2.42') voldoen – sien byvoorbeeld STOKER(1955) vergelykings (3.35) en

(3.45) met betrekking tot Van der Waerden en Terry se toetse. Hierdie voorwaarde word dus nie verder eksplisiet vermeld nie.

2). Vervang ons voorwaarde (2.42) deur

(2.47) $\sum_h |\Psi_{Nh}-\bar{\Psi}_N|^{2+\delta} = O(N)$ vir $\delta>0$,

dan geld hierdie lemma ook (sien opmerking 7 na stelling 2.3).

STELLING 2.5.

Indien

(2.40) (Iedere diskontinuïteit van $F(x)) \leq \theta < 1$,

(2.42) $\text{maks}_h(\Psi_{Nh}-\bar{\Psi}_N)^2 = o(N)$ en

(2.44) $\nu_i>0$ vir $i=1,2,\ldots,k$,

dan besit

$$T_k = \sum_i (N-n_i)N^{-1}(t_i-Et_i)^2[\text{var}(t_i)]^{-1}$$

asimptoties, vir $N\to\infty$, 'n χ^2-verdeling met $(k-1)$ g.v.v.

Bewys: Uit stelling 2.3 volg dat $\sum_i c_i t_i$ (vir nie alle c_i's gelyk nie)* asimptoties normaal verdeel is as $N\to\infty$. $[\sum_i c_i t_i - \sum_i c_i Et_i][\text{var}(\sum_i c_i t_i)]^{-\frac{1}{2}}$ is asimptoties normaal (0,1) verdeel as $N\to\infty$.

Maar

$$[\sum_i c_i t_i - \sum_i c_i Et_i][\text{var}(\sum_i c_i t_i)]^{-\frac{1}{2}} = \sum_i c_i[t_i-Et_i][\text{var}(\sum_{i'} c_{i'}t_{i'})]^{-\frac{1}{2}}$$

$$= \sum_i \frac{c_i[N\text{var}(t_i)]^{\frac{1}{2}}}{[(N-n_i)\text{var}(\sum_{i'} c_{i'}t_{i'})]^{\frac{1}{2}}}\hat{t}_i .$$

Neem nou

(2.52) $e_i = c_i[\text{var}(t_i)]^{\frac{1}{2}}[(1-\nu_{Ni})\text{var}(\sum_{i'} c_{i'}t_{i'})]^{-\frac{1}{2}}$

dan is e_i asimptoties onafhanklik van N en eindig.

Gevolglik is $\sum_i e_i\hat{t}_i$ asimptoties normaal verdeel as $N\to\infty$.

Uit stelling 2.4 en lemmas 2.5 en 2.6 volg verder dat die variante $\underline{\hat{t}}_i$, $i=1,2,\ldots,k$ gesamentlik, vir $N\to\infty$,

---

*). Anders is die limietverdeling ontaard.

asimptoties 'n normaalverdeling besit met momentematriks $M_1 = I - \vec{\nu}\vec{\nu}'$ van rang $(k-1)$. Uit stelling 1.1 en opmerking 2 daarna met $\hat{t}_i$ in plaas van x_i, volg dat $T_k = \sum_i \hat{t}_i^2$, vir $N \to \infty$, asimptoties 'n χ^2-verdeling met $(k-1)$ g.v.v. besit [9].

Opmerkings.

1). Indien (2.42) vervang word deur (2.47), geld die stelling ook, omdat (2.47) voldoende is vir die geldigheid van (2.42).

2). Volgens die teorie van die χ^2-verdeling moet vir T_k geld: $\lim\limits_{N \to \infty} E(T_k) = (k-1)$ en

$$\lim_{N \to \infty} var(T_k) = 2(k-1).$$

Eersgenoemde is in lemma 2.4 aangetoon en laasgenoemde kan soos volg bewys word:

Omdat die variant t_i asimptoties normaal verdeel is as $N \to \infty$ (stelling 2.3 opmerking 3), volg dat

$$(2.53)\quad E(t_i - Et_i)^4 [var(t_i)]^{-2} = \mu_{4(i)} / \mu_{2(i)}^2 \overset{def}{=} \beta_{2(i)} = 3 + o(1)$$

waar $\mu_{r(i)}$ die r^{de} orde sentrale moment van t_i is.

Omdat t_i en $t_{i'}$ asimptoties gesamentlik 'n tweeveranderlike normaalverdeling besit (sien die bewys van hierdie stelling), volg uit KENDALL en STUART (1958) vol.I p. 83 dat

$$(2.54)\quad E(t_i - Et_i)^2 (t_{i'} - Et_{i'})^2 \overset{def}{=} \mu_{22(i,i')} = (1 + 2\rho_{ii'}^2)\sigma_i^2 \sigma_{i'}^2 (1 + o(1))$$

waar $\rho_{ii'}$ die (eksakte) korrelasiekoëffisiënt is tussen t_i en $t_{i'}$, $\sigma_i^2 = var(t_i)$ en $\sigma_{i'}^2 = var(t_{i'})$.

Gevolglik is

$$(2.55)\quad \frac{E[(t_i - Et_i)^2 (t_{i'} - Et_{i'})^2]}{var(t_i) var(t_{i'})} = 1 + 2\rho_{ii'}^2 + o(1).$$

9). Vergelyk ook lemma 1.1 opmerking 1 en CRAMÉR(1946) p.419.

Maar

(2.56) $\rho^2_{ii'} = [\text{kov}(t_i, t_{i'})]^2[\text{var}(t_i)\text{var}(t_{i'})]^{-1}$

$$= \frac{\nu_{Ni}\nu_{Ni'}}{(1-\nu_{Ni})(1-\nu_{Ni'})} \text{ m.b.v. (2.18), (2.22) en (2.32).}$$

Nou volg:

(2.57) $\text{var}(T_k) = E(T_k^2) - [E(T_k)]^2$

$= E[\sum_i (N-n_i)N^{-1}(t_i-Et_i)^2(\text{var } t_i)^{-1}]^2 - (k-1)^2$ uit lemma 2.4.

$$= \sum_i (1-\nu_{Ni})^2 E(t_i-Et_i)^4[\text{var}(t_i)]^{-2} - (k-1)^2 + \sum_{i\neq i'}\sum (1-\nu_{Ni})(1-\nu_{Ni'}) \frac{E(t_i-Et_i)^2(t_{i'}-Et_{i'})^2}{\text{var}(t_i)\text{var}(t_{i'})}$$

$$= 3\sum_i(1-\nu_{Ni})^2 - (k-1)^2 + \sum_{i\neq i'}\sum(1-\nu_{Ni})(1-\nu_{Ni'})[1+\frac{2\nu_{Ni}\nu_{Ni'}}{(1-\nu_{Ni})(1-\nu_{Ni'})}] + o(1)$$

$$= 2\sum_i(1-\nu_{Ni})^2+\sum_{ii'}\sum(1-\nu_{Ni})(1-\nu_{Ni'})+2\sum_{i\neq i'}\sum\nu_{Ni}\nu_{Ni'} - (k-1)^2+o(1)$$

$$= 2\sum_i(1-2\nu_{Ni}+\nu^2_{Ni})+(k-1)^2+2\sum_{i\neq i'}\sum\nu_{Ni}\nu_{Ni'} - (k-1)^2+o(1)$$

$$= 2(k-1) + o(1).$$

3). Indien aan voorwaardes (2.40) en (2.42) voldoen word en p van die ν_i gelyk is aan nul, dan volg op 'n soortgelyke wyse as in stelling 2.5 dat $T_{k-p} = \sum \hat{t}_i^2$, waarby die sommasie gaan oor dié (k-p) indekse uit 1,2,...,k waarvoor $\nu_i>0$, asimptoties 'n χ^2-verdeling besit met (k-p-1) g.v.v. as $N\to\infty$. Soos in lemma 2.4 kan bewys word dat $E(T_{k-p}) = (k-p-1)$.

2.4. OMSKRYWING VAN T_k.

(2.58) $T_k = \sum_i \hat{t}_i^2$

$$= \sum_i (N-n_i)N^{-1}(t_i-Et_i)^2[\text{var}(t_i)]^{-1}$$

$$= (N-1)(\sum_i t_i^2/n_i - N\bar{\Psi}_N^2)[\sum_h(a_h-\bar{\Psi}_N)^2]^{-1}.$$

2.5. SPESIALE GEVALLE VAN T_k.

2.5.1. Inleiding.

Deur spesiale gevalle van die funksie $\Psi_N(\delta)$ te neem, kan toetse vir verskuiwing, verskil in verspreiding, skeefheid, ens. uit T_k afgelei word. Ons bespreek alleen toetse vir verskuiwing en verskil in verspreiding. Die toetse vir verskil in verspreiding is alleen sinvol as aangeneem word dat 'n lokaliteitsparameter (bv. gemiddeld, mediaan) van die populasies gelyk is.

2.5.2. Toetse vir Verskuiwing.

2.5.2.1. $\Psi_N(\delta_{ij}) \equiv \delta_{ij}$ (vergelyk WILCOXON (1945) en STOKER (1955)).

Hier is

(2.59) $t_i = \sum_j r_{ij}/(N+1)$ uit (2.1),

(2.60) $E(t_i) = n_i N^{-1}\sum_h h/(N+1)$ uit (2.14)

$$= \tfrac{1}{2}n_i.$$

STOKER (1955) het vir hierdie geval bewys dat

$$(N+1)^2\sigma_a^2 = (N^3-\sum_\alpha g_\alpha^3)/12N \quad \text{sodat}$$

(2.61) $$\sigma_a^2 = [N(N^2-1)+\sum_\alpha g_\alpha-\sum_\alpha g_\alpha^3][12N(N+1)^2]^{-1}$$

$$= \frac{N-1}{12(N+1)} - \frac{\sum_\alpha g_\alpha(g_\alpha^2-1)}{12N(N+1)^2}$$

$$= \frac{N-1}{12(N+1)} - \frac{\gamma}{12N(N+1)^2} \quad \text{waar}$$

(2.62) $$\gamma = \sum_\alpha g_\alpha(g_\alpha^2-1).$$

Uit (2.18) volg nou dat

(2.63) $$\operatorname{var}(t_i) = \frac{n_i(N-n_i)}{12(N+1)}\left[1-\frac{\gamma}{N(N^2-1)}\right]$$

$$= \frac{n_i(N-n_i)}{12(N+1)}.\beta \quad \text{waar}$$

(2.64) $$\beta = 1 - \frac{\gamma}{N(N^2-1)}.$$

Die toetsingsgrootheid T_k word nou

$$(2.65)\quad T_W = \sum_i \frac{12(N+1)}{Nn_i\beta}(t_i-\tfrac{1}{2}n_i)^2 \quad \text{uit (2.4) en (2.5)}$$

$$= \frac{12(N+1)}{N\beta}\sum_i t_i^2/n_i - \frac{3(N+1)}{\beta}\,.$$

Omdat $0<h/(N+1)<1$, is $\underset{h}{\text{maks}}(\Psi_{Nh}-\bar{\Psi}_N)^2 = O(1)$.

Daar word dus voldoen aan voorwaarde (2.42).

Indien

(2.40) (Iedere diskontinuïteit van $F(x)$) $\leq \theta < 1$ en

(2.44) $\nu_i>0$ vir $i=1,2,\ldots,k$,

dan besit T_W asimptoties 'n χ^2-verdeling met $(k-1)$ g.v.v. as $N\rightarrow\infty$.

Opmerkings.

1). Bostaande is 'n uitbreiding van WILCOXON(1945) se toets van twee na k steekproewe.

2). Stel $R_i = \sum_j r_{ij}$ en $t_i = R_i/(N+1)$, dan volg uit (2.63) dat

$$(2.66)\quad \text{var}(R_i) = \frac{n_i(N-n_i)(N+1)}{12}\left[1-\frac{\gamma}{N(N^2-1)}\right]$$

wat dieselfde is as die uitdrukking deur KRUSKAL (1952) verkry. (Vergelyk ook RIJKOORT (1952) p. 395).

Vervolgens herlei (2.65) na:

$$(2.67)\quad T_W = \frac{12}{N(N+1)}\left[1-\frac{\gamma}{N(N^2-1)}\right]^{-1}\sum_i\frac{R_i^2}{n_i} - 3(N+1)\left[1-\frac{\gamma}{N(N^2-1)}\right]^{-1}$$

$= H^*$ van KRUSKAL (1952).

2.5.2.2. $\Psi_N(\delta_{ij}) \equiv \Xi_{r_{ij}} \equiv E(\xi_{r_{ij}})$

waar ξ_h $(\xi_1<\xi_2<\ldots<\xi_N)$ die waardes in rang van 'n ewekansige steekproef van grootte N uit 'n normaal (0,1) populasie is (vergelyk TERRY (1952)).

In hierdie geval is:

$$(2.68)\quad t_i = \sum_j \Xi_{r_{ij}}$$

$$(2.69)\quad E(t_i) = n_iN^{-1}\sum_h \Xi_h$$

$= 0$ weens simmetrie.

Verderis

(2.70) $\text{var}(t_i) = n_i(N-n_i)N^{-1}(N-1)^{-1} \sum_h a_h^2$

waar a_h in (2.11) gedefinieer is met $\Psi_N(\delta_h) = \Xi_h$.

As toetsingsgrootheid volg nou uit (2.58):

(2.71) $T_T = \dfrac{(N-1)}{\sum_h a_h^2} \sum_i \dfrac{t_i^2}{n_i}$.

Vir die normaal (0,1) verdeling geld (sien STOKER (1955) p.57):

(2.72) $\max_h (\Psi_{Nh}-\bar{\Psi}_N)^2 = \max_h (\Xi_h - \bar{\Xi})^2$

$= \max_h \Xi_h^2$

$= O(2 \log_e N)$

waar $\bar{\Xi} = N^{-1} \sum_h \Xi_h = 0$ sodat (2.42) bevredig word.

Onder die voorwaardes

(2.40) (Iedere diskontinuïteit van $F(x)$) $\leq \Theta < 1$ en

(2.44) $\nu_i > 0$ vir $i=1,2,\ldots,k$

besit T_T asimptoties, vir $N\to\infty$, 'n χ^2-verdeling met $(k-1)$ g.v.v.

Opmerking.

Meer algemeen kan ons beweer dat as ξ_h, $h=1,2,\ldots,N$ die waardes in rang is uit 'n willekeurige populasie met streng stygende kontinue verdelingsfunksie met kontinue tweede afgeleide en met eindige derde orde absolute sentrale moment, bly die bewerings van §2.5.2.2 geldig. Hierdie resultaat volg m.b.v. hulpstelling 3.1 en stelling 2.8 van STOKER (1955) waaruit volg dat (2.47) bevredig word en dus (2.42) - sien opmerking 7 na stelling 2.3.

2.5.2.3. $\Psi_N(\delta_{ij}) \equiv Q(\delta_{ij})$ waar $Q(q)$ die q^{de} kwantiel van 'n normaal (0,1) -verdeling is(vergelyk VAN DER WAERDEN (1957)).

Hiervoor is:

(2.73) $t_i = \sum_j Q[r_{ij}/(N+1)]$,

(2.74) $E(t_i) = n_i N^{-1} \sum_h Q(\delta_h)$

$= 0$ uit simmetrie (sien STOKER(1955) p. 31),

(2.75) $\mathrm{var}(t_i) = n_i(N-n_i)N^{-1}(N-1)^{-1} \sum_h a_h^2$ waar a_h gedefinieer is in (2.11) met $\Psi_N(\delta_h) = Q(\delta_h)$.

Die toetsingsgrootheid is nou:

(2.76) $$T_Q = \frac{(N-1)}{\sum_h a_h^2} \sum_i \frac{t_i^2}{n_i} \quad .$$

Uit STOKER (1955) p. 57 volg nou vir die normaal (0,1) -verdeling:

(2.77) $$\begin{aligned}\underset{h}{\mathrm{maks}}(\Psi_{Nh}-\bar{\Psi}_N)^2 &= \underset{h}{\mathrm{maks}}[Q(\delta_h)-\bar{Q}]^2 \\ &= \underset{h}{\mathrm{maks}}\ Q^2(\delta_h) \\ &= O(2\log_e(N+1))\end{aligned}$$

waar $\bar{Q} = N^{-1} \sum_h Q(\delta_h) = 0$.

Daar word dus voldoen aan voorwaarde (2.42).

Indien

(2.40) (Iedere diskontinuïteit van $F(x)$) $\leq \Theta < 1$ en

(2.44) $\nu_i > 0$ vir $i=1,2,\ldots,k$,

dan besit T_Q asimptoties, vir $N \to \infty$, 'n χ^2-verdeling met (k-1) g.v.v.

2.5.3. Toetse vir Verskil in Verspreiding.

Tot dusver is deurgaans veronderstel dat $\Psi_N(\delta)$ monotoon stygend is in δ, $0<\delta<1$. Die voorgaande stellings kan egter maklik aangepas word om gevalle in te sluit waar $\Psi_N(\delta)$ byvoorbeeld monotoon dalend is, of monotoon dalend oor 'n deel van die interval (0,1) en monotoon stygend oor die res van die interval, of omgekeerd.

Lemma 2.7. (Uitbreiding van STOKER (1955) stelling 2.4).

Stel dat, vir iedere N, die funksie $\Psi_N(\delta)$ kontinu en monotoon in δ is vir $0\leq\delta\leq 1$ (met „monotoon" word hier bedoel i) stygend of ii) dalend of iii) stygend oor die eerste deel van die interval (0,1) en dalend oor die tweede deel van die interval, of omgekeerd) en dat daar 'n kontinue en streng monotone funksie $\Psi(\delta)$ in δ vir $0<\delta<1$ bestaan (sien omskrywing van „monotoon" hierbo) sodanig dat vir alle δ met $0<\delta<1$ geld $\lim_{N\to\infty} \Psi_N(\delta) = \Psi(\delta)$ en wel, vir iedere $\varepsilon>0$, gelykmatig op $\varepsilon\leq\delta\leq 1-\varepsilon$. Hierby word veronderstel dat, in geval iii), as die funksie $\Psi_N(\delta)$ in die punt $\delta = \delta_{N_0}$, $0 < \delta_{N_0} < 1$, die „keerpunt" (maksimumpunt of minimumpunt) bereik en $\Psi(\delta)$ in die punt $\delta = \delta_0$, $0<\delta_0<1$ die keerpunt bereik, dat $|\delta_{N_0} - \delta_0| \xrightarrow[\infty]{N} 0$ en wel gelykmatig in N.

Indien $\lim_{N\to\infty} \bar{\Psi}_N$ (sien (2.17)) bestaan en gelyk is aan a(eindig) en a in die bereik lê van $\Psi(\delta)$ oor die gebiede $0<\delta<\delta_0$ en $\delta_0<\delta<1$ en indien iedere diskontinuïteit van F(x) hoogstens gelyk is aan 'n konstant $\Theta < 1$, dan bestaan 'n positiewe konstante c^2 sodanig dat

$$\liminf_{N\to\infty} \sigma^2_{F,a}(g_1, g_2, \ldots, g_\lambda) \geq c^2 \quad \text{spr } 0$$

en wel is c^2 gelykmatig van 0 geskei op alle klasse van verdelings F(x) waarvoor die grootste diskontinuïteit gelykmatig van 1 geskei bly.

Bewys: Die stelling volg op 'n soortgelyke wyse as dié van STOKER (1955) stelling 2.4 met onder andere die volgende aanpassings:

Laat $\Psi(\delta_{01}) = a = \Psi(\delta_{02})$ met $0<\delta_{01}<\delta_{02}<1$.

Kies $\varepsilon<\min(\delta_{01}, 1-\delta_{02}, \frac{1-\Theta}{8})$. Verdeel die gebied $(\varepsilon, 1-\varepsilon)$ in vier dele, nl. $(\varepsilon,\delta_{01}),(\delta_{01},\delta_0),(\delta_0,\delta_{02})$ en $(\delta_{02}, 1-\varepsilon)$.

2.5.3.1. $\Psi_N(\delta_{ij}) \equiv \delta_{ij}$ waarby rangnommers nou van weerskante af toegeken word. Hierdie prosedure is deur FREUND en ANSARI (1957) gevolg vir die twee-steekproef geval, en is die volgende: Die waarnemings word in stygende volgorde gerangskik en rangnommers word van weerskante af toegeken. Hierdeur is die stel waarnemings (sê 2n in aantal) in twee dele verdeel, elk waarvan rangnommers 1,2,...,n bevat. (Soortgelyk vir 'n onewe aantal 2n+1 waarby die middelste een die rangnommer n+1 het - sien later). Die som van al die rangnommers van die een steekproef word dan as toetsingsgrootheid geneem - sien ANSARI en BRADLEY (1960).

Bogenoemde metode ten opsigte van die toekenning van rangnommers word nou ook in hierdie studie ingevoer. Omdat die rangnommers weereens natuurlike getalle is, geld alle resultate van §2.1 tot §2.4 met die nodige aanpassing (sien lemma 2.7).

In hierdie geval is

$$(2.78)\quad t_i = \sum_j r_{ij}/(N+1).$$

a). N ewe.

$$(2.79)\quad E(t_i) = n_i N^{-1} \sum_{h=1}^{N} r_h/(N+1)$$

$$= \frac{n_i}{N(N+1)} \cdot 2 \sum_{h=1}^{\frac{1}{2}N} h$$

$$= \frac{n_i(N+2)}{4(N+1)} .$$

Vir var(t_i) bereken ons eers:

$$(2.80)\quad \sigma_c^2 = N^{-1} \sum_{h=1}^{N} [r_h/(N+1) - N^{-1} \sum_{h=1}^{N} r_h/(N+1)]^2 \text{ uit } (2.20)$$

$$= N^{-1} \sum_h [r_h/(N+1) - (N+2)/4(N+1)]^2$$

$$= \frac{1}{16N(N+1)^2} \sum_{h=1}^{N} [16r_h^2 - 8(N+2)r_h + (N+2)^2]$$

$$= \frac{1}{16N(N+1)^2} [32 \sum_{h=1}^{\frac{1}{2}N} h^2 - 16(N+2) \sum_{h=1}^{\frac{1}{2}N} h + N(N+2)^2]$$

$$= \frac{(N^2-4)}{48(N+1)^2} .$$

Stel nou dat die lengtes van die knope in volgorde gegee word deur $g_1, g_2, \ldots, g_\beta, g_{\beta+1}, \ldots, g_\lambda$ waar die $\frac{1}{2}N^{\text{de}}$

waarneming in die knoop van lengte g_β en die $(\frac{1}{2}N+1)^{de}$ waarneming in die knoop van lengte $g_{\beta+1}$ bevat is, d.w.s. $\sum_{\alpha=1}^{\beta} g_\alpha = \frac{1}{2}N$ en $\sum_{\alpha=\beta+1}^{\lambda} g_\alpha = \frac{1}{2}N$. (Indien die waarnemings in genoemde twee knope gelyk is, word dit nogtans as twee verskillende knope beskou. Ongelyke waardes word beskou as knope van lengte een).

Definieer

(2.81) $G_\alpha = g_1+g_2+\ldots+g_\alpha$ waar $\alpha \in \{1,2,\ldots,\beta\}$ en laat $G_o = 0$.

Stel

(2.82) $G'_{\alpha+1} = g_{\alpha+1}+g_{\alpha+2}+\ldots+g_\lambda$ waar $\alpha \in \{\beta+1,\beta+2,\ldots,\lambda\}$ en $G'_{\lambda+1} = 0$.

Nou is m.b.v. (2.16) en (2.20):

(2.83) $(N+1)^2(\sigma_c^2-\sigma_a^2) = (N+1)^2N^{-1}[\sum_h \Psi_{Nh}^2 - \sum_h a_h^2]$ met $\Psi_N(\delta_h) \equiv \delta_h$

$$= (N+1)^2N^{-1}\Big[\sum_{h=1}^{\frac{1}{2}N}\Psi_{Nh}^2+\sum_{h=\frac{1}{2}N+1}^{N}\Psi_{Nh}^2-\sum_{h=1}^{\frac{1}{2}N}a_h^2-\sum_{h=\frac{1}{2}N+1}^{N}a_h^2\Big]$$

$$= \frac{(N+1)^2}{N}\Big[\sum_{\alpha=1}^{\beta}\sum_{\mu=1}^{g_\alpha}\left(\frac{G_{\alpha-1}+\mu}{N+1}\right)^2+\sum_{\alpha=\beta+1}^{\lambda}\sum_{\mu=1}^{g_\alpha}\left(\frac{G'_{\alpha+1}+\mu}{N+1}\right)^2 -$$

$$-\sum_{\alpha=1}^{\beta}g_\alpha\left(\frac{1}{g_\alpha}\sum_{\mu=1}^{g_\alpha}\frac{G_{\alpha-1}+\mu}{N+1}\right)^2-\sum_{\alpha=\beta+1}^{\lambda}g_\alpha\left(\frac{1}{g_\alpha}\sum_{\mu=1}^{g_\alpha}\frac{G'_{\alpha+1}+\mu}{N+1}\right)^2\Big]$$

$$= \frac{1}{N}\Bigg[\sum_{\alpha=1}^{\beta}\Big\{\sum_{\mu=1}^{g_\alpha}(G_{\alpha-1}+\mu)^2-g_\alpha^{-1}[\sum_{\mu=1}^{g_\alpha}(G_{\alpha-1}+\mu)]^2\Big\} +$$

$$+\sum_{\alpha=\beta+1}^{\lambda}\Big\{\sum_{\mu=1}^{g_\alpha}(G'_{\alpha+1}+\mu)^2-g_\alpha^{-1}[\sum_{\mu=1}^{g_\alpha}(G'_{\alpha+1}+\mu)]^2\Big\}\Bigg]$$

$$= \frac{1}{N}\Bigg[\sum_{\alpha=1}^{\beta}\Big\{\sum_{\mu=1}^{g_\alpha}(G_{\alpha-1}^2+2\mu G_{\alpha-1}+\mu^2)-g_\alpha^{-1}[\sum_{\mu=1}^{g_\alpha}(G_{\alpha-1}+\mu)\sum_{\mu'=1}^{g_\alpha}(G_{\alpha-1}+\mu')]\Big\}+$$

$$+\sum_{\alpha=\beta+1}^{\lambda}\Big\{\sum_{\mu=1}^{g_\alpha}(G'^{2}_{\alpha+1}+2\mu G'_{\alpha+1}+\mu^2)-g_\alpha^{-1}[\sum_{\mu=1}^{g_\alpha}(G'_{\alpha+1}+\mu)\sum_{\mu'=1}^{g_\alpha}(G'_{\alpha+1}+\mu')]\Big\}\Bigg]$$

$$= \frac{1}{N}\left[\sum_{\alpha=1}^{\beta}\left\{g_\alpha G^2_{\alpha-1} + G_{\alpha-1}g_\alpha(g_\alpha+1)+\tfrac{1}{6}g_\alpha(g_\alpha+1)(2g_\alpha+1) - \right.\right.$$

$$\left. -g_\alpha^{-1}[g_\alpha G_{\alpha-1}+\tfrac{1}{2}g_\alpha(g_\alpha+1)]^2\right\}+$$

$$+\sum_{\alpha=\beta+1}^{\lambda}\left\{g_\alpha G'^2_{\alpha+1}+G'_{\alpha+1}g_\alpha(g_\alpha+1)+\tfrac{1}{6}g_\alpha(g_\alpha+1)(2g_\alpha+1) - \right.$$

$$\left.\left. -g_\alpha^{-1}[g_\alpha G'_{\alpha+1}+\tfrac{1}{2}g_\alpha(g_\alpha+1)]^2\right\}\right]$$

$$= \frac{1}{N}\left[\sum_{\alpha=1}^{\beta}\left(\frac{1}{12}g_\alpha^3 - \frac{1}{12}g_\alpha\right) + \sum_{\alpha=\beta+1}^{\lambda}\left(\frac{1}{12}g_\alpha^3 - \frac{1}{12}g_\alpha\right)\right]$$

$$= \frac{1}{12N}\sum_{\alpha=1}^{\lambda} g_\alpha(g_\alpha^2 - 1)$$

wat dieselfde is as die ooreenkomstige uitdrukking in die geval waar rangnommers gewoonweg toegeken word (sien vergelyking (2.61)).

Met behulp van die uitdrukking vir σ_c^2 volg nou dat

$$(2.84)\quad \sigma_a^2 = \frac{(N^2-4)}{48(N+1)^2} - \frac{1}{12N(N+1)^2}\sum_\alpha g_\alpha(g_\alpha^2-1)$$

$$= \frac{N^3-4\sum_\alpha g_\alpha^3}{48N(N+1)^2}\ .$$

Uit (2.18) volg dus dat

$$(2.85)\quad \mathrm{var}(t_i) = \frac{n_i(N-n_i)(N^3-4\sum_\alpha g_\alpha^3)}{48N(N-1)(N+1)^2}$$

en uit (2.22) dat

$$(2.86)\quad \mathrm{kov}(t_i,t_{i'}) = -\frac{n_i n_{i'}(N^3-4\sum_\alpha g_\alpha^3)}{48N(N-1)(N+1)^2}\ ,\quad i\neq i'\ .$$

Opmerkings.

1). Indien daar geen knope voorkom nie, d.w.s. as alle $g_\alpha=1$, is

$$(2.87)\quad \mathrm{var}(t_i) = \frac{n_i(N-n_i)(N^2-4)}{48(N-1)(N+1)^2}\ .$$

2). Neem die geval van twee steekproewe. Stel $n_i=m$, $N-n_i=n$ en $(N+1)t_i=W$, dan herlei (2.79) na:

$$E(W) = m(m+n+2)/4$$

en $\mathrm{var}(t_i)$ uit (2.87) na:

$$\mathrm{var}(W) = \frac{mn(m+n+2)(m+n-2)}{48(m+n-1)}$$

wat presies dieselfde is as die resultate aangegee deur ANSARI en BRADLEY (1960).

Die toetsingsgrootheid vir N ewe, is

$$(2.88)\quad T_E = \frac{48(N-1)(N+1)^2}{N^3-4\sum_\alpha g_\alpha^3}\sum_i \frac{t_i^2}{n_i} - \frac{3N(N-1)(N+2)^2}{N^3-4\sum_\alpha g_\alpha^3}\ .$$

Let op dat $N^3-4\sum_\alpha g_\alpha^3 = 0$ alleenlik as daar slegs een knoop is in welke geval $g_1=\frac{1}{2}N$ en $g_2=\frac{1}{2}N$. Onder voorwaarde (2.40) kan dié geval met waarskynlikheid één nie voorkom nie.

Aan voorwaarde (2.42) word voldoen - sien §2.5.2.1.

Indien

(2.40) (Iedere diskontinuïteit van $F(x)$) $\leq \Theta < 1$ en

(2.44) $\nu_i>0$ vir $i=1,2,\ldots,k$,

dan besit T_E asimptoties, vir $N\to\infty$, 'n χ^2-verdeling met $(k-1)$ g.v.v.

b). N onewe.

Nou is

$$(2.89)\quad E(t_i) = n_i N^{-1}\sum_{h=1}^{N} r_h/(N+1)$$

$$= \frac{n_i}{N(N+1)}\Big[\ 2\sum_{h=1}^{\frac{1}{2}(N-1)} h + \tfrac{1}{2}(N+1)\Big]$$

$$= \frac{n_i(N+1)}{4N}\ .$$

In hierdie geval is

$$(2.90)\quad \sigma_c^2 = N^{-1}\sum_h[r_h/(N+1) - N^{-1}\sum_h r_h/(N+1)]^2 \text{ uit } (2.20)$$

$$= \frac{1}{N}\Big[\sum_{h=1}^{N}\frac{r_h^2}{(N+1)^2} - \frac{1}{2N}\cdot\sum_{h=1}^{N} r_h + \frac{N(N+1)^2}{16N^2}\Big]$$

$$= \frac{1}{N}\Bigg[\ 2\sum_{h=1}^{\frac{1}{2}(N-1)}\frac{h^2}{(N+1)^2} + \frac{1}{4} - \frac{1}{N}\cdot\sum_{h=1}^{\frac{1}{2}(N-1)} h - \frac{N+1}{4N} + \frac{(N+1)^2}{16N}\ \Bigg]$$

$$= \frac{(N-1)(N^2+3)}{48N^2(N+1)} .$$

Neem nou aan dat die middelste waarde (d.i. die $\frac{1}{2}(N+1)^{de}$ waarde) as laaste of enigste waarde bevat word in die knoop van lengte g_β, dan volg op soortgelyke wyse as in (2.83) dat

$$(2.91)\quad (N+1)^2(\sigma_c^2-\sigma_a^2) = \frac{1}{12N} \sum_{\alpha=1}^{\lambda} g_\alpha(g_\alpha^2-1).$$

(Dieselfde resultaat volg as die $\frac{1}{2}(N+1)^{de}$ waarde bevat word in die knoop van lengte $g_{\beta+1}$).

Uit die voorafgaande twee resultate volg dat

$$(2.92)\quad \sigma_a^2 = \frac{(N-1)(N^2+3)}{48N^2(N+1)} - \frac{1}{12N(N+1)^2} \sum_\alpha g_\alpha(g_\alpha^2-1)$$

$$= \frac{N^4+6N^2-3-4N\sum_\alpha g_\alpha^3}{48N^2(N+1)^2} .$$

Uit (2.18) volg dus dat

$$(2.93)\quad \mathrm{var}(t_i) = \frac{n_i(N-n_i)(N^4+6N^2-3-4N\sum_\alpha g_\alpha^3)}{48(N-1)N^2(N+1)^2}$$

en uit (2.22) dat

$$(2.94)\quad \mathrm{kov}(t_i,t_{i'}) = -\frac{n_i n_{i'}(N^4+6N^2-3-4N\sum_\alpha g_\alpha^3)}{48N^2(N-1)(N+1)^2} , \quad i'\neq i.$$

Opmerkings.

1). Indien daar geen knope voorkom nie, is

$$(2.95)\quad \mathrm{var}(t_i) = \frac{n_i(N-n_i)(N^4+6N^2-3-4N^2)}{48N^2(N-1)(N+1)^2}$$

$$= \frac{n_i(N-n_i)(N^2+3)}{48N^2(N+1)} .$$

2). Neem die geval van twee steekproewe. Stel $n_i=m$, $N-n_i=n$ en $(N+1)t_i=W$, dan herlei $E(t_i)$ na:

$$E(W) = \tfrac{1}{4}m(m+n+1)^2/(m+n)$$

en $\mathrm{var}(t_i)$ na:

$$\mathrm{var}(W) = \frac{mn[(m+n)^2+3](m+n+1)}{48(m+n)^2}$$

wat dieselfde is as die resultate gegee deur ANSARI en BRADLEY (1960).

Die toetsingsgrootheid vir N onewe, is

$$(2.96)\quad T_o = \frac{48N(N-1)(N+1)^2}{N^4+6N^2-3-4N\sum\limits_\alpha g_\alpha^3} \sum_i \frac{t_i^2}{n_i} - \frac{3(N-1)(N+1)^4}{N^4+6N^2-3-4N\sum\limits_\alpha g_\alpha^3} .$$

Soos in geval a) word aan voorwaarde (2.42) voldoen.

Indien

(2.40) (Iedere diskontinuïteit van $F(x)$) $\leq \theta < 1$ en

(2.44) $\nu_i > 0$ vir $i=1,2,\ldots,k$,

dan besit T_o asimptoties, vir $N \to \infty$, 'n χ^2-verdeling met $(k-1)$ g.v.v.

Samevatting: T_E en T_o word gesamentlik gegee deur:

$$(2.97)\quad T_\delta = \frac{48N(N-1)(N+1)^2}{N^4-4N\sum\limits_\alpha g_\alpha^3+\delta(6N^2-3)} \sum_i \frac{t_i^2}{n_i} - \frac{3(N-1)(N+\delta)^2(N+2-\delta)^2}{N^4-4N\sum\limits_\alpha g_\alpha^3+\delta(6N^2-3)}$$

waar

$$(2.98)\quad \delta = \begin{cases} 1 & \text{as } N \text{ onewe is} \\ 0 & \text{as } N \text{ ewe is.} \end{cases}$$

Verder is

$$(2.99)\quad E(t_i) = \frac{n_i(N+2-\delta)}{4(N+1-\delta)} \qquad \text{en}$$

$$(2.100)\quad \mathrm{var}(t_i) = \frac{n_i(N-n_i)(N^4-4N\sum\limits_\alpha g_\alpha^3+\delta[6N^2-3])}{48N^2(N-1)(N+1)^2}$$

2.5.3.2. $\Psi_N(\delta_{ij}) = (\delta_{ij}-\frac{1}{2})^2$ (gewone toekenning van rangnommers).

Hier is

$$(2.101)\quad t_i = \sum_j [r_{ij}/(N+1) - \tfrac{1}{2}]^2 ,$$

$$(2.102)\quad E(t_i) = n_i N^{-1} \sum_h \Psi_N(\delta_h)$$

$$= n_i N^{-1} \sum_h [h/(N+1) - \tfrac{1}{2}]^2$$

$$= \frac{n_i(N-1)}{12(N+1)} .$$

Uit (2.20) volg:

$$(2.103)\quad \sigma_c^2 = N^{-1}\sum_h[(\frac{h}{N+1}-\tfrac{1}{2})^2 - \frac{N-1}{12(N+1)}]^2$$

$$= \frac{(N^2-4)(N-1)}{180(N+1)^3} .$$

Soos in (2.83) volg dat

$$(N+1)^4(\sigma_c^2-\sigma_a^2)=N^{-1}\left[\sum_{\alpha=1}^{\lambda}\left\{\sum_{\mu=1}^{g_\alpha}\left(G_{\alpha-1}+\mu-\tfrac{1}{2}(N+1)\right)^4 - \right.\right.$$

$$\left.\left.-g_\alpha^{-1}\left(\sum_{\mu=1}^{g_\alpha}\left\{G_{\alpha-1}+\mu-\tfrac{1}{2}(N+1)\right\}^2\right)^2\right\}\right]$$

$$= \frac{1}{180N}\sum_\alpha g_\alpha(g_\alpha^2-1)[(8g_\alpha+11)(2g_\alpha+1)+60(G_{\alpha-1}-\frac{N+1}{2})(g_\alpha+1) +$$

$$+60(G_{\alpha-1}-\frac{N+1}{2})^2] .$$

Uit bostaande twee resultate volg dus:

$$(2.104)\quad \sigma_a^2 = \frac{(N^2-4)(N-1)}{180(N+1)^3} - \frac{1}{180N(N+1)^4}\sum_\alpha g_\alpha(g_\alpha^2-1)\ x$$

$$x[(8g_\alpha+11)(2g_\alpha+1)+60(G_{\alpha-1}-\frac{N+1}{2})(g_\alpha+1)+60(G_{\alpha-1}-\frac{N+1}{2})^2]$$

met

$$(2.105)\quad var(t_i) = \frac{n_i(N-n_i)}{(N-1)}\sigma_a^2 .$$

Die toetsingsgrootheid T_k word nou:

$$(2.106)\quad T_M = \frac{(N-1)}{N\sigma_a^2}\sum_i\frac{t_i^2}{n_i} - \frac{(N-1)^3}{144(N+1)^2\sigma_a^2}$$

met σ_a^2 soos gegee in (2.104).

Omdat $0 < h/(N+1) < 1$, is $\max_h(\Psi_{Nh}-\bar{\Psi}_N)^2 = O(1)$.

Daar word dus voldoen aan voorwaarde (2.42).

Indien

(2.40) (Iedere diskontinuïteit van $F(x)$) $\leq \theta < 1$ en

(2.44) $\nu_i > 0$ vir $i=1,2,..,k$,

volg uit stelling 2.5 dat T_M asimptoties, vir $N\to\infty$, 'n χ^2-verdeling met (k-1) g.v.v. besit.

Opmerkings.

1). Indien F kontinu is sodat met waarskynlikheid een geen knope voorkom nie, is alle $g_\alpha=1$ (met waarskynlikheid een) en kry ons:

$$(2.107) \quad \sigma_a^2 = \frac{(N^2-4)(N-1)}{180(N+1)^3} \quad ,$$

$$(2.108) \quad \mathrm{var}(t_i) = \frac{n_i(N-n_i)(N^2-4)}{180(N+1)^3} \quad \text{en}$$

$$(2.109) \quad T_M = \frac{180(N+1)^3}{N(N^2-4)} \sum_i \frac{t_i^2}{n_i} - \frac{5(N-1)^2(N+1)}{4(N^2-4)} \; .$$

In dié geval word ook aan voorwaarde (2.40) voldoen. T_M besit asimptoties, vir $N\to\infty$, 'n χ^2-verdeling met (k-1) g.v.v. indien voorwaarde (2.44) bevredig word.

2). T_M is 'n uitbreiding van MOOD (1954) se toetsingsgrootheid van twee na k steekproewe. Sien ook §3.6.3.3 waar die asimptotiese eienskappe van T_M bespreek word.

3). Vir die geval k=2 met $n_i=m$, $N-n_i=n$ en $t_i=M/(N+1)^{-2}$ is

$$E(M) = \frac{1}{12}\, m(N-1)(N+1) \text{ uit } (2.102) \quad \text{en}$$

$$\mathrm{var}(M) = \frac{1}{180}\, mn(N^2-4)(N+1) \text{ uit } (2.108)$$

wat dieselfde is as die ooreenkomstige resultate van MOOD (1954).

2.5.3.3. $\Psi_N(\delta_{ij}) = (\delta_{ij}-\frac{1}{2})^2$ (rangnommers van weerskante af toegeken).

Hiervoor is

$$(2.110) \quad t_i = \sum_j [r_{ij}/(N+1) - \tfrac{1}{2}]^2 .$$

Ons onderskei twee gevalle.

a). N ewe.

$$(2.111) \quad E(t_i) = n_i N^{-1} \sum_h [r_h/(N+1) - \tfrac{1}{2}]^2$$

$$= \frac{n_i}{N(N+1)^2} \sum_{h=1}^{N} [r_h^2 - (N+1)r_h + \tfrac{1}{4}(N+1)^2]$$

$$= \frac{n_i}{N(N+1)^2}\left[2\sum_{h=1}^{\frac{1}{2}N} h^2 - 2(N+1)\sum_{h=1}^{\frac{1}{2}N} h + \tfrac{1}{4}N(N+1)^2\right]$$

$$= \frac{n_i(N-1)}{12(N+1)} .$$

$$(2.112)\quad \sigma_c^2 = \frac{1}{N}\sum_h\left[\left(\frac{r_h}{N+1} - \tfrac{1}{2}\right)^2 - \frac{N-1}{12(N+1)}\right]^2$$

$$= N^{-1}(N+1)^{-4}\sum_h[r_h^2 - (N+1)r_h + \tfrac{1}{6}(N+1)(N+2)]^2$$

$$= N^{-1}(N+1)^{-4}\sum_h[r_h^4 - 2(N+1)r_h^3 + \tfrac{1}{3}(N+1)(4N+5)r_h^2 -$$

$$- \tfrac{1}{3}(N+1)^2(N+2)r_h + \tfrac{1}{36}(N+1)^2(N+2)^2]$$

$$= \frac{1}{N(N+1)^4}\left[2\sum_{h=1}^{\frac{1}{2}N} h^4 - 4(N+1)\sum_{h=1}^{\frac{1}{2}N} h^3 + \frac{2(N+1)(4N+5)}{3}\sum_{h=1}^{\frac{1}{2}N} h^2 -\right.$$

$$\left. - \frac{2(N+1)^2(N+2)}{3}\sum_{h=1}^{\frac{1}{2}N} h + \frac{N(N+1)^2(N+2)^2}{36}\right]$$

$$= \frac{1}{N(N+1)^4}\left[\frac{6(\frac{1}{2}N)^5 + 15(\frac{1}{2}N)^4 + 10(\frac{1}{2}N)^3 - \frac{1}{2}N}{15} -\right.$$

$$- (N+1)(\tfrac{1}{2}N)^2(\tfrac{1}{2}N+1)^2 + \frac{(N+1)(4N+5)}{18}N(\tfrac{1}{2}N+1)(N+1)$$

$$\left. - \tfrac{1}{3}(N+1)^2(N+2)(\tfrac{1}{2}N)(\tfrac{1}{2}N+1) + \frac{N(N+1)^2(N+2)^2}{36}\right]$$

$$= \frac{(N^2-4)(N-1)}{180(N+1)^3} .$$

Netsoos in §2.5.3.2 vind ons

$$(N+1)^4(\sigma_c^2 - \sigma_a^2) = \frac{1}{180N}\left\{\sum_{\alpha=1}^{\beta} g_\alpha(g_\alpha^2-1)[(8g_\alpha+11)(2g_\alpha+1) + \right.$$

$$+60(G_{\alpha-1} - \tfrac{N+1}{2})(g_\alpha+1) + 60(G_{\alpha-1} - \tfrac{N+1}{2})^2] +$$

$$+ \sum_{\alpha=\beta+1}^{\lambda} g_\alpha(g_\alpha^2-1)[(8g_\alpha+11)(2g_\alpha+1) +$$

$$\left. +60(G'_{\alpha+1} - \tfrac{N+1}{2})(g_\alpha+1) + 60(G'_{\alpha+1} - \tfrac{N+1}{2})^2]\right\}$$

waar $G'_{\alpha+1}$ in (2.82) gedefinieer is en β dieselfde betekenis het as in §2.5.3.1.

Nou is

$$(2.113)\quad \sigma_a^2 = \frac{(N^2-4)(N-1)}{180(N+1)^3} - \frac{1}{180N(N+1)^4}\Big\{\sum_{\alpha=1}^{\beta} g_\alpha(g_\alpha^2-1)[(8g_\alpha+11).$$

$$.(2g_\alpha+1)+60(G_{\alpha-1}-\tfrac{N+1}{2})(g_\alpha+1)+60(G_{\alpha-1}-\tfrac{N+1}{2})^2] +$$

$$+\sum_{\alpha=\beta+1}^{\lambda} g_\alpha(g_\alpha^2-1)[(8g_\alpha+11)(2g_\alpha+1) +$$

$$+60(G'_{\alpha+1}-\tfrac{N+1}{2})(g_\alpha+1)+60(G'_{\alpha+1}-\tfrac{N+1}{2})^2]\Big\}$$

en

$$(2.114)\quad \mathrm{var}(t_i) = \frac{n_i(N-n_i)\sigma_a^2}{(N-1)} .$$

b). N onewe.

$$(2.115)\quad E(t_i) = n_i N^{-1}(N+1)^{-2}\sum_{h=1}^{N}[r_h^2-(N+1)r_h+\tfrac{1}{4}(N+1)^2]$$

$$= \frac{n_i}{N(N+1)^2}\Big[2\sum_{h=1}^{\frac{1}{2}(N-1)} h^2+(\tfrac{N+1}{2})^2-2(N+1)\sum_{h=1}^{\frac{1}{2}(N-1)} h -$$

$$-\tfrac{1}{2}(N+1)^2+\tfrac{N}{4}(N+1)^2\Big]$$

$$= \frac{n_i(N-1)}{12(N+1)} .$$

$$(2.116)\quad \sigma_c^2 = \frac{1}{N(N+1)^4}\sum_{h=1}^{N}\Big[r_h^4-2(N+1)r_h^3+\tfrac{1}{3}(N+1)(4N+5)r_h^2 -$$

$$-\tfrac{1}{3}(N+1)^2(N+2)r_h+\frac{(N+1)^2(N+2)^2}{36}\Big]$$

$$= \frac{1}{N(N+1)^4}\Big[2\sum_{h=1}^{\frac{1}{2}(N-1)} h^4+(\tfrac{N+1}{2})^4-4(N+1)\sum_{h=1}^{\frac{1}{2}(N-1)} h^3-\tfrac{1}{4}(N+1)^4$$

$$+\frac{2(N+1)(4N+5)}{3}\sum_{h=1}^{\frac{1}{2}(N-1)} h^2+\frac{(N+1)(4N+5)}{3}(\tfrac{N+1}{2})^2$$

$$-\frac{2(N+1)^2(N+2)}{3}\sum_{h=1}^{\frac{1}{2}(N-1)} h-\frac{(N+1)^3(N+2)}{6}+\frac{N(N+1)^2(N+2)^2}{36}\Big]$$

$$= \frac{(N^2-4)(N-1)}{180(N+1)^3} .$$

Sowel $E(t_i)$ as σ_c^2 is dieselfde as in geval a) waarby N ewe is. Bewys kan word dat ook $\sigma_c^2-\sigma_a^2$ dieselfde uitdrukking lewer (sien bv. §2.5.3.1.b.). Gevolglik is σ_a^2 en $\mathrm{var}(t_i)$ dieselfde as in (2.113) en (2.114).

Ons kry as toetsingsgrootheid

$$(2.117)\quad T_M' = \frac{(N-1)}{N\sigma_a^2}\sum_i \frac{t_i^2}{n_i} - \frac{(N-1)^3}{144(N+1)^2\sigma_a^2}$$

met σ_a^2 soos gegee in (2.113).

Soos in §2.5.3.2. besit T_M' onder voorwaardes (2.40) en (2.44) asimptoties, vir $N\to\infty$, 'n χ^2-verdeling met (k-1) g.v.v.

Opmerkings.

1). Indien alle $g_\alpha=1$, kry ons

$$(2.118)\quad \sigma_a^2 = \frac{(N^2-4)(N-1)}{180(N+1)^3},$$

$$(2.119)\quad \mathrm{var}(t_i) = \frac{n_i(N-n_i)(N^2-4)}{180(N+1)^3} \quad \text{en}$$

$$(2.120)\quad T_M' = \frac{180(N+1)^3}{N(N^2-4)}\sum_i \frac{t_i^2}{n_i} - \frac{5(N-1)^2(N+1)}{4(N^2-4)}$$

wat presies dieselfde is as die ooreenkomstige uitdrukkings in §2.5.3.2.

2). Vir die geval dat $\Psi_N(\delta_{ij}) = (\delta_{ij}-\frac{1}{2})^2$ maak dit nie saak op watter van bogenoemde twee maniere rangnommers toegeken word nie. Inderdaad kan bewys word dat die twee gevalle identies dieselfde toetsingsgrootheid lewer.

2.5.3.4. $\Psi_N(\delta_{ij}) = \delta_{ij}^2$ (gewone toekenning van rangnommers).

Hier is

$$(2.121)\quad t_i = \sum_j r_{ij}^2/(N+1)^2,$$

$$(2.122)\quad E(t_i) = n_i N^{-1}(N+1)^{-2}\sum_h h^2$$

$$= \frac{n_i(2N+1)}{6(N+1)} \quad \text{en}$$

$$(2.123)\quad \sigma_c^2 = N^{-1}(N+1)^{-4}\sum_h [h^2 - \tfrac{1}{6}(N+1)(2N+1)]^2$$

$$= \frac{(2N+1)(8N+11)(N-1)}{180(N+1)^3}.$$

Op analoë wyse as in die vorige gevalle kry ons:

$$(N+1)^4(\sigma_c^2-\sigma_a^2) = \frac{1}{180N}\sum_{\alpha=1}^{\lambda} g_\alpha(g_\alpha^2-1)[(8g_\alpha+11)(2g_\alpha+1)+60G_{\alpha-1}(g_\alpha+1)+ +60G_{\alpha-1}^2]$$

sodat

$$(2.124)\quad \sigma_a^2 = \frac{(2N+1)(8N+11)(N-1)}{180(N+1)^3} - \frac{1}{180N(N+1)^4}\sum_{\alpha=1}^{\lambda} g_\alpha(g_\alpha^2-1)\ \times$$

$$\times\ [(8g_\alpha+11)(2g_\alpha+1)+60G_{\alpha-1}(g_\alpha+1)+60G_{\alpha-1}^2]\ ,$$

$$(2.125)\quad \mathrm{var}(t_i) = \frac{n_i(N-n_i)\sigma_a^2}{(N-1)}\ .$$

Die toetsingsgrootheid T_k word nou:

$$(2.126)\quad T_R = \frac{(N-1)}{N\sigma_a^2}\left[\sum_i \frac{t_i^2}{n_i} - \frac{N(2N+1)^2}{36(N+1)^2}\right]$$

met σ_a^2 soos gegee in (2.124).

Indien aan voorwaardes (2.40) en (2.44) voldoen word, besit T_R asimptoties, vir $N\to\infty$, 'n χ^2-verdeling met (k-1) g.v.v.

Opmerkings.

1). Indien alle $g_\alpha=1$, is

$$(2.127)\quad \sigma_a^2 = \frac{(2N+1)(8N+11)(N-1)}{180(N+1)^3}\ ,$$

$$(2.128)\quad \mathrm{var}(t_i) = \frac{n_i(N-n_i)(2N+1)(8N+11)}{180(N+1)^3} \qquad \text{en}$$

$$(2.129)\quad T_R = \frac{180(N+1)^3}{N(2N+1)(8N+11)}\left[\sum_i \frac{t_i^2}{n_i} - \frac{N(2N+1)^2}{36(N+1)^2}\right]\ .$$

2). Omdat groot rangnommers in hierdie toets relatief baie meer gewig dra as klein rangnommers (die toets is gebaseer op die vierkante van die rangnommers), is hierdie toets nie juis sinvol as 'n toets vir verskil in verspreiding nie (sien ook §3.6.3.5).

2.5.3.5. $\Psi_N(\delta_{ij}) = \delta_{ij}^2$ (rangnommers van weerskante af toegeken).

Hier is

$$(2.130)\quad t_i = \sum_j r_{ij}^2/(N+1)^2.$$

Ons onderskei weer tussen N ewe en onewe.

a). N ewe.

(2.131) $$E(t_i) = n_i N^{-1}(N+1)^{-2} \sum_h r_h^2$$

$$= \frac{n_i}{N(N+1)^2} \cdot 2 \sum_{h=1}^{\frac{1}{2}N} h^2$$

$$= \frac{n_i(N+2)}{12(N+1)},$$

(2.132) $$\sigma_c^2 = N^{-1}(N+1)^{-4} \sum_{h=1}^{N} [r_h^2 - \frac{(N+1)(N+2)}{12}]^2$$

$$= \frac{1}{N(N+1)^4}[2 \sum_{h=1}^{\frac{1}{2}N} h^4 - \frac{1}{3}(N+1)(N+2) \sum_{h=1}^{\frac{1}{2}N} h^2 + \frac{(N+1)^2(N+2)^2 N}{144}]$$

$$= \frac{(N^2-4)(4N+11)}{720(N+1)^3}.$$

Soos tevore bereken ons $(\sigma_c^2 - \sigma_a^2)$ en kry dieselfde uitdrukking as in §2.5.3.4. Gevolglik is

(2.133) $$\sigma_a^2 = \frac{(N^2-4)(4N+11)}{720(N+1)^3} - \frac{1}{180N(N+1)^4}\left\{\sum_{\alpha=1}^{\beta} g_\alpha(g_\alpha^2-1) \times \right.$$

$$\times[(8g_\alpha+11)(2g_\alpha+1)+60G_{\alpha-1}(g_\alpha+1)+60G_{\alpha-1}^2] +$$

$$\left. + \sum_{\alpha=\beta+1}^{\lambda} g_\alpha(g_\alpha^2-1)[(8g_\alpha+11)(2g_\alpha+1)+60G'_{\alpha+1}(g_\alpha+1)+60G'^2_{\alpha+1}]\right\}.$$

Nou is

(2.134) $$\text{var}(t_i) = \frac{n_i(N-n_i)\sigma_a^2}{(N-1)}$$ met σ_a^2 soos in (2.133) gegee.

Ons het dus as toetsingsgrootheid

(2.135) $$T_G = \frac{(N-1)}{N\sigma_a^2}\left[\sum_i \frac{t_i^2}{n_i} - \frac{N(N+2)^2}{144(N+1)^2}\right]$$

wat onder voorwaardes (2.40) en (2.44) asimptoties χ^2 verdeel is met (k-1) g.v.v. as $N \to \infty$.

Opmerking.

Indien alle $g_\alpha = 1$, is

(2.136) $$\sigma_a^2 = \frac{(N^2-4)(4N+11)}{720(N+1)^3},$$

(2.137) $$\mathrm{var}(t_i) = \frac{n_i(N-n_i)(N^2-4)(4N+11)}{720(N+1)^3(N-1)}$$ en

(2.138) $$T_G = \frac{720(N-1)(N+1)^3}{N(N^2-4)(4N+11)} \sum_i \frac{t_i^2}{n_i} - \frac{5(N^2-1)(N+2)}{(N-2)(4N+11)} .$$

b). N onewe.

(2.139) $$E(t_i) = n_i N^{-1}(N+1)^{-2} \sum_{h=1}^{N} r_h^2$$

$$= \frac{n_i}{N(N+1)^2}\left[2 \sum_{h=1}^{\frac{1}{2}(N-1)} h^2 + \left(\frac{N+1}{2}\right)^2\right]$$

$$= \frac{n_i(N^2+2N+3)}{12N(N+1)} ,$$

(2.140) $$\sigma_c^2 = \frac{1}{N} \sum_{h=1}^{N} \left[\frac{r_h^2}{(N+1)^2} - \frac{N^2+2N+3}{12N(N+1)}\right]^2$$

$$= \frac{1}{N(N+1)^4}\left[2 \sum_{h=1}^{\frac{1}{2}(N-1)} h^4 + \left(\frac{N+1}{2}\right)^4 - \frac{(N+1)(N^2+2N+3)}{3N} \sum_{h=1}^{\frac{1}{2}(N-1)} h^2 \right.$$

$$\left. - \frac{(N+1)^3(N^2+2N+3)}{24N} + \frac{(N+1)^2(N^2+2N+3)^2}{144N}\right]$$

$$= \frac{(N-1)(4N^4+15N^3+59N^2+105N+45)}{720N^2(N+1)^3} .$$

$(\sigma_c^2-\sigma_a^2)$ is dieselfde as vir die geval N ewe, sodat:

(2.141) $$\sigma_a^2 = \frac{(N-1)(4N^4+15N^3+59N^2+105N+45)}{720N^2(N+1)^3} -$$

$$- \frac{1}{180N(N+1)^4}\left\{\sum_{\alpha=1}^{\beta} g_\alpha(g_\alpha^2-1)[(8g_\alpha+11)(2g_\alpha+1)+60(g_\alpha+1)G_{\alpha-1} + +60G_{\alpha-1}^2] + \right.$$

$$\left. + \sum_{\alpha=\beta+1}^{\lambda} g_\alpha(g_\alpha^2-1)[(8g_\alpha+11)(2g_\alpha+1)+60G'_{\alpha+1}(g_\alpha+1)+60G'^2_{\alpha+1}]\right\}.$$

Ons het as toetsingsgrootheid

(2.142) $$T_H = \frac{(N-1)}{N\sigma_a^2}\left[\sum_i \frac{t_i^2}{n_i} - \frac{(N^2+2N+3)^2}{144N(N+1)^2}\right]$$

wat onder voorwaardes (2.40) en (2.44) asimptoties χ^2 verdeel is met (k-1) g.v.v. as $N \to \infty$.

Opmerking.

Indien alle $g_\alpha=1$, is

$$(2.143)\quad \sigma_a^2 = \frac{(N-1)(4N^4+15N^3+59N^2+105N+45)}{720N^2(N+1)^3} \quad ,$$

$$(2.144)\quad \mathrm{var}(t_i)= \frac{n_i(N-n_i)(4N^4+15N^3+59N^2+105N+45)}{720N^2(N+1)^3} \qquad \text{en}$$

$$(2.145)\quad T_H= \frac{720N(N+1)^3}{4N^4+15N^3+59N^2+105N+45}\left[\sum_i \frac{t_i^2}{n_i} - \frac{(N^2+2N+3)^2}{144N(N+1)^2}\right] .$$

2.5.3.6. $\Psi_N(\delta_{ij}) = Q^2(\delta_{ij})$ (sien LEMMER en STOKER(1960) en KLOTZ(1962). Rangnommers word voortaan gewoonweg toegeken).

Hier is

$$(2.146)\quad t_i = \sum_j Q^2(\delta_{ij}),$$

$$(2.147)\quad E(t_i) = n_i N^{-1} \sum_h Q^2(\delta_h) \text{ en}$$

$$(2.148)\quad \mathrm{var}(t_i) = n_i(N-n_i)N^{-1}(N-1)^{-1} \sum_h (a_h-\bar{\Psi}_N)^2$$

waar a_h in (2.11) gedefinieer is met $\Psi_N(\delta_h) = Q^2(\delta_h)$.

Die toetsingsgrootheid is

$$(2.149)\quad T_{Q_2} = \frac{(N-1)}{\sum_h (a_h-\bar{\Psi}_N)^2} \sum_i n_i^{-1}[t_i-n_i N^{-1} \sum_h Q^2(\delta_h)]^2 .$$

Uit STOKER (1955) p. 57 volg:

$$\begin{aligned}\underset{h}{\mathrm{maks}}(\Psi_{Nh}-\bar{\Psi}_N)^2 &= \underset{h}{\mathrm{maks}}[Q^2(\delta_h)-\bar{Q}]^2 \text{ waar } \bar{Q} = N^{-1} \sum_h Q^2(\delta_h)\\ &\leq \underset{h}{\mathrm{maks}}[Q^4(\delta_h)]\\ &= O[2 \log_e(N+1)]^2\\ &= o(N) \text{ sodat (2.42) bevredig word.}\end{aligned}$$

Onder voorwaardes (2.40) en (2.44) besit T_{Q_2} asimptoties, vir $N\to\infty$, 'n χ^2-verdeling met $(k-1)$ g.v.v.

Opmerking.

T_{Q_2} is 'n uitbreiding van twee na k steekproewe van 'n toetsingsgrootheid bespreek deur LEMMER en STOKER (1960) en KLOTZ (1962).

2.5.3.7. $\Psi_N(\delta_{ij}) = E(\xi^2_{r_{ij}})$ - sien §2.5.2.2 en CROUSE (1960) §4.14.

Hier is

(2.150) $t_i = \sum_j E(\xi^2_{r_{ij}})$,

(2.151) $E(t_i) = n_i N^{-1} \sum_h E(\xi_h^2)$ en

(2.152) $\text{var}(t_i) = n_i(N-n_i)N^{-1}(N-1)^{-1} \sum_h (a_h - \bar{\Psi}_N)^2$

waar a_h in (2.11) gedefinieer is met $\Psi_N(\delta_h) = E(\xi_h^2)$.

Die toetsingsgrootheid T_k is in hierdie geval:

(2.153) $T_{T_2} = \dfrac{(N-1)}{\sum_h (a_h - \bar{\Psi}_N)^2} \sum_i n_i^{-1}[t_i - n_i N^{-1} \sum_h E(\xi_h^2)]^2.$

Omdat die ξ_h's rangparameters uit 'n normaal(0,1)-populasie is, geld vir $\xi_N = \text{maks}_h \{\xi_h\}$ dat

$$\xi_N = O_p(2 \log_e N)^{\frac{1}{2}} \quad \text{(CRAMÉR (1946) p. 374).}$$

$$\xi_N^2 = O_p(2 \log_e N),$$

$$E(\xi_N^2) = O(2 \log_e N),$$

$$\begin{aligned} \text{maks}_h (\Psi_{Nh} - \bar{\Psi}_N)^2 &\le \text{maks}_h \Psi^2_{Nh} \\ &= \text{maks}_h [E(\xi_N^2)]^2 \\ &= O(2 \log_e N)^2 \end{aligned}$$

sodat voorwaarde (2.42) dus bevredig word.

Onder die voorwaardes

(2.40) (Iedere diskontinuïteit van $F(x)$) $\le \theta < 1$ en

(2.44) $\nu_i > 0$ vir $i=1,2,\ldots,k$,

besit T_{T_2} asimptoties, vir $N \to \infty$, 'n χ^2-verdeling met $(k-1)$ g.v.v.

Opmerking.

T_{T_2} is 'n uitbreiding van twee na k steekproewe van 'n toetsingsgrootheid bespreek deur CAPON (1961) § 5 (C).

2.6. ALGEMENE GEVAL.

2.6.1. Inleiding.

Neem vervolgens die meer algemene funksie $\Psi_N(x_{ij},\delta_{ij})$, wat eindig veronderstel word vir alle eindige x_{ij} en $0<\delta_{ij}<1$.

Definieer

$$(2.154)\quad a_h = g_\alpha^{-1} \sum_{\mu=1}^{g_\alpha} \Psi_N\left[z_{G_{\alpha-1}+\mu} , \frac{G_{\alpha-1}+\mu}{N+1} \right]$$

vir $G_{\alpha-1}+1 \leq h \leq G_\alpha$,

$$(2.155)\quad a_{ij} = a_h \text{ indien } x_{ij} = z_h \text{ en}$$

$$(2.156)\quad \bar{\Psi}_N = N^{-1} \sum_h a_h .$$

Neem weer

$$(2.157)\quad t_i = \sum_j a_{ij},$$

$$(2.158)\quad \hat{t}_i = (N-n_i)^{\frac{1}{2}}(t_i-Et_i)[N\text{var}(t_i)]^{-\frac{1}{2}} \quad \text{en}$$

$$(2.159)\quad T_k = \sum_i \hat{t}_i^2 .$$

Laat $Z = (z_1,z_2,\ldots,z_N)$, die waargenome stel waardes wees.

Soos in lemma 2.1 kan bewys word (sien ook STOKER (1955) p.27) dat:

$$(2.160)\quad E(t_i|H_{o\sigma}; Z) = n_i N^{-1} \sum_h a_h .$$

Die verwagtingswaarde is voorwaardelik.

Vir gegewe Z kan bewys word (sien lemma 2.2 en STOKER (1955) p. 27) dat:

$$(2.161)\quad \text{var}(t_i|H_{o\sigma}; Z) = n_i(N-n_i)N^{-1}(N-1)^{-1} \sum_h (a_h-\bar{\Psi}_N)^2$$

$$= n_i(N-n_i)(N-1)^{-1}\sigma_a^2 \quad \text{waar}$$

$$(2.162)\quad \sigma_a^2 = N^{-1} \sum_h (a_h-\bar{\Psi}_N)^2 \quad {}^{10)}.$$

Soortgelyk:

$$(2.163)\quad \text{kov}(t_i,t_{i'}|H_{o\sigma}; Z) = -n_i n_{i'}(N-1)^{-1}\sigma_a^2 .$$

Die meeste stellings wat bewys is vir die funksie $\Psi_N(\delta_{ij})$, geld voorwaardelik vir $\Psi_N(x_{ij},\delta_{ij})$. Ons bespreek

---

10). Ons veronderstel deurgaans $\sigma_a^2>0$ vir eindige N.

hulle kortliks.

STELLING 2.6.

Indien

$$(2.35)\quad \frac{\max_h (a_h-\bar{\Psi}_N)^2}{N^{-1}\sum_h (a_h-\bar{\Psi}_N)^2} = o_p(N) \text{ vir } N\to\infty,$$

(2.36) minstens twee van die c_i's in (2.31) verskillend is,

(2.44) $\nu_i>0$ vir $i=1,2,\ldots,k$,

dan is $\sum_i c_i t_i$ asimptoties normaal verdeel as $N\to\infty$.

Bewys: Stellings 2.1 en 2.2 geld voorwaardelik vir gegewe Z (STOKER (1955) p. 33 opmerking). Soos in stelling 2.3 volg dan dat $\sum_i c_i t_i$ asimptoties normaal verdeel is as $N\to\infty$ (sien HOEFFDING (1952) en STOKER (1955) p. 34).

Opmerkings.

1). Onder die voorwaarde

(2.164) $\liminf_{N\to\infty} .\ \sigma_a^2 \geq \varepsilon>0$ in waarskynlikheid,

is die voorwaarde

(2.165) $\max_h (a_h-\bar{\Psi}_N)^2 = o_p(N)$

voldoende vir die geldigheid van (2.35). Stelling 2.6 geld dus onder voorwaardes (2.36), (2.44), (2.164) en (2.165). Voldoende voorwaardes dat t_i asimptoties normaal verdeel is (sien stelling 2.3 opmerking 3), is: (2.44), (2.164) en (2.165).

2). Laasgenoemde drie voorwaardes is ook voldoende vir die geldigheid in waarskynlikheid van:

(2.49) $\lim_{N\to\infty} \text{var}(\hat{t}_i) > 0$ en

(2.50) $\lim_{N\to\infty} E|\hat{t}_i|^{2+\delta} < \infty$ vir $\delta>0$ (sien lemma 2.6).

STELLING 2.7.

Die asimptotiese momentematriks van die k variante $\underline{\hat{t}}_i$ is van rang (k-1) mits (2.44) geld en

(2.164) $\liminf_{N\to\infty} .\ \sigma_a^2 \geq \varepsilon>0$ in waarskynlikheid.

Bewys: Soos stelling 2.4.

Opmerking.

Hierdie stelling geld voorwaardelik vir gegewe Z.

STELLING 2.8.

Indien

(2.164) $\lim\limits_{N\to\infty} \inf.\ \sigma_a^2 \geq \varepsilon > 0$ in waarskynlikheid,

(2.165) $\text{maks}_h (a_h - \bar{\Psi}_N)^2 = o_p(N)$ en

(2.44) $\nu_i > 0$ vir $i=1,2,\ldots,k$,

dan besit

$$T_k = \sum_i (N-n_i)(t_i - Et_i)^2 [N\text{var}(t_i)]^{-1}$$

asimptoties, vir $N\to\infty$, 'n χ^2-verdeling met $(k-1)$ g.v.v.

Bewys: Soos stelling 2.5.

2.6.2. Spesiale geval.

Neem $\Psi_N(x_{ij},\delta_{ij}) = x_{ij}$ (vergelyk PITMAN (1937)).

Hiervoor is:

(2.166) $t_i = \sum_j x_{ij}$

(2.167) $E(t_i | H_{00}; Z) = n_i N^{-1} \sum_h z_h$

$= n_i \bar{z}$ waar

(2.168) $\bar{z} = N^{-1} \sum_h z_h$.

In dié geval is

(2.169) $a_h = g_\alpha^{-1} \sum\limits_{\mu=1}^{g_\alpha} z_{G_{\alpha-1}+\mu}$ vir $G_{\alpha-1}+1 \leq h \leq G_\alpha$

$= z_h$.

Verder is

(2.170) $\text{var}(t_i | H_{00}; Z) = n_i (N-n_i) N^{-1} (N-1)^{-1} \sum_h (z_h - \bar{z})^2$

$= n_i (N-n_i)(N-1)^{-1} S_z^2$ waar

(2.171) $S_z^2 = N^{-1} \sum_h (z_h - \bar{z})^2$.

Die toetsingsgrootheid T_k word dus:

(2.172) $T_p = (N-1) N^{-1} S_z^{-2} \sum_i [\sum_j x_{ij} - n_i \bar{z}]^2 / n_i$

$$= \frac{(N-1)}{NS_z^2} \left[\sum_i \frac{t_i^2}{n_i} - N\bar{z}^2 \right] .$$

Gestel $\underline{z}$ is 'n variant uit die populasie waaruit die oorspronklike waarnemings afkomstig is en neem aan dat var($\underline{z}$)>0 (sien voetnoot 10) en $E|\underline{z}|^{2+\delta} < \infty$, $\delta>0$.

Volgens CRAMéR (1946) pp. 346-347 (sien ook STOKER (1955) p. 38) is:

$$(2.173)\quad \lim_{N\to\infty} N^{-1} \sum_h (z_h-\bar{z})^2 = \text{var}(\underline{z}) \quad \text{(in waarskynlikheid)}$$
$$>0 \quad \text{(sodat voorwaarde (2.164)}$$

bevredig word), en

$$(2.174)\quad \lim_{N\to\infty} N^{-1} \sum_h |z_h-\bar{z}|^{2+\delta} \leq E|\underline{z}-\mu|^{2+\delta} \quad \text{(in waarskynlikheid)}$$
$$< \infty \text{ vir } \delta>0 \text{ en } \mu=E(\underline{z}).$$

Soos in stelling 2.3 opmerking 7 volg onder voorwaarde (2.173) dat voorwaarde (2.174) voldoende is vir die geldigheid van (2.165).

Gevolglik besit T_p vir $N\to\infty$ asimptoties 'n χ^2-verdeling met (k-1) g.v.v. mits voorwaarde (2.44) geld.

Opmerking.

Stel alle n_i=b, dan is N=kb en herlei T_p na:

$$(2.175)\quad T_p' = \frac{(kb-1)}{k\,S_z^2} \sum_{i=1}^{k} \left(b^{-1} \sum_{j=1}^{b} x_{ij}-\bar{z}\right)^2.$$

2.7. SLOTOPMERKINGS.

1). In hierdie hoofstuk is alleen onder H_o gewerk. In hoofstuk III word soortgelyke resultate bewys onder 'n algemene hipotese H.* Daardeur word dit moontlik om byvoorbeeld uitdrukkings te vind vir die asimptotiese relatiewe doeltreffendheid van bogenoemde toetse met betrekking tot die F-toets, ens.

2). In hierdie hoofstuk is 'n toetsingsgrootheid gedefinieer wat gebruik kan word in die eenfaktor variansie-analise. Om aan te toon watter noue verband daar tussen hierdie toets T_k en die gewone variansie-analise toets bestaan, gee ons 'n kort uiteensetting van laasgenoemde.

* Onder H word verstaan die hipotese wat H_o en H_a insluit.

Beskou 'n eenfaktor klassifikasie met k rye en n_i waarnemings in die i^{de} ry. Laat $N=\sum_i n_i$, $x_{i.}=n_i^{-1}\sum_j x_{ij}$ en $x_{..}=N^{-1}\sum_{ij}\sum x_{ij}$.

Skematies sien die eenfaktor variansie-analise daar soos volg uit:

TABEL 2.1.

Die eenfaktor Variansie-analise.

Variasiebron	Som van Vierkante	G.v.v.
Tussen rye	$q_1=\sum_i n_i(x_{i.}-x_{..})^2$	$(k-1)$
Binne rye	$q_2=\sum_i\sum_j(x_{ij}-x_{i.})^2$	$(N-k)$
Totaal	$Q=\sum_i\sum_j(x_{ij}-x_{..})^2$	$(N-1)$

As toets vir ry-effekte word gebruik

$F_1=q_1(N-k)/[q_2(k-1)]$ wat 'n F-verdeling met $(k-1)$ en $(N-k)$ g.v.v. besit onder die aanname dat die waarnemings binne alle rye uit normaalpopulasies met gelyke variansies afkomstig is.

Nou is

$$(k-1)F_1 = q_1(N-k)/q_2 = \frac{\sum_i n_i(x_{i.}-x_{..})^2}{\sum_i\sum_j(x_{ij}-x_{i.})^2}\cdot(N-k)$$

$$= \frac{\sum_i n_i^{-1}(\sum_j x_{ij})^2-Nx_{..}^2}{(N-k)^{-1}\sum_{ij}\sum(x_{ij}-x_{i.})^2}\ .$$

Asimptoties, as alle $n_i\to\infty$, besit $(k-1)F_1$ 'n χ^2-verdeling met $(k-1)$ g.v.v. en word presies dieselfde resultate as in die verdelingsvrye geval hierbo gevind waar T_k asimptoties χ^2 verdeel is met $(k-1)$ g.v.v. (vergelyk bv. (2.58) en (2.172)).

3). In hoofstuk V word toetse behandel wat op dieselfde wyse as hierbo ooreenstemming toon met die tweefaktor variansie-analise.

4). Sien hoofstuk V §5.3.1 ten opsigte van 'n verdere moontlike toepassing van die teorie van dié hoofstuk.

HOOFSTUK III.

'n ALGEMENE TOETSINGSGROOTHEID. LIMIETVERDELING ONDER 'n ALGEMENE HIPOTESE H.

3.1. INLEIDING.

Gegee $N = \sum_{i=1}^{k} n_i$ onderling onafhanklike variante $\underline{x}_{ij}$; $j=1,2,\ldots,n_i$; $i=1,2,\ldots,k$ waar $\underline{x}_{ij}$, vir vaste i, afkomstig is uit 'n populasie met kontinue verdelingsfunksie F_i, $i=1,2,\ldots,k$.

Die probleem wat hier beskou word, is om die nulhipotese H_o: $F_1 \equiv F_2 \equiv \ldots \equiv F_k$, naamlik dat alle verdelingsfunksies F_i, $i=1,2,\ldots,k$, identies is, te toets teen 'n alternatiewe hipotese aangedui deur H_a, naamlik dat nie alle verdelingsfunksies identies is nie.

Die N waarnemings $\underline{x}_{ij}$ [1] is spr 0 [2] almal verskillend [3] omdat F_i kontinu veronderstel word.

In hierdie hoofstuk word 'n algemene toets vir bogenoemde probleem behandel en sy asimptotiese eienskappe ondersoek.

3.2. DIE TOETSINGSGROOTHEID S_N.

Laat

(3.1) $\nu_{Ni} = n_i N^{-1}$, dan is $\sum_i \nu_{Ni} = 1$.

Laat verder

(3.2) $\nu_i = \lim_{N \to \infty} \nu_{Ni}$ en neem aan dat

---

1). Tensy anders gespesifiseer, deurloop i,i',i",g en g' in hierdie hoofstuk die waardes 1,2,...,k ; j,j' die waardes 1,2,...,n_i en h die waardes 1,2,...,N.

2). Sien hoofstuk II voetnoot 8.

3). Op die geval van moontlike eventuele gelyke waarnemings word in hierdie hoofstuk nie ingegaan nie.

(3.3) $0<\nu_o\leq\nu_{Ni}\leq 1-\nu_o<1$ vir alle N, vir alle $i=1,2,\ldots,k$ en vir 'n vaste getal $\nu_o\leq\frac{1}{2}$.

Dan geld ook

(3.4) $0<\nu_o\leq\nu_i\leq 1-\nu_o<1$ vir alle i.

Stel

(3.5) $F_{in_i}(x)$ = (aantal $x_{ij}\leq x)/n_i$ vir vaste $i;j=1,2,..,n_i$.

$F_{in_i}(x)$ is dus die empiriese verdelingsfunksie van die i^{de} stel variante $x_{i1},x_{i2},\ldots,x_{in_i}$, d.i. van die i^{de} steekproef.

Definieer

(3.6) $H_N(x) = \sum_i \nu_{Ni}F_{in_i}(x)$, d.i. die gekombineerde empiriese verdelingsfunksie.

Laat

(3.7) $H(x) = \sum_i \nu_{Ni}F_i(x)$, die gekombineerde populasie verdelingsfunksie vir die gegewe steekproefgroottes.

Beskou die grootheid

(3.8) $S_N = \sum_i c_i n_i^{-1} \sum_{j=1}^{n_i} E_{Nij}$ waar die E_{Nij} gegewe getalle is en die c_i willekeurige eindige konstantes, nie almal nul nie.

Neem die volgende vorm van die definisie, naamlik:

(3.9) $S_N = \sum_i c_i \int_{-\infty}^{\infty} J_N[H_N(x)]dF_{in_i}(x)$

waar J_N 'n willekeurige funksie is wat voldoen aan die volgende voorwaardes:

(3.10) $J(H) = \lim_{N\to\infty} J_N(H)$ bestaan vir $0<H<1$,

(3.11) $J(H)$ is nie konstant nie [4],

(3.12) $\int_{I_N} [J_N(H_N)-J(H_N)]dF_{in_i} = o_p(N^{-\frac{1}{2}})$, $i=1,2,\ldots,k$,

waar I_N die interval is waarin $0<H_N(x)<1$ en o_p gedefinieer

---

4). Sien ook voorwaarde (3.28) wat 'n verdere beperking lê, naamlik dat $J'(H)$ nie altyd = 0 kan wees oor die oorvleuelingsgebied van die verdelings nie.

is in §1.2.2,

(3.13) $J_N(1) = o(N^{\frac{1}{2}})$ en

(3.14) $|J^{(m)}(H)| = \left|\frac{d^m J}{dH^m}\right| \leq K[H(1-H)]^{-m-\frac{1}{2}+\delta}$ vir m=0,1,2 en

vir 'n $\delta>0$, K 'n eindige positiewe konstante.

Die twee vorms van S_N is identies as

(3.15) $E_{Nij} = J_N[N^{-1} \sum_{i'} \sum_{j'} \mathcal{L}(x_{ij}-x_{i'j'})]$

waar

$$\mathcal{L}(u) = \begin{cases} 1 & \text{as } u \geq 0 \\ 0 & \text{as } u<0 \end{cases}$$

d.i. as $E_{Nij} = J_N(r_{ij}/N)$

waar r_{ij} die rangnommer van x_{ij} in die gesamentlike rangskikking van die N waardes x_{ij} is.

Laat I die interval wees waarin $0<H<1$.

3.3. LIMIETVERDELING VAN S_N ONDER H (SIEN CHERNOFF EN SAVAGE (1958)).

Laat, vir $g=1,2,\ldots,k$,

(3.16) $B_g(x) = \int_{x_0}^{x} J'[H(x)]dF_g(x)$ en

(3.17) $\tilde{B}_g(\underline{x}_{ij}) = B_g(\underline{x}_{ij}) - EB_g(\underline{x}_i)$

waar $\underline{x}_i$ 'n variant is met verdelingsfunksie F_i en x_o 'n konstante is, byvoorbeeld so dat $H(x_o) = \frac{1}{2}$.

Definieer verder

(3.18) $\xi_{ij} = \nu_{Ni}^{-1} \sum_g (c_i\nu_{Ng} - c_g\nu_{Ni})\tilde{B}_g(x_{ij})$ en

(3.19) $\xi_i = \sum_j \xi_{ij}$.

Lemma 3.1. Die verwagtingswaarde van ξ_{ij} word gegee deur

(3.20) $E(\xi_{ij}|H) = 0$.

Bewys: Omdat $E\tilde{B}_g(x_{ij}) = 0$, is

$$\begin{aligned} E(\xi_{ij}) &= E[\nu_{Ni}^{-1} \sum_g (c_i\nu_{Ng} - c_g\nu_{Ni})\tilde{B}_g(x_{ij})] \\ &= \nu_{Ni}^{-1} \sum_g (c_i\nu_{Ng} - c_g\nu_{Ni})E\tilde{B}_g(x_{ij}) \\ &= 0. \end{aligned}$$

Opmerking.

In hierdie hoofstuk werk ons deurgaans, tensy anders gespesifiseer, onder H. $E(\xi_{ij})$ beteken dus $E(\xi_{ij}|H)$.

Lemma 3.2. $\left[\{H_N(x)-H(x)\}\int_{x_0}^{x} J'[H(x)]dF_i(x)\right]_{-\infty}^{\infty} = 0$ met waarskynlikheid een.

Bewys: $\int_{x_0}^{x} J'[H(x)]dF_i(x) = O\left(\int_{x_0}^{x} J'[H(x)]dH(x)\right)$

$$= O(J[H(x)])$$

$$= O([H(x)\{1-H(x)\}]^{\delta-\frac{1}{2}}) \text{ uit } (3.14).$$

$H_N(x)$ is die empiriese verdelingsfunksie van 'n steekproef uit 'n populasie met verdelingsfunksie $H(x)$. As $z_1<z_2< \ldots <z_N$ die gerangskikte steekproefwaardes is, dan is $H_N(x) = 0$ vir $x<z_1$

en $H_N(x) = 1$ vir $x\geq z_N$.

Indien $H(z_1-0) > 0$, geld

$$[H_N(x)-H(x)]\ O([H(x)\{1-H(x)\}]^{\delta-\frac{1}{2}})$$
$$= O([H(x)]^{\delta+\frac{1}{2}}[1-H(x)]^{\delta-\frac{1}{2}})$$
$$\xrightarrow[-\infty]{x} 0.$$

Indien $H(z_N)<1$, geld

$$[H_N(x)-H(x)]\ O([H(x)\{1-H(x)\}]^{\delta-\frac{1}{2}})$$
$$= O([H(x)]^{\delta-\frac{1}{2}}[1-H(x)]^{\delta+\frac{1}{2}})$$
$$\xrightarrow[+\infty]{x} 0.$$

Aangesien $H(x)$ kontinu is, geld

$H(z_1-0) > 0$ en $H(z_N) < 1$ met waarskynlikheid een vir alle eindige N.

Dus: $\left[\{H_N(x)-H(x)\}\int_{x_0}^{x} J'[H(x)]dF_i(x)\right]_{-\infty}^{\infty} = 0$ met waarskynlikheid een.

Lemma 3.3. Die variansie van $N^{-\frac{1}{2}} \sum\limits_i \xi_i$ word gegee deur:

(3.21) $\operatorname{var}(N^{-\frac{1}{2}}\sum\limits_i \xi_i) = \sum\limits_i \nu_{Ni}^{-1} \sum\limits_{g}\sum\limits_{g'} (c_i\nu_{Ng} - c_g\nu_{Ni})(c_i\nu_{Ng'} - c_{g'}\nu_{Ni})\sigma_{B_{gi}B_{g'i}}$

waar

(3.22) $\sigma_{B_{gi}B_{g'i}} = \int\int_{-\infty<x<y<\infty} F_i(x)[1-F_i(y)]J'[H(x)]J'[H(y)]\cdot$
$\cdot[dF_g(x)dF_{g'}(y) + dF_{g'}(x)dF_g(y)].$

Bewys: Omdat

(3.20) $E(\xi_{ij}) = 0$, is

(3.23) $\operatorname{var}(\xi_{ij}) = E(\xi_{ij}^2)$

$= E[\nu_{Ni}^{-1}\sum\limits_g (c_i\nu_{Ng} - c_g\nu_{Ni})\tilde{B}_g(x_{ij})\nu_{Ni}^{-1}\sum\limits_{g'}(c_i\nu_{Ng'} - c_{g'}\nu_{Ni})\tilde{B}_{g'}(x_{ij})]$

$= \nu_{Ni}^{-2}\sum\limits_g \sum\limits_{g'}(c_i\nu_{Ng} - c_g\nu_{Ni})(c_i\nu_{Ng'} - c_{g'}\nu_{Ni})E\tilde{B}_g(x_{ij})\tilde{B}_{g'}(x_{ij}).$

Verder is

(3.24) $\operatorname{var}(N^{-\frac{1}{2}}\xi_i) = N^{-1}\operatorname{var}(\xi_i)$

$= N^{-1}\operatorname{var}(\sum\limits_j \xi_{ij})$

$= N^{-1}\sum\limits_j \operatorname{var}(\xi_{ij})$ omdat ξ_{ij} en $\xi_{ij'}$ vir $j' \neq j$ onderling onafhanklik is (want x_{ij} en $x_{ij'}$ is onafhanklik)

$= \frac{1}{N}\sum\limits_j \nu_{Ni}^{-2}\sum\limits_g \sum\limits_{g'}(c_i\nu_{Ng} - c_g\nu_{Ni})(c_i\nu_{Ng'} - c_{g'}\nu_{Ni})E\tilde{B}_g(x_{ij})\tilde{B}_{g'}(x_{ij})$

$= \nu_{Ni}^{-1}\sum\limits_g \sum\limits_{g'}(c_i\nu_{Ng} - c_g\nu_{Ni})(c_i\nu_{Ng'} - c_{g'}\nu_{Ni})E\tilde{B}_g(x_{ij})\tilde{B}_{g'}(x_{ij}).$

Die variante $N^{-\frac{1}{2}}\xi_i$ en $N^{-\frac{1}{2}}\xi_{i'}$ is onderling onafhanklik indien $i' \neq i$. Gevolglik is

(3.25) $\operatorname{var}(\sum\limits_i N^{-\frac{1}{2}}\xi_i) = \sum\limits_i \operatorname{var}(N^{-\frac{1}{2}}\xi_i)$

$= \sum\limits_i \nu_{Ni}^{-1}\sum\limits_g \sum\limits_{g'}(c_i\nu_{Ng} - c_g\nu_{Ni})(c_i\nu_{Ng'} - c_{g'}\nu_{Ni})E\tilde{B}_g(x_{ij})\tilde{B}_{g'}(x_{ij}).$

Nou moet $E\tilde{B}_g(x_{ij})\tilde{B}_{g'}(x_{ij})$ nog bereken word.

(3.26) $B_g(x_{ij}) - EB_g(x_i) = \int_{-\infty}^{\infty} B_g(x)dF_{i1}(x) - \int_{-\infty}^{\infty} B_g(x)dF_i(x)$

vir $j = 1, 2, \ldots, n_i$

$= \int_{-\infty}^{\infty} B_g(x)d[F_{i1}(x) - F_i(x)]$

$= \int_{-\infty}^{\infty}\left\{\int_{x_0}^{x} J'[H(x)]dF_g(x)\right\}d[F_{i1}(x) - F_i(x)]$

$$= [\{\int_{x_0}^{x} J'[H(x)]dF_g(x)\}\{F_{il}(x) - F_i(x)\}]_{-\infty}^{\infty} -$$

$$- \int_{-\infty}^{\infty} [F_{il}(x) - F_i(x)]J'[H(x)]dF_g(x)$$ deur parsiële integrasie. M.b.v. lemma 3.2 [5)] volg verder

$$= - \int_{-\infty}^{\infty} [F_{il}(x)-F_i(x)]J'[H(x)]dF_g(x)$$ met waarskynlikheid één, sodat

$$(3.27)\quad E\tilde{B}_g(x_{ij})\tilde{B}_{g'}(x_{ij}) = E\Big\{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[F_{il}(x)-F_i(x)][F_{il}(y)-F_i(y)]\cdot$$
$$\cdot J'[H(x)]J'[H(y)]dF_g(x)dF_{g'}(y)\Big\}$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} E[F_{il}(x)-F_i(x)][F_{il}(y)-F_i(y)]J'[H(x)]\cdot$$
$$\cdot J'[H(y)]dF_g(x)dF_{g'}(y)$$
$$= \iint_{-\infty<x<y<\infty} F_i(x)[1-F_i(y)]J'[H(x)]J'[H(y)]dF_g(x)dF_{g'}(y) +$$
$$+ \iint_{-\infty<y<x<\infty} F_i(y)[1-F_i(x)]J'[H(x)]J'[H(y)]dF_g(x)dF_{g'}(y)$$
$$= \sigma_{B_{gi}B_{g'i}}$$ uit (3.22), want:

Neem byvoorbeeld $x<y$:

As $\underline{x}_i \leq x$, is $F_{il}(x) = 1$ en $F_{il}(y) = 1$.

As $x<\underline{x}_i\leq y$, is $F_{il}(x) = 0$ en $F_{il}(y) = 1$.

As $\underline{x}_i>y$, is $F_{il}(x) = 0$ en $F_{il}(y) = 0$.

$$E[F_{il}(x)-F_i(x)][F_{il}(y)-F_i(y)]$$
$$= [1-F_i(x)][1-F_i(y)]P[\underline{x}_i\leq x] - F_i(x)[1-F_i(y)]P[x<\underline{x}_i\leq y] +$$
$$+ F_i(x)F_i(y)P[\underline{x}_i>y]$$
$$= [1-F_i(x)][1-F_i(y)]F_i(x) - F_i(x)[1-F_i(y)][F_i(y)-F_i(x)] +$$
$$+ F_i(x)F_i(y)[1-F_i(y)]$$
$$= [1-F_i(y)]F_i(x).$$

'n Soortgelyke resultaat volg vir $y<x$.

Hiermee is die lemma bewys.

Opmerking.

Met behulp van (3.27) herlei (3.23) tot:

$$(3.23a)\quad var(\xi_{ij}) = \nu_{Ni}^{-2}\sum_g\sum_{g'}(c_i\nu_{Ng}-c_g\nu_{Ni})(c_i\nu_{Ng'}-c_{g'}\nu_{Ni})\sigma_{B_{gi}B_{g'i}}\ .$$

---

5). Sien vergelykings (3.3), (3.50) en (3.51).

Gevolglik besit t_i en t_i'/n_i asimptoties dieselfde verdeling onder H. Alle resultate ten opsigte van t_i geld dus ook asimptoties vir t_i'/n_i en ons kry uit §3.4 opmerking 3 en (3.59), (3.60), (3.62) en (3.63):

$$E(t_i'/n_i|H) = \int_{-\infty}^{\infty} J[H(x)]dF_i(x) + o(N^{-\frac{1}{2}}).$$

Opmerkings.

1). Onder H_o: $F_1 \equiv F_2 \equiv \ldots \equiv F_k \equiv F$ geld

$$(3.95)\quad E(t_i'/n_i|H_o) = \int_{-\infty}^{\infty} J[F(x)]dF(x) + o(N^{-\frac{1}{2}}).$$

2). Vergelyk

$$(2.42)\quad \underset{h}{\text{maks}}(\Psi_{Nh}-\overline{\Psi}_N)^2 = o(N), \text{ d.w.s. } \underset{h}{\text{maks}}|\Psi_{Nh}| = o(N^{\frac{1}{2}}),$$

wat soortgelyk is aan (3.92).

3). Onder voorwaarde (3.93) sal, vir 0<H<1,

$$(3.96)\quad \lim_{N\to\infty} \Psi_N[NH/(N+1)] = \lim_{N\to\infty} J_N(H)$$
$$= J(H) \text{ uit (3.10).}$$

4). Voorwaarde (3.93) lê die verband tussen die toetsingsgroothede van hoofstukke II en III. Ons kon, as alternatiewe prosedure, ook geneem het

$$J_N[H_N(x)] = \Psi_N[NH_N(x)/(N+1)].$$

3.6.2. Toetse vir Verskuiwing.

3.6.2.1. 'n Algemene uitdrukking vir die asimptotiese relatiewe doeltreffendheid van een toets met betrekking tot 'n tweede toets.

Beskou die alternatiewe hipotese

$$(3.97)\quad H_a: F_i(x) = F(x+d_iN^{-\frac{1}{2}}) \text{ vir } i=1,2,\ldots,k$$

waar die d_i 's willekeurige eindige reële getalle is.

Neem aan dat $f(x) = F'(x)$ oral bestaan, kontinu en gelykmatig begrens is.

STELLING 3.7.

Indien

(3.92) $\Psi_N[N/(N+1)] = o(N^{\frac{1}{2}})$,

(3.93) $\int_{I_N} |J_N[H_N(x)]-\Psi_N[NH_N(x)/(N+1)]|dF_{in_i}(x) = o_p(N^{-\frac{1}{2}})$

vir $i=1,2,\ldots,k$ en

(3.98) $|f(x+\delta_i N^{-\frac{1}{2}})J'[\sum_{i'}\nu_{Ni'}F(x+\delta_{i'} N^{-\frac{1}{2}})]| < M(x)$ vir N voldoende groot en $\int_{-\infty}^{\infty} M(x)dF(x) < \infty$, waar die δ_i 's willekeurige eindige reële getalle is, dan geld

(3.99) $\lim_{N\to\infty} N^{\frac{1}{2}}[E(t_i'/n_i|H_a)-E(t_i'/n_i|H_o)] =$

$$= (d.-d_i) \int_{-\infty}^{\infty} f(x)J'[F(x)]dF(x) \quad \text{waar}$$

(3.100) $d. = \sum_{i'}\nu_{i'} d_{i'}$.

Bewys: Onder H_a is

(3.101) $F_{i'}(x) = F(x+d_{i'} N^{-\frac{1}{2}})$

$$= F[x+d_i N^{-\frac{1}{2}}+(d_{i'}-d_i)N^{-\frac{1}{2}}]$$

$$= F_i[x+(d_{i'}-d_i)N^{-\frac{1}{2}}] \text{ vir alle } i'$$

$$= F_i(x+\theta_{i'} N^{-\frac{1}{2}}) \quad \text{waar}$$

(3.102) $\theta_{i'} = d_{i'}-d_i$, $i' =1,2,\ldots,k$, sodat

(3.103) $H(x) = \sum_{i'}\nu_{Ni'} F_i(x+\theta_{i'} N^{-\frac{1}{2}})$ uit (3.7).

Uit die eerste middelwaardestelling in k veranderlikes geld vir die funksie $G(\theta_1,\theta_2,\ldots,\theta_k)$:

(3.104) $G(\theta_1,\theta_2,\ldots,\theta_k) = G(0,0,\ldots,0)+\sum_{i'}\theta_{i'}\left.\frac{\partial G(\theta_1,\theta_2,\ldots,\theta_k)}{\partial\theta_{i'}}\right|_{\vec{\theta}=\hat{\vec{\theta}}}$

waar $\vec{\theta} = (\theta_1,\theta_2,\ldots,\theta_k)$ en $|\hat{\theta}_{i'}|<|\theta_{i'}|$ vir alle i' (sien STEWART, C. A. (1940) „Advanced Calculus").

Stel nou

(3.105) $G(\theta_1,\theta_2,\ldots,\theta_k) = N^{\frac{1}{2}} \int_{-\infty}^{\infty} J[H(x)]dF_i(x)$

$$= N^{\frac{1}{2}} \int_{-\infty}^{\infty} J[\sum_{i'}\nu_{Ni'} F_i(x+\theta_{i'} N^{-\frac{1}{2}})]dF_i(x) \text{ uit (3.103)},$$

dan is:

(3.106) $G(0,0,..,0) = N^{\frac{1}{2}} \int_{-\infty}^{\infty} J[F(x)]dF(x)$

en met behulp van (3.104) volg dan

(3.107) $G(\theta_1,\theta_2,..,\theta_k) - G(0,0,..,0) =$

$$= \sum_{i'} \nu_{Ni'}\, \theta_{i'} \int_{-\infty}^{\infty} f_i(x+\hat{\theta}_{i'} N^{-\frac{1}{2}})J'[\sum_{i''}\nu_{Ni''} F_i(x+\hat{\theta}_{i''} N^{-\frac{1}{2}})]dF_i(x)$$

sodat m.b.v. (3.94), (3.105) en (3.106) volg

$$\lim_{N\to\infty} N^{\frac{1}{2}}[E(t_i'/n_i|H_a)-E(t_i'/n_i|H_o)] = \sum_{i'}\nu_{i'}\,\theta_{i'} \int_{-\infty}^{\infty} f(x)J'[F(x)]dF(x)$$

$$= (d_{\cdot}-d_i) \int_{-\infty}^{\infty} f(x)J'[F(x)]dF(x).$$

Die differensiasie m.b.t. die $\theta_{i'}$ onder die integraalteken en die neem van die limiet vir $N\to\infty$ onder die integraalteken, is toelaatbaar onder voorwaarde (3.98) (sien CRAMéR (1946) pp. 66-67, GIBSON (1954) p. 441 stelling III) aangesien $J'[F(x)]$ en $f(x)$ kontinu veronderstel word.

Hiermee is die stelling bewys.

Opmerkings.

1). Voorwaarde (3.98) kan in die volgende vorm omskrywe word, naamlik:

(3.108) $|f(x+cN^{-\frac{1}{2}})J'[\lambda_o F(x+cN^{-\frac{1}{2}})+(1-\lambda_o)F(x+c''N^{-\frac{1}{2}})]| < M(x)$

vir N voldoende groot met $\int_{-\infty}^{\infty} M(x)dF(x) < \infty$ waar $\lambda_o>0$ en die c 's eindige reële getalle is.

2). 'n Voldoende voorwaarde vir die geldigheid van (3.98), is

(3.109) $|f(x+cN^{-\frac{1}{2}})J'[F(x+c' N^{-\frac{1}{2}})]| \leq M < \infty$ gelykmatig in x vir N voldoende groot waar c en c' willekeurige eindige reële getalle is.

STELLING 3.8.

Onder die voorwaardes (3.10)-(3.14), (3.92), (3.93) en (3.98) besit

(3.110) $T_k' = \sum_i (N-n_i)[t_i'-E(t_i'|H_o)]^2[N\ \mathrm{var}(t_i'|H_o)]^{-1}$

onder H_a asimptoties, vir $N\to\infty$, 'n nie-sentrale χ^2-verdeling met (k-1) grade van vryheid en nie-sentraliteitsparameter

$$(3.111)\quad \lambda_k^2 = [\int_{-\infty}^{\infty} f(x)J'[F(x)]dF(x)]^2 \sum_i \nu_i (d.-d_i)^2 \cdot$$

$$\cdot [\lim_{N\to\infty} N^{-1} \sum_h (\Psi_{Nh} - \bar{\Psi}_N)^2]^{-1}.$$

Bewys: Definieer, soos in vergelyking (2.4),

$$(3.112)\quad \hat{t}_i' = (N-n_i)^{\frac{1}{2}}[t_i' - E(t_i'|H_o)][N\ \mathrm{var}(t_i'|H_o)]^{-\frac{1}{2}}.$$

Beskou ook

$$(3.56)\quad \hat{t}_i = (N-n_i)^{\frac{1}{2}}[t_i - E(t_i|H_a)][N\ \mathrm{var}(t_i|H_a)]^{-\frac{1}{2}}.$$

Uit stelling 3.6 volg dat die momentematriks van die $\hat{t}_i$'s onder H_a asimptoties gegee word deur $\Lambda = I - \vec{\nu}\vec{\nu}'$ met karakteristieke wortels 1,1,...,1 en 0 as $N\to\infty$ [9].

Nou is:

$$(3.113)\quad \hat{t}_i' = (N-n_i)^{\frac{1}{2}}[t_i'/n_i - E(t_i'/n_i|H_o)][N\mathrm{var}(t_i'/n_i|H_o)]^{-\frac{1}{2}}$$

uit (3.112)

$$= (N-n_i)^{\frac{1}{2}}[t_i'/n_i - E(t_i'/n_i|H_a)][N\mathrm{var}(t_i'/n_i|H_o)]^{-\frac{1}{2}} +$$

$$+ (N-n_i)^{\frac{1}{2}}[E(t_i'/n_i|H_a) - E(t_i'/n_i|H_o)][N\mathrm{var}(t_i'/n_i|H_o)]^{-\frac{1}{2}}$$

$$= \hat{t}_i'' V_{Ni} + E_{Ni} \quad \text{waar}$$

$$(3.114)\quad E_{Ni} = (N-n_i)^{\frac{1}{2}}[E(t_i'/n_i|H_a) - E(t_i'/n_i|H_o)]\cdot$$

$$\cdot [N\mathrm{var}(t_i'/n_i|H_o)]^{-\frac{1}{2}},$$

$$(3.115)\quad \hat{t}_i'' = (N-n_i)^{\frac{1}{2}}[t_i'/n_i - E(t_i'/n_i|H_a)][N\mathrm{var}(t_i'/n_i|H_a)]^{-\frac{1}{2}} \quad \text{en}$$

$$(3.116)\quad V_{Ni} = [\mathrm{var}(t_i'/n_i|H_a)]^{\frac{1}{2}}[\mathrm{var}(t_i'/n_i|H_o)]^{-\frac{1}{2}}.$$

Hier is

$$(3.117)\quad \lim_{N\to\infty} E_{Ni} = \lim_{N\to\infty} (N-n_i)^{\frac{1}{2}} N^{\frac{1}{2}} [E(t_i'/n_i|H_a) - E(t_i'/n_i|H_o)]\cdot$$

$$\cdot [N^2 \mathrm{var}(t_i'/n_i|H_o)]^{-\frac{1}{2}}$$

$$= \nu_i^{\frac{1}{2}}(d.-d_i) \int_{-\infty}^{\infty} f(x)J'[F(x)]dF(x) \cdot \lim_{N\to\infty} [N^{-1}\sum_h (\Psi_{Nh} - \bar{\Psi}_N)^2]^{-\frac{1}{2}}$$

uit (2.20), (2.21) en (3.99)

$= \text{konstant} < \infty$ m.b.v. (3.98) en (2.41) omdat F(x) kontinu is.

---

9). Onder voorwaarde (3.97) word daar voldoen aan voorwaarde (3.72) vir N voldoende groot.

Nou volg dat (vergelyk onder andere lemma 3.23)

$$\lim_{N\to\infty} P\left[\left\{\frac{N-n_i}{N}\right\}^{\frac{1}{2}} \left|\frac{t_i'/n_i-E(t_i'/n_i\,|H_a)}{[\mathrm{var}(t_i'/n_i\,|H_a)]^{\frac{1}{2}}} - \frac{t_i-E(t_i\,|H_a)}{[\mathrm{var}(t_i\,|H_a)]^{\frac{1}{2}}}\right| > \varepsilon\,|H_a\right] = 0$$

waar $\varepsilon > 0$.

Gevolglik besit $\hat{t}_i$ en $\hat{t}_i''$ asimptoties dieselfde verdeling onder H_a en volg met behulp van (3.113) dat

$$(3.118)\quad \lim_{N\to\infty} \mathrm{var}(\hat{t}_i'\,|H_a) = \lim_{N\to\infty} V_{Ni}^2\ \mathrm{var}(\hat{t}_i\,|H_a).$$

Maar

$$(3.119)\quad \lim_{N\to\infty} V_{Ni}^2 = \lim_{N\to\infty} [\mathrm{var}(t_i'/n_i\,|H_a)][\mathrm{var}(t_i'/n_i\,|H_o)]^{-1}$$

$$= \lim_{N\to\infty} [\mathrm{var}(t_i\,|H_a)][\mathrm{var}(t_i\,|H_o)]^{-1} \text{ met behulp van lemma 3.23}$$

$$= 1 \text{ uit lemma 3.22 .}$$

Dus

$$(3.120)\quad \lim_{N\to\infty} \mathrm{var}(\hat{t}_i'\,|H_a) = \lim_{N\to\infty} \mathrm{var}(\hat{t}_i\,|H_a)$$

waarby $0 < \lim_{N\to\infty} \mathrm{var}(\hat{t}_i\,|H_a) < \infty$ uit (3.77).

Soortgelyk is

$$(3.121)\quad \lim_{N\to\infty} \mathrm{kov}(\hat{t}_i',\hat{t}_{i'}'\,|H_a) = \lim_{N\to\infty} \mathrm{kov}(\hat{t}_i,\hat{t}_{i'}\,|H_a).$$

Die momentematriks van die $\hat{t}_i'$'s onder H_a is dus ook asimptoties gelyk aan $\Lambda = I-\vec{\nu}\vec{\nu}'$ met karakteristieke wortels 1,1,...,1 en 0 (lemma 1.2).

Uit (3.113) volg:

$$\sum_i \nu_i^{\frac{1}{2}} E(\hat{t}_i'\,|H_a) = \sum_i \nu_i^{\frac{1}{2}}(N-n_i)^{\frac{1}{2}}[E(t_i'/n_i\,|H_a)-E(t_i'/n_i\,|H_o)]\cdot$$
$$\cdot[N\ \mathrm{var}(t_i'/n_i\,|H_o)]^{-\frac{1}{2}}$$
$$= \sum_i \nu_i^{\frac{1}{2}} E_{Ni} \quad \text{uit (3.114), sodat}$$

$$(3.122)\quad \lim_{N\to\infty} \sum_i \nu_i^{\frac{1}{2}} E(\hat{t}_i'\,|H_a) = \sum_i \nu_i^{\frac{1}{2}} \lim_{N\to\infty} E_{Ni}$$

$$= \sum_i \nu_i(d_{.}-d_i) \int_{-\infty}^{\infty} f(x)J'[F(x)]dF(x) \lim_{N\to\infty} [N^{-1}\sum_h(\Psi_{Nh}-\bar{\Psi}_N)^2]^{-\frac{1}{2}}$$

$= 0$ met behulp van (3.100), sodat aan (1.27) voldoen word.

Op analoë wyse as in lemma 3.21 volg dat die $\hat{t}_i'$'s se gesamentlike verdeling onder H_a asimptoties normaal is (vergelyk ook lemma 3.23 met sy bewys).

Uit stelling 1.1 opmerking 1 volg dat $T_k' = \sum_i \hat{t}_i'^2$ asimptoties, vir $N \to \infty$, 'n nie-sentrale χ^2-verdeling besit met (k-1) g.v.v. en nie-sentraliteitsparameter

$$(3.123)\quad \lambda_k^2 = \lim_{N\to\infty} \sum_{i=1}^{k} [E(\hat{t}_i' | H_a)]^2$$

$$= \lim_{N\to\infty} \sum_i \left[\frac{(N-n_i)^{\frac{1}{2}}[E(t_i'|H_a)-E(t_i'|H_o)]}{[N\mathrm{var}(t_i'|H_o)]^{\frac{1}{2}}}\right]^2$$

$$= \lim_{N\to\infty} \sum_i \left[\frac{(N-n_i)[E(t_i'/n_i|H_a)-E(t_i'/n_i|H_o)]^2}{N\mathrm{var}(t_i'/n_i|H_o)}\right]$$

$$= \frac{[\int_{-\infty}^{\infty} f(x)J'[F(x)]dF(x)]^2 \sum_i \nu_i(d.-d_i)^2}{\lim_{N\to\infty} N^{-1} \sum_h (\Psi_{Nh}-\bar{\Psi}_N)^2} \text{ uit } (3.117).$$

Opmerkings.

1). Voorwaarde (3.93) lê die verband tussen J_N en Ψ_N.

2). λ_k^2 kan ook in 'n ander vorm geskryf word.

Onder voorwaarde (3.93) kan, soos in lemma 3.23, aangetoon word dat

$$(3.124)\quad \lim_{N\to\infty} \int_{-\infty}^{\infty} J_N[H_N(x)]dH_N(x) = \lim_{N\to\infty} N^{-1} \sum_h \Psi_N(\delta_h).$$

Indien verder ook geld

$$(3.125)\quad \int_{-\infty}^{\infty} |J_N^2[H_N(x)]-\Psi_N^2[NH_N(x)/(N+1)]|dH_N(x) = o_p(1) ,$$

dan volg op 'n analoë wyse as in die vorige geval dat

$$(3.126)\quad \lim_{N\to\infty} \int_{-\infty}^{\infty}[J_N[H_N(x)] - \int_{-\infty}^{\infty}J_N[H_N(x)]dH_N(x)]^2 dH_N(x)$$

$$= \lim_{N\to\infty} N^{-1} \sum_h (\Psi_{Nh}-\bar{\Psi}_N)^2$$

mits die uitdrukkings almal bestaan en eindig is.

Die nie-sentraliteitsparameter word dan

$$(3.127)\quad \lambda_k^2 = \frac{[\int_{-\infty}^{\infty} f(x)J'[F(x)]dF(x)]^2 \sum_i \nu_i(d.-d_i)^2}{\lim_{N\to\infty} \int_{-\infty}^{\infty}\{J_N[H_N(x)]-\int_{-\infty}^{\infty}J_N[H_N(x)]dH_N(x)\}^2 dH_N(x)}$$

STELLING 3.9.

Onder voorwaardes (3.10) - (3.14), (3.92), (3.93) en (3.98) is die asimptotiese relatiewe doeltreffendheid van T_k' met betrekking tot die variansieverhoudingstoets F:

$$(3.128)\quad E_{T_k',F} = [\int_{-\infty}^{\infty} f(x)J'[F(x)]dF(x)]^2\sigma_F^2 \lim_{N\to\infty} [\frac{1}{N}\sum_h(\Psi_{Nh}-\bar{\Psi}_N)^2]^{-1}$$

waar

$$(3.129)\quad \sigma_F^2 = \int_{-\infty}^{\infty} x^2dF(x) - [\int_{-\infty}^{\infty} x\, dF(x)]^2.$$

Bewys: Indien σ_F^2 bestaan en eindig is, besit die F-toets onder H_a ook asimptoties, vir $N\to\infty$, 'n nie-sentrale χ^2-verdeling met (k-1) g.v.v. en nie-sentraliteitsparameter

$$(3.130)\quad \lambda_F^2 = \sigma_F^{-2} \sum_i \nu_i(d.-d_i)^2$$ - sien byvoorbeeld TANG(1938) pp. 136-139.

Die a.r.d. van T_k' m.b.t. F is dan (sien §1.4):

$$(3.131)\quad E_{T_k',F} = \lambda_k^2/\lambda_F^2$$

$$= [\int_{-\infty}^{\infty} f(x)J'[F(x)]dF(x)]^2\sigma_F^2 \lim_{N\to\infty} [\frac{1}{N}\sum_h(\Psi_{Nh}-\bar{\Psi}_N)^2]^{-1}$$

$$= \frac{[\int_{-\infty}^{\infty} f(x)J'[F(x)]dF(x)]^2\sigma_F^2}{\lim_{N\to\infty}\int_{-\infty}^{\infty}\{J_N[H_N(x)] - \int_{-\infty}^{\infty} J_N[H_N(x)]dH_N(x)\}^2 dH_N(x)} \;*$$

indien aan voorwaarde (3.125) ook nog voldoen word.

Opmerkings.

1). As aan (3.98) voldoen word behalwe in 'n versameling punte van maat nul m.b.t. F, geld alle voorgaande resultate ook. Dit impliseer dat f(x) nie gelykmatig begrens hoef te wees en J'[F(x)] nie aan voorwaarde (3.14) hoef te voldoen in 'n versameling punte x van maat nul nie. (Vergelyk byvoorbeeld die opmerking hieroor in ANSARI en BRADLEY (1960) p. 1183).

2). Indien J'[F(x)] begrens is vir alle x, word voorwaarde (3.98) onmiddellik bevredig.

* Hierdie uitdrukking is onafhanklik van k sodat dieselfde resultaat vir die a.r.d. verkry word as in die geval van twee steekproewe. Sien ook die voetnoot op bladsy 92.

Vervolgens bespreek ons enkele spesiale gevalle van T_k'. In ons verwysings na toetsingsgroothede van hoofstuk II, hou ons in gedagte dat alle $g_\alpha = 1$ met waarskynlikheid een omdat die verdelingsfunksies F_i hier kontinu veronderstel word.

3.6.2.2. $\Psi_N(\delta_{ij}) \equiv \delta_{ij}$ (sien §2.5.2.1).

Hiervoor is

(2.59) $t_i' = \sum_j r_{ij}/(N+1)$ met toetsingsgrootheid

(2.65) $T_W = 12(N+1)N^{-1} \sum_i t_i^2/n_i - 3(N+1)$ indien alle g_α=1.

Stel

(3.132) $$J_N(F) = \Psi_N[NF/(N+1)] \text{ vir } 0<F<1$$
$$= NF/(N+1),$$

dan is $J(F) = F$ vir $0<F<1$.

Aan voorwaardes (3.10) - (3.14), (3.92) en (3.93) word voldoen (sien ook CAPON(1961) §5 (F)).

Nou is
$$f(x)J'[F(x)] = f(x)\,\frac{dJ[F(x)]}{dF(x)}$$
$$= f(x)\,\frac{dF(x)}{dF(x)}$$
$$= f(x).$$

$f(x)$ is 'n funksie waarvoor ons aangeneem het $f(x)\leq K<\infty$ (sien §3.6.2.1). Daar word dus voldoen aan voorwaarde (3.98) - sien stelling 3.7 opmerking 2.

Verder is
$$\int_{-\infty}^{\infty} f(x)J'[F(x)]dF(x) = \int_{-\infty}^{\infty} f(x)dF(x)$$
$$= \int_{-\infty}^{\infty} f^2(x)\,dx \qquad \text{en}$$
$$\lim_{N\to\infty} N^{-1} \sum_h (\Psi_{Nh}-\bar{\Psi}_N)^2 = \lim_{N\to\infty} \frac{N-1}{12(N+1)} \quad \text{m.b.v. (2.61)}$$
$$= \frac{1}{12}\ .$$

Die nie-sentraliteitsparameter van (3.123) herlei na:

(3.133) $\lambda_W^2 = 12[\int_{-\infty}^{\infty} f^2(x)\,dx]^2 \sum_i \nu_i(d.-d_i)^2.$

Die a.r.d. van T_W m.b.t. F is dus

(3.134) $E_{T_W,F} = 12\sigma_F^2[\int_{-\infty}^{\infty} f^2(x)\,dx]^2$ m.b.v. (3.130) en (3.131).

Opmerkings.

1). $E_{T_W,F}$ is dieselfde as die a.r.d. van WILCOXON(1945) se toetsingsgrootheid m.b.t. die t-toets in die twee-steekproef geval - sien HODGES en LEHMANN(1956).

2). Vir die normaalverdeling is $E_{T_W,F} = 3/\pi$.

Vir die homogene verdeling is $E_{T_W,F} = 1$.

HODGES en LEHMANN(1956) het bewys dat $E_{T_W,F}$ sy minimum-waarde van 0.864 bereik vir die verdeling met

$$f(x) = \begin{cases} 3\sqrt{5}(5-x^2)/20 & \text{vir } x^2 \le 5 \\ 0 & \text{vir } x^2 > 5 \end{cases}.$$

3). CHACKO(1963) het vir die k-steekproef geval die uitdrukking in (3.134) gevind vir die meer beperkende geval dat alle $n_i = n$.

3.6.2.3. $\Psi_N(\delta_{ij}) \equiv \Xi_{r_{ij}} \equiv E(\xi_{r_{ij}})$ - sien §2.5.2.2.

In hierdie geval is

(2.68) $t_i' = \sum_j \Xi_{r_{ij}}$ met toetsingsgrootheid T_T gegee in (2.71) waarby alle $g_\alpha=1$ (sien die slot van §3.6.2.1 m.b.t. die kontinuïteit van die F_i en die feit dat alle $g_\alpha=1$).

Volgens CHERNOFF en SAVAGE(1958) §5 is J gelyk aan J_o, die inverse funksie van die normaal (0,1) verdelingsfunksie Φ. Vergelyk ook CAPON(1961) §5 (B). Gevolglik geld: $u = \Phi[J(u)]$ sodat

$1 = \emptyset[J(u)]J'(u)$ deur differensiasie na u

waarby $\emptyset(u)$ die frekwensiefunksie van die normaal (0,1) - verdeling is.

Dus

(3.135) $J'(u) = 1/\emptyset[J(u)]$.

Uit lemma 3.12 volg dat voorwaardes (3.10) - (3.13) bevredig word indien (3.14) geld. Volgens CAPON(1961) §5 word ook aan (3.14) voldoen. Ook voorwaardes (3.92) en (3.93) word bevredig (m.b.v. C+S(1958) §4 en §6 en vergelyking (3.12)).

Omdat daar met waarskynlikheid een geen knope voorkom nie, volg m.b.v. STOKER(1955) p. 55 dat

$$\begin{aligned}\lim_{N\to\infty} N^{-1} \sum_h (\Psi_{Nh}-\bar{\Psi}_N)^2 &= \lim_{N\to\infty} N^{-1} \sum_h \Xi_h^2 \\ &= E(\xi^2) \\ &= \text{var}(\xi) \\ &= 1.\end{aligned}$$

Vir enige willekeurige verdelingsfunksie F (waarvoor aan (3.98) voldoen word) geld nou uit (3.123) dat

$$(3.136)\quad \lambda_T^2 = [\int_{-\infty}^{\infty} f(x)J'[F(x)]dF(x)]^2 \sum_i \nu_i(d.-d_i)^2$$

$$= \left[\int_{-\infty}^{\infty} \frac{f(x)}{\emptyset[J\{F(x)\}]} dF(x)\right]^2 \sum_i \nu_i(d.-d_i)^2.$$

Die a.r.d. van TERRY(1952) se toets m.b.t. die F-toets word gegee deur

$$(3.137)\quad E_{T_T,F} = \sigma_F^2[\int_{-\infty}^{\infty} f^2(x)J'[F(x)]dx]^2$$

$$= \sigma_F^2\left[\int_{-\infty}^{\infty} \frac{f^2(x)}{\emptyset[J\{F(x)\}]} dx\right]^2.$$

Hierdie resultaat is presies dieselfde as dié verkry deur C+S(1958) vir die a.r.d. van die Terrytoets m.b.t. die t-toets in die twee-steekproef geval [10]. *

CHERNOFF en SAVAGE(1958) het bewys dat genoemde uitdrukking groter of gelyk aan een is met gelykheid in

10). Let op dat daar 'n drukfout in hulle uitdrukking (5.5) is. Die σ^2 moet in die teller staan en nie in die noemer nie.

geval van die normaalverdeling, sodat ons ook direk kan beweer dat $E_{T_T,F} \geq 1$.

Opmerkings.

1). Die gelykheid $E_{T_T,F} = 1$ in die geval van die normaalverdeling kan maklik aangetoon word, want indien $F(x)$ die normaal (μ,σ^2) - verdelingsfunksie is, volg:

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} \text{eksp}[-\tfrac{1}{2}(y-\mu)^2/\sigma^2]\, dy$$

$$= \Phi(\tfrac{x-\mu}{\sigma}) \text{ en frekwensiefunksie}$$

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \text{eksp}[-\tfrac{1}{2}(x-\mu)^2/\sigma^2]$$

$$= \tfrac{1}{\sigma}\phi(\tfrac{x-\mu}{\sigma}) \quad \text{sodat}$$

$$J[F(x)] = \Phi^{-1}[\Phi(\tfrac{x-\mu}{\sigma})]$$

$$= \frac{x-\mu}{\sigma} \text{ en}$$

$$\phi[J\{F(x)\}] = \phi(\tfrac{x-\mu}{\sigma})$$

$$= \sigma\, f(x) \quad \text{met } \sigma = \sigma_F .$$

Ten slotte is:

$$(3.138)\quad \lambda_T^2 = \left[\int_{-\infty}^{\infty} \frac{f(x)}{\sigma_F f(x)}\, dF(x)\right]^2 \sum_i \nu_i(d.-d_i)^2 \text{ uit } (3.136)$$

$$= \left[\int_{-\infty}^{\infty} dF(x)\right]^2 \sigma_F^{-2} \sum_i \nu_i(d.-d_i)^2$$

$$= \sigma_F^{-2} \sum_i \nu_i(d.-d_i)^2$$

$$= \lambda_F^2 \quad \text{sodat}$$

$$(3.139)\quad E_{T_T,F} = 1.$$

2). Met betrekking tot voorwaarde (3.109) ondersoek ons twee gevalle.

a). $F(x) = \Phi(x)$, die normaal (0,1) verdelingsfunksie.

Ons het bewys $J'[F(x)] = 1/\phi[J\{F(x)\}]$ sodat vir z_N, die grootste van die x_{ij}-waardes, geld:

$$f(z_N+cN^{-\frac{1}{2}})J'[F(z_N+c'N^{-\frac{1}{2}})] = \phi(z_N+cN^{-\frac{1}{2}})/[\phi(z_N+c'N^{-\frac{1}{2}})]$$

omdat $f(x) = \phi(x)$

$$= \text{eksp}[-\tfrac{1}{2}(z_N+cN^{-\frac{1}{2}})^2+\tfrac{1}{2}(z_N+c'N^{-\frac{1}{2}})^2]$$

$$= \text{eksp}[O(z_N.N^{-\frac{1}{2}})].$$

Vir die normaalverdeling is $z_N = O[(2\log_e N)^{\frac{1}{2}}]$ volgens CRAMéR(1946) p. 374.

Dus $f(z_N+cN^{-\frac{1}{2}})J'[F(z_N+c'N^{-\frac{1}{2}})] \xrightarrow[\infty]{N} e^0 = 1$ waarmee aangetoon is dat aan (3.109) voldoen word, en dus ook aan (3.98) - sien stelling 3.7 opmerking 2.

b). Neem $F(x)$ die Laplace (dubbel eksponensiaal) verdelingsfunksie met frekwensiefunksie

$$f(x) = \tfrac{1}{2}\lambda^{-1}\text{eksp}(-|x|/\lambda) .$$

Volgens CRAMéR(1946) p. 373 is $z_N = O(\lambda\log_e\tfrac{1}{2}N)$.

Laat $u = \Phi^{-1}[F(z_N+c'N^{-\frac{1}{2}})]$, $z_N>0$.

$$\Phi(u) = F(z_N+c'N^{-\frac{1}{2}})$$

$$1-\Phi(u) = 1-F(z_N+c'N^{-\frac{1}{2}}) .$$

Neem die linkerkant en regterkant afsonderlik.

$$\text{L.K.} = \frac{1}{\sqrt{2\pi}}\int_u^\infty \text{eksp}(-\tfrac{1}{2}t^2)\,dt$$

$$= \frac{1}{\sqrt{2\pi}}\,u^{-1}[\text{eksp}(-\tfrac{1}{2}u^2)][1 + O(u^{-2})].$$

$$\text{R.K.} = \frac{1}{2\lambda}\int_{z_N+c'N^{-\frac{1}{2}}}^\infty \text{eksp}(-t/\lambda)\,dt$$

$$= \frac{1}{2\lambda}[-\lambda\text{eksp}(-t/\lambda)]_{z_N+c'N^{-\frac{1}{2}}}^\infty$$

$$= \tfrac{1}{2}\text{eksp}[(-z_N-c'N^{-\frac{1}{2}})/\lambda].$$

Dus $\dfrac{1}{\sqrt{2\pi}}\,u^{-1}[\text{eksp}(-\tfrac{1}{2}u^2)][1 + O(u^{-2})] = \tfrac{1}{2}\text{eksp}[(-z_N-c'N^{-\frac{1}{2}})/\lambda].$

Vir N groot kry ons benaderd

$$u^{-1}\text{eksp}(-\tfrac{1}{2}u^2) = (\pi/2)^{\frac{1}{2}}\text{eksp}(-z_N/\lambda)$$

$$\ln u + \tfrac{1}{2}u^2 = z_N/\lambda - \ln(\pi/2)^{\frac{1}{2}} \qquad \text{waar } \ln u \equiv \log_e u$$

$$\tfrac{1}{2}u^2\Big(1+\frac{\ln u}{\tfrac{1}{2}u^2}\Big) = z_N/\lambda - \ln(\pi/2)^{\frac{1}{2}}$$

$$u\Big(1 + \frac{\ln u}{u^2} +\ldots\Big)= (2z_N/\lambda - \ln\tfrac{\pi}{2})^{\frac{1}{2}}.$$

As eerste benadering is $u = O(\sqrt{z_N})$. Hiervan maak ons gebruik om te skryf

$$u \doteqdot (2z_N/\lambda - \ln \tfrac{\pi}{2})^{\frac{1}{2}} - \frac{\ln u}{u}$$

$$\doteqdot (2z_N/\lambda - \ln \tfrac{\pi}{2})^{\frac{1}{2}} - O\left(\frac{\ln z_N}{\sqrt{z_N}}\right)$$

$$f(z_N+cN^{-\frac{1}{2}})J'[F(z_N+c'N^{-\frac{1}{2}})] = f(z_N+cN^{-\frac{1}{2}})/\ \phi[J\{F(z_N+c'N^{-\frac{1}{2}})\}]$$

uit (3.135)

$$= f(z_N+cN^{-\frac{1}{2}})/\ \phi[\Phi^{-1}\{F(z_N+c'N^{-\frac{1}{2}})\}]$$

$$= f(z_N+cN^{-\frac{1}{2}})/\ \phi(u)$$

$$= \frac{\frac{1}{2}\lambda^{-1}\text{eksp}[-(z_N+cN^{-\frac{1}{2}})/\lambda]}{(2\pi)^{-\frac{1}{2}}\text{eksp}[-\frac{1}{2}\{(2z_N/\lambda - \ln \frac{\pi}{2})^{\frac{1}{2}} - O(z_N^{-\frac{1}{2}} \ln z_N)\}^2]}$$

$$= (\tfrac{\pi}{2})^{\frac{1}{2}}\lambda^{-1}\text{eksp}[-\lambda^{-1}(z_N+cN^{-\frac{1}{2}})+\tfrac{1}{2}\{(2z_N/\lambda - \ln\tfrac{\pi}{2})^{\frac{1}{2}}-O(z_N^{-\frac{1}{2}}\ln z_N)\}^2]$$

$$= (\tfrac{\pi}{2})^{\frac{1}{2}}\lambda^{-1}\text{eksp}[-\lambda^{-1}(z_N+cN^{-\frac{1}{2}})+\tfrac{1}{2}\{2z_N/\lambda - \ln\tfrac{\pi}{2} + O(z_N^{-\frac{1}{2}}\ln z_N)^2 - O(\ln z_N)\}]$$

$$= (\tfrac{\pi}{2})^{\frac{1}{2}}\lambda^{-1}\text{eksp}[-z_N/\lambda - cN^{-\frac{1}{2}}+z_N/\lambda - \tfrac{1}{2}\ln\tfrac{\pi}{2} + O(z_N^{-\frac{1}{2}}\ln z_N)^2]\cdot \cdot\text{eksp}[-O(\ln z_N)]$$

$$\xrightarrow[\infty]{N} 0$$

waarmee aangetoon is dat (3.109), en dus ook (3.98), bevredig word.

3.6.2.4. $\Psi_N(\delta_{ij}) \equiv Q(\delta_{ij})$ - sien §2.5.2.3.

Hier is

(2.73) $t_i' = \sum_j Q[r_{ij}/(N+1)]$ met toetsingsgrootheid T_Q gegee in (2.76). Stel nou

(3.140) $J_N(H) = \Psi[NH/(N+1)]$ vir $0<H<1$, dan is

$$J(H) = \lim_{N\to\infty} \Psi[NH/(N+1)]$$

$= \Phi^{-1}(H)$ waar Φ die normaal (0,1) - verdelingsfunksie is, sodat

(3.135) $J'(u) = 1/\ \phi[J(u)]$ soos in TERRY(1952) se geval.

Aan alle voorwaardes van stelling 3.9 word voldoen - sien onder andere C+S(1958) §6 B voorbeeld 2.

Uit STOKER(1955) p. 57 volg nou dat

$$\lim_{N\to\infty} N^{-1} \sum_h (\Psi_{Nh}-\bar{\Psi}_N)^2 = \lim_{N\to\infty} N^{-1} \sum_h Q^2(\delta_h) = 1 .$$

Vir enige willekeurige toelaatbare verdelingsfunksie F geld nou uit (3.123) dat

$$(3.141)\quad \lambda_Q^2 = [\int_{-\infty}^{\infty} f(x)J'[F(x)]dF(x)]^2 \sum_i \nu_i(d.-d_i)^2$$

$$= \left[\int_{-\infty}^{\infty} \frac{f(x)}{\phi[J\{F(x)\}]}\, dF(x)\right]^2 \sum_i \nu_i(d.-d_i)^2$$

wat dieselfde is as λ_T^2 .

Opmerkings.

1). T_T en T_Q is albei, onder dieselfde voorwaardes, asimptoties χ^2 verdeel met (k-1) g.v.v. as $N\to\infty$ en nie-sentraliteitsparameter

$$\left[\int_{-\infty}^{\infty} \frac{f(x)}{\phi[J\{F(x)\}]}\, dF(x)\right]^2 \sum_i \nu_i(d.-d_i)^2.$$

Die a.r.d.-eienskappe van T_Q is dus dieselfde as dié van T_T, soos in die twee-steekproef geval - sien C+S(1958) §6 B voorbeeld 2.

2). Die resultate van §3.6.2.3 opmerking 2 geld ook vir hierdie geval.

3). Gestel dat ons met 'n gewysigde Van der Waerden toetsingsgrootheid werk, en wel waar $J = F^{-1}$ 11).

Dan is $u = F[J(u)]$

$$1 = f[J(u)]J'(u)$$

$$J'(u) = 1/f[J(u)] .$$

$$f(x+cN^{-\frac{1}{2}})J'[F(x+c'N^{-\frac{1}{2}})] = f(x+cN^{-\frac{1}{2}})/ f[F^{-1}\{F(x+c'N^{-\frac{1}{2}})\}]$$

$$= f(x+cN^{-\frac{1}{2}})/ f(x+c'N^{-\frac{1}{2}}) .$$

Neem F die Laplace-verdelingsfunksie, dan is volgens CRAMéR(1946) p. 373 $z_N = O(\lambda \log_e \frac{N}{2})$ sodat

$$f(z_N+cN^{-\frac{1}{2}})J'[F(z_N+c'N^{-\frac{1}{2}})] = K\ \text{eksp}[-(z_N+cN^{-\frac{1}{2}})/\lambda+(z_N+c'N^{-\frac{1}{2}})/\lambda]$$

$$= K\ \text{eksp}[O(N^{-\frac{1}{2}})]$$

$$\xrightarrow[\infty]{N} K < \infty.$$

11). Hierdie F moet voldoen aan (3.10) - (3.14) en (3.93).

Ook in hierdie geval word aan voorwaarde (3.109), en dus ook aan (3.98), voldoen.

4). HODGES en LEHMANN(1961) het vir die twee-steekproef geval die a.r.d. van Wilcoxon se toets m.b.t. Van der Waerden se toets $[E_{W,Q}(F)]$ bereken vir verskillende verdelingsfunksies F. Hulle het bevind dat

$$E_{W,Q}(F) \leq 6/\pi \ .$$

3.6.3. Toetse vir Verskil in Verspreiding.

3.6.3.1. 'n Algemene uitdrukking vir die asimptotiese relatiewe doeltreffendheid van een toets m.b.t. 'n tweede toets.

Beskou die alternatiewe hipotese

(3.142) H_a: $F_i(x) = F[x(1+d_i N^{-\frac{1}{2}})]$ met

$\int_{-\infty}^{\infty} x \, dF_i(x) = 0, \ i=1,2,\ldots,k$ 12) waar die d_i 's willekeurige eindige reële getalle is.

Om slegs verspreidingseffekte te toets (en nie ook verskuiwingseffekte nie), moet alle verdelings gelyke mediane of gemiddeldes besit, of alle waarnemings moet om 'n lokaliteitsparameter gereduseer word. Die probleem in hierdie geval is egter dat die populasie-parameters meestal onbekend is - sien byvoorbeeld §5.3.1, SUKHATME(1958), ANSARI en BRADLEY(1960) §9 en CROUSE(1960) §4.12 .

Neem aan dat $f(x) = F'(x)$ orals bestaan en kontinu is en dat f(x) gelykmatig begrens is.

---

12). CROUSE(1960) werk met die alternatiewe hipotese

$$F_i(x) = F[(x-\mu)(1+d_i N^{-\frac{1}{2}})]$$

waar $\mu = \int x \, dF_i(x), \ \ i=1,2,\ldots,k.$

STELLING 3.10.

Indien

(3.92) $\Psi_N[N/(N+1)] = o(N^{\frac{1}{2}})$,

(3.93) $\int_{I_N} |J_N[H_N(x)]-\Psi_N[NH_N(x)/(N+1)]|dF_{in_i}(x) = o_p(N^{-\frac{1}{2}})$,

i=1,2,...,k en

(3.143) $|f(x+\delta_i xN^{-\frac{1}{2}})J'[\sum_{i'}\nu_{Ni'} F(x+\delta_{i'} xN^{-\frac{1}{2}})]| < M(x)$ vir

N voldoende groot en $\int_{-\infty}^{\infty} x\ M(x)\ dF(x) < \infty$ waar die δ_i 's willekeurige eindige reële getalle is,

dan geld

(3.144) $$\lim_{N\to\infty} N^{\frac{1}{2}}[E(t_i'/n_i|H_a) - E(t_i'/n_i|H_o)] =$$

$$= (d.-d_i) \int_{-\infty}^{\infty} xf(x)J'[F(x)]dF(x).$$

Bewys: Onder H_a is

(3.145) $$\begin{aligned} F_{i'}(x) &= F(x+d_{i'}\ xN^{-\frac{1}{2}}) \\ &= F[x+d_i xN^{-\frac{1}{2}}+(d_{i'}-d_i)xN^{-\frac{1}{2}}] \\ &= F_i(x+\Theta_{i'}\ xN^{-\frac{1}{2}}) \quad \text{waar} \end{aligned}$$

(3.102) $\Theta_{i'} = d_{i'}-d_i$, $i' =1,2,...,k$, sodat

(3.146) $H(x) = \sum_{i'}\nu_{Ni'}\ F_i(x+\Theta_{i'}\ xN^{-\frac{1}{2}})$ uit (3.7).

Stel nou

(3.147) $$G(\Theta_1,\Theta_2,...,\Theta_k) = N^{\frac{1}{2}} \int_{-\infty}^{\infty} J[H(x)]dF_i(x)$$

$$= N^{\frac{1}{2}} \int_{-\infty}^{\infty} J[\sum_{i'}\nu_{Ni'}\ F_i(x+\Theta_{i'}\ xN^{-\frac{1}{2}})]dF_i(x) \text{ uit (3.146),}$$

dan is

(3.148) $$G(0,0,...,0) = N^{\frac{1}{2}} \int_{-\infty}^{\infty} J[F(x)]dF(x).$$

Op soortgelyke wyse as in stelling 3.7 volg met behulp van (3.94), (3.104), (3.147) en (3.148) dat

$$\lim_{N\to\infty} N^{\frac{1}{2}}[E(t_i'/n_i|H_a)-E(t_i'/n_i|H_o)] =$$

$$= (d.-d_i) \int_{-\infty}^{\infty} xf(x)J'[F(x)]dF(x).$$

STELLING 3.1.

Indien voorwaardes (3.10) tot (3.14) geld en

(3.28) $\operatorname{var}(\xi_{ij}) > 0$ vir alle i en j,

dan geld onder H vir vaste F_i en ν_{Ni} $(i=1,2,\ldots,k)$ dat

$$\lim_{N\to\infty} P\left[\frac{\underline{S}_N - E\underline{S}_N}{\sigma_{\underline{S}_N}} \le t\right] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{1}{2}x^2}\, dx$$

waar

(3.29) $\lim_{N\to\infty} E\underline{S}_N = \sum_i c_i \int_I J(H)\,dF_i$ en

(3.30) $\lim_{N\to\infty} N\sigma^2_{\underline{S}_N} = \lim_{N\to\infty} \sum_i \nu_{Ni}^{-1} \sum_g \sum_{g'} (c_i\nu_{Ng} - c_g\nu_{Ni})(c_i\nu_{Ng'} - c_{g'}\nu_{Ni}) \cdot \sigma_{B_{gi}B_{g'i}}$.

Bewys: Die bewys gaan aan die hand van 'n aantal hulpteoremas.

Opmerkings.

1). Stelling 3.1 is 'n uitbreiding van stelling 1 van CHERNOFF en SAVAGE (1958) van twee na k steekproewe. Die bewys geskied op 'n analoë wyse en verskeie resultate van bogenoemde stelling word hier gebruik.

2). Voorwaardes (3.10) tot (3.14) ten opsigte van J is fundamenteel en word deurgaans in hierdie hoofstuk aanvaar (by alle lemmas en stellings).

3). In lemma 3.9 opmerking 1 word aangetoon dat

(3.31) $\lim_{N\to\infty} N\sigma^2_{S_N} > 0$ onder (3.3) en (3.28).

4). Voorwaarde (3.28) vereis dat $\operatorname{var}(\xi_{ij})>0$ vir alle i, waarby

(3.23a) $\operatorname{var}(\xi_{ij}) = \nu_{Ni}^{-2} \sum_g \sum_{g'} (c_i\nu_{Ng} - c_g\nu_{Ni})(c_i\nu_{Ng'} - c_{g'}\nu_{Ni})\sigma_{B_{gi}B_{g'i}}$

en

(3.22) $\sigma_{B_{gi}B_{g'i}} = \iint_{-\infty<x<y<\infty} F_i(x)[1-F_i(y)]J'[H(x)]J'[H(y)] \cdot [dF_g(x)dF_{g'}(y) + dF_{g'}(x)dF_g(y)]$.

Dui nou die frekwensiefunksie van die verdeling met verdelingsfunksie F_g aan deur f_g, aannemende die afgeleides van die F_i $(i=1,2,\ldots,k)$ bestaan. Indien die variasiegebiede van die verdelings met verdelingsfunksies F_g en $F_{g'}$ 'n gemeenskaplike gebied bevat van maat (ten opsigte van beide verdelings) > 0 (m.a.w. nie disjunk is nie), sê ons die frekwensiefunksies f_g en $f_{g'}$ oorvleuel mekaar.

Indien f_i geeneen van f_g of $f_{g'}$ oorvleuel nie, is $\sigma_{B_{gi}B_{g'i}} = 0$.* Indien f_i slegs f_g oorvleuel (en nie ook $f_{g'}$ nie), is $\sigma^2_{B_{gi}} \equiv \sigma_{B_{gi}B_{gi}} > 0$ en $\sigma_{B_{gi}B_{g'i}} = 0$ (selfs indien f_g en $f_{g'}$ mekaar oorvleuel).

Alleen indien f_i sowel f_g as $f_{g'}$ oorvleuel, is $\sigma_{B_{gi}B_{g'i}} > 0$, $g' \neq g$.

$\mathrm{var}(\xi_{ij}) > 0$ impliseer dus dat daar tenminste een waarde van $g \neq i$ bestaan (L.W. as $g=i$ is die koëffisiënt van $\sigma_{B_{gi}B_{g'i}}$ in (3.23a) gelyk aan nul en ook as $g'=i$) waarvoor $\sigma^2_{B_{gi}} > 0$. Dit beteken dus dat die frekwensiefunksies f_i en f_g mekaar oorvleuel. Elke frekwensiefunksie f_i moet deur tenminste één ander frekwensiefunksie oorvleuel word $(\mathrm{var}(\xi_{ij})>0$ vir alle $i=1,2,\ldots,k)$. Daar kan dus onder voorwaarde (3.28) geen frekwensiefunksies wees wat alleen lê nie.

Sien ook die opmerking na die bewys van stelling 3.3.

Lemma 3.4. S_N kan soos volg opgesplits word:

$$(3.32)\quad S_N = A + B_{1N} + B_{2N} + \sum_{m=1}^{5} C_{mN}$$

waar

$$(3.33)\quad A = \sum_i c_i \int_I J(H)\,dF_i$$

$$(3.34)\quad B_{1N} = \sum_i c_i \int_I J(H)\,d(F_{in_i} - F_i)$$

$$(3.35)\quad B_{2N} = \sum_i c_i \int_I (H_N - H)J'(H)\,dF_i$$

$$(3.36)\quad C_{1N} = \sum_i c_i \int_I (H_N - H)J'(H)\,d(F_{in_i} - F_i)$$

$$(3.37)\quad C_{2N} = \sum_i c_i \int_{I_N} \tfrac{1}{2}(H_N - H)^2 J''[\phi H_N + (1-\phi)H]\,dF_{in_i}, \quad 0<\phi<1$$

---

* Die geval waar die variasiegebied van een verdeling bevat word in die variasiegebied van 'n tweede verdeling sonder dat hulle mekaar oorvleuel, word uitgesluit.

(3.38) $C_{3N} = -\sum_i c_i \int_{H_N=1} [J(H)+(H_N-H)J'(H)]dF_{in_i}$

(3.39) $C_{4N} = \sum_i c_i \int_{I_N} [J_N(H_N)-J(H_N)]dF_{in_i}$

(3.40) $C_{5N} = \sum_i c_i \int_{H_N=1} J_N(H_N)dF_{in_i}$.

Bewys: $S_N = \sum_i c_i \int_{-\infty}^{\infty} J_N[H_N(x)]dF_{in_i}(x)$

$$= \sum_i c_i \int_{I_N} [J_N(H_N)-J(H_N)]dF_{in_i} + \sum_i c_i \int_{I_N} J(H_N)dF_{in_i} + \sum_i c_i \int_{H_N=1} J_N(H_N)dF_{in_i}$$

$$= C_{4N}+\sum_i c_i \int_{I_N} \left\{J(H)+(H_N-H)J'(H)+\tfrac{1}{2}(H_N-H)^2J''[\emptyset H_N+(1-\emptyset)H]\right\}dF_{in_i} + C_{5N} \quad \text{met } 0<\emptyset<1$$

$$= C_{4N}+\sum_i c_i \int_{0<H<1} [J(H)+(H_N-H)J'(H)]d(F_{in_i}-F_i+F_i) - \sum_i c_i \int_{H_N=1} [J(H)+(H_N-H)J'(H)]dF_{in_i} + \sum_i c_i \int_{I_N} \tfrac{1}{2}(H_N-H)^2J''[\emptyset H_N+(1-\emptyset)H]dF_{in_i} + C_{5N}$$

$$= C_{4N}+\sum_i c_i \int_{0<H<1} J(H)dF_i+\sum_i c_i \int_{0<H<1} J(H)d(F_{in_i}-F_i) + \sum_i c_i \int_{0<H<1} (H_N-H)J'(H)dF_i+\sum_i c_i \int_{0<H<1} (H_N-H)J'(H)d(F_{in_i}-F_i) + C_{3N} + C_{2N} + C_{5N}$$

$$= C_{4N}+A+B_{1N}+B_{2N}+C_{1N}+C_{3N}+C_{2N}+C_{5N} .$$

Opmerking.

Die A-, B- en C-terme stel die „konstante", „eerste orde stogastiese" en „tweede orde stogastiese" terme van S_N voor.

Lemma 3.5. Die term A is eindig.

Bewys: $A = \sum_i c_i \int_I J(H)dF_i$ uit (3.33).

Uit (3.7) volg dat $F_i(x) \leq \nu_{Ni}^{-1}H(x)$ – sien vergelyking (3.50) – sodat:

$$(3.41) \quad \left| \int_{-\infty}^{\infty} J[H(x)]dF_i(x) \right| \leq K \int_0^1 [H(1-H)]^{\delta-\frac{1}{2}} dH \quad \text{uit (3.3) en (3.14)}$$

$$= K\, B(\delta+\tfrac{1}{2}, \delta+\tfrac{1}{2}) \quad {}^{6)}$$

$$= K' < \infty.$$

Omdat die c_i's eindig is en die sommasie oor 'n eindige aantal terme gaan, volg uit bostaande dat A bestaan en eindig is.

Lemma 3.6. Vir I: $0<H<1$ geld

$$(3.42) \quad \int_I \left[\int_{x_0}^{x} J'[H(x)]dH(x)\right] d[F_{in_i}(x)-F_i(x)]$$

$$= \int_I J[H(x)]d[F_{in_i}(x)-F_i(x)].$$

Bewys: $\int_I \left[\int_{x_0}^{x} J'[H(x)]dH(x)\right] d[F_{in_i}(x) - F_i(x)]$

$$= \int_I J[H(x)]d[F_{in_i}(x) - F_i(x)]$$

$$-J[H(x_0)] \int_I d[F_{in_i}(x) - F_i(x)]$$

$$= \int_I J(H)d(F_{in_i} - F_i).$$

Lemma 3.7. Vir $\xi_{ij} = \nu_{Ni}^{-1} \sum_g (c_i \nu_{Ng} - c_g \nu_{Ni}) \tilde{B}_g(x_{ij})$ geld onder voorwaarde (3.28):

$$(3.43) \quad 0 < \operatorname{var}(\xi_{ij}) < \infty.$$

Bewys: Uit (3.22) volg

$$\sigma_{B_{gi}B_{g'i}} = \iint_{-\infty<x<y<\infty} F_i(x)[1-F_i(y)]J'[H(x)]J'[H(y)] \cdot [dF_g(x)dF_{g'}(y) + dF_{g'}(x)dF_g(y)]$$

$$\leq 2K \iint_{-\infty<x<y<\infty} H(x)[1-H(y)]J'[H(x)]J'[H(y)]dH(x)dH(y)$$

(sien (3.52) en (3.53)), sodat m.b.v. (3.14) volg:

$\sigma_{B_{gi}B_{g'i}} < \infty$ vir alle g, g' en i (vergelyk CHERNOFF en SAVAGE(1958) p. 989).

---

6). B dui die gewone betafunksie aan en die K 's konstante getalle.

Uit (3.3) en (3.23a) volg dus dat

(3.44) $\mathrm{var}(\xi_{ij}) < \infty$.

Die resultaat volg m.b.v. (3.28).

STELLING 3.2.

Indien

(3.28) $\mathrm{var}(\xi_{ij}) > 0$ vir alle i en j,

dan is $\sqrt{N}(B_{1N} + B_{2N})$ onder H asimptoties ormaal verdeel indien $N \to \infty$.

Bewys: Uit (3.34) en (3.35) volg

(3.45) $B_{1N}+B_{2N} = \sum_i c_i \int_I J(H)d(F_{in_i}-F_i) + \sum_i c_i \int_I (H_N-H)J'(H)dF_i$

$$= \sum_i c_i \int_I [\int_{x_0}^{x} J'[H(x)]dH(x)]d[F_{in_i}(x) - F_i(x)] +$$

$$+ \sum_i c_i [\{H_N(x)-H(x)\} \int_{x_0}^{x} J'[H(x)]dF_i(x)]_{-\infty}^{\infty} -$$

$$- \sum_i c_i \int_I [\int_{x_0}^{x} J'[H(x)]dF_i(x)]d[H_N(x)-H(x)]$$ deur parsiële integrasie en m.b.v. lemma 3.6. Uit lemma 3.2 en vergelykings (3.6) en (3.7) volg verder met waarskynlikheid een

$$= \sum_i c_i \int_I [\int_{x_0}^{x} J'[H(x)]d\sum_g \nu_{Ng}F_g(x)]d[F_{in_i}(x) - F_i(x)] -$$

$$- \sum_i c_i \int_I [\int_{x_0}^{x} J'[H(x)]dF_i(x)]d[\sum_g \nu_{Ng}F_{gn_g}(x) - \sum_g \nu_{Ng}F_g(x)]$$

$$= \sum_i c_i \sum_g \nu_{Ng} \int_I [\int_{x_0}^{x} J'[H(x)]dF_g(x)]d[F_{in_i}(x) - F_i(x)] -$$

$$- \sum_i c_i \int_I [\int_{x_0}^{x} J'[H(x)]dF_i(x)] \sum_g \nu_{Ng} d[F_{gn_g}(x) - F_g(x)]$$

want $d(\sum_g \nu_{Ng}F_g) = \sum_g \nu_{Ng}dF_g$

$$= \sum_i c_i \sum_g \nu_{Ng} \int_I B_g(x)d[F_{in_i}(x) - F_i(x)] -$$

$$- \sum_i c_i \sum_g \nu_{Ng} \int_I B_i(x)d[F_{gn_g}(x) - F_g(x)]$$ uit (3.16)

$$= \sum_i c_i \sum_g \nu_{Ng}[\int_I B_g dF_{in_i} - \int_I B_g dF_i - \int_I B_i dF_{gn_g} + \int_I B_i dF_g]$$

$$= \sum_i c_i \sum_g \nu_{Ng}[n_i^{-1} \sum_{j=1}^{n_i} \{B_g(x_{ij})-EB_g(x_{ij})\} - n_g^{-1} \sum_{j=1}^{n_g} \{B_i(x_{gj})-EB_i(x_{gj})\}]$$

$$= \sum_i c_i \sum_g \nu_{Ng} n_i^{-1} \sum_{j=1}^{n_i} \tilde{B}_g(x_{ij}) - \sum_i c_i \sum_g \nu_{Ng} n_g^{-1} \sum_{j=1}^{n_g} \tilde{B}_i(x_{gj}) \quad \text{uit (3.17)}$$

$$= \sum_i c_i \sum_g \nu_{Ng} n_i^{-1} \sum_{j=1}^{n_i} \tilde{B}_g(x_{ij}) - \sum_g c_g \sum_i \nu_{Ni} n_i^{-1} \sum_{j=1}^{n_i} \tilde{B}_g(x_{ij}) \; .$$

Omdat die sommasies in die vorige reël onafhanklik vanmekaar is, kan die indekse in die tweede uitdrukking verwissel word. Verder kan eindige sommasies verwissel word, sodat bostaande

$$= \sum_i \sum_g c_i \nu_{Ng} n_i^{-1} \sum_j \tilde{B}_g(x_{ij}) - \sum_i \sum_g c_g \nu_{Ni} n_i^{-1} \sum_j \tilde{B}_g(x_{ij})$$

$$= \sum_i [n_i^{-1} \sum_j \sum_g (c_i \nu_{Ng} - c_g \nu_{Ni}) \tilde{B}_g(x_{ij})]$$

$$= N^{-1} \sum_i \sum_j \xi_{ij} \quad \text{m.b.v. (3.18)}$$

$$= N^{-1} \sum_i \xi_i \quad \text{sodat}$$

(3.46) $\sqrt{N}(B_{1N} + B_{2N}) = N^{-\frac{1}{2}} \sum_i \xi_i$ met waarskynlikheid een.

Vir vaste i is daar n_i onderling onafhanklike, identies verdeelde variante ξ_{ij}, $j=1,2,\ldots,n_i$, elk met eindige positiewe variansie (sien lemma 3.7) soos gegee in (3.23a).

Volgens die sentrale limietstelling (CRAMéR(1946) p. 215) is die som $N^{-\frac{1}{2}}\xi_i = N^{-\frac{1}{2}} \sum_{j=1}^{n_i} \xi_{ij}$ asimptoties normaal verdeel as $n_i \to \infty$ $(0<\nu_0 \leq n_i/N \leq 1-\nu_0<1)$.

Die variante $N^{-\frac{1}{2}}\xi_i$ en $N^{-\frac{1}{2}}\xi_{i'}$ is onderling onafhanklik indien $i' \neq i$. Omdat elk van die k variante $N^{-\frac{1}{2}}\xi_i$ asimptoties normaal verdeel is, besit die som $\sqrt{N}(B_{1N}+B_{2N})=N^{-\frac{1}{2}}\sum_i \xi_i$ asimptoties 'n normaalverdeling as $N \to \infty$ [7].

---

7). Voorwaarde (3.3) impliseer dat alle $n_i \to \infty$ as $N \to \infty$.

Lemma 3.8. Die asimptotiese verwagtingswaarde van S_N word gegee deur

(3.29) $\lim_{N\to\infty} E(S_N) = \sum_i c_i \int_I J(H)dF_i$.

Bewys: $E(B_{1N}+B_{2N}) = E(N^{-1} \sum_i \sum_j \xi_{ij})$ uit (3.45)

$= N^{-1} \sum_i \sum_j E(\xi_{ij})$

$= 0$ uit (3.20).

Elk van die C-terme $= o_p(N^{-\frac{1}{2}})$ - sien lemma 3.10 .

Gevolglik is $\lim_{N\to\infty} E(S_N) = A$ uit (3.32)

$= \sum_i c_i \int_I J(H)dF_i$ uit (3.33).

Opmerking.

$\lim_{N\to\infty} E(S_N) = A$ is eindig - sien lemma 3.5.

Lemma 3.9. Die asimptotiese waarde van die variansie van $N^{\frac{1}{2}}S_N$ word gegee deur:

(3.30) $\lim_{N\to\infty} N\sigma^2_{S_N} = \lim_{N\to\infty} \sum_i \nu^{-1}_{Ni} \sum_g \sum_{g'} (c_i\nu_{Ng}-c_g\nu_{Ni})(c_i\nu_{Ng'}-c_{g'}\nu_{Ni})\cdot \sigma_{B_{gi}B_{g'i}}$

waar

(3.22) $\sigma_{B_{gi}B_{g'i}} = \int\int_{-\infty<x<y<\infty} F_i(x)[1-F_i(y)]J'[H(x)]J'[H(y)]\cdot [dF_g(x)dF_{g'}(y) + dF_{g'}(x)dF_g(y)]$.

Bewys: $\mathrm{var}[\sqrt{N}(B_{1N}+B_{2N})] = \mathrm{var}(N^{-\frac{1}{2}}\sum_i \xi_i)$ uit (3.46)

$= \sum_i \nu^{-1}_{Ni} \sum_g \sum_{g'} (c_i\nu_{Ng}-c_g\nu_{Ni})(c_i\nu_{Ng'}-c_{g'}\nu_{Ni})\sigma_{B_{gi}B_{g'i}}$ uit (3.25) en (3.27).

Omdat A konstant en $N^{\frac{1}{2}}$ x die C-terme asimptoties in waarskynlikheid gelyk is aan nul (sien lemma 3.10), geld:

(3.47) $\lim_{N\to\infty} N\sigma^2_{S_N} = \lim_{N\to\infty} N\mathrm{var}(B_{1N}+B_{2N})$

$= \lim_{N\to\infty} \sum_i \nu^{-1}_{Ni} \sum_g \sum_{g'} (c_i\nu_{Ng}-c_g\nu_{Ni})(c_i\nu_{Ng'}-c_{g'}\nu_{Ni})\sigma_{B_{gi}B_{g'i}}$

waar $\sigma_{B_{gi}B_{g'i}}$ gegee is in (3.22).

Opmerkings.

1). Met behulp van (3.23a) volg uit (3.47):

(3.48) $\lim\limits_{N\to\infty} N\sigma^2_{S_N} = \lim\limits_{N\to\infty} \sum\limits_i \nu_{Ni}\mathrm{var}(\xi_{ij})$.

Deurdat $0<\nu_o\leq\nu_{Ni}\leq 1-\nu_o<1$ vir alle i, en die sommasie oor 'n eindige aantal terme gaan, volg m.b.v. (3.43) dat

(3.49) $0 < \lim\limits_{N\to\infty} N\sigma^2_{S_N} < \infty$.

2). Stel $k=2$, $c_1=1$, $c_2=0$, $\nu_1=\lambda$, $\nu_2=1-\lambda$, $F_1=F$ en $F_2=G$, dan reduseer $\lim\limits_{N\to\infty} N\sigma^2_{S_N}$ tot:

$$\lim_{N\to\infty} N\sigma^2_{S_N} = (1-\lambda)\sigma^2_{B_{12}} + \frac{(1-\lambda)^2}{\lambda}\sigma^2_{B_{21}} \quad \text{waar}$$

$$\sigma^2_{B_{12}} = 2 \iint\limits_{-\infty<x<y<\infty} G(x)[1-G(y)]J'[H(x)]J'[H(y)]dF(x)dF(y) \quad \text{en}$$

$$\sigma^2_{B_{21}} = 2 \iint\limits_{-\infty<x<y<\infty} F(x)[1-F(y)]J'[H(x)]J'[H(y)]dG(x)dG(y)$$

wat presies dieselfde is as $\lim\limits_{N\to\infty} N\sigma^2_{T_N}$ van CHERNOFF en SAVAGE(1958) [8].

Lemma 3.10. Die C-terme is almal asimptoties in waarskynlikheid van kleiner orde as $N^{-\frac{1}{2}}$.

Bewys: Die volgende resultate word gebruik, naamlik (sien C+S(1958)):

Vir alle waardes van i geld:

(3.50) $H \geq \nu_{Ni}F_i \geq \nu_o F_i$,

(3.51) $1-F_i \leq (1-H)/\nu_{Ni} \leq (1-H)/\nu_o$,

(3.52) $F_i(1-F_i) \leq H(1-H)\nu_{Ni}^{-2} \leq H(1-H)\nu_o^{-2}$ en

(3.53) $dH \geq \nu_{Ni}dF_i \geq \nu_o dF_i$.

Laat (a_N, b_N) die interval $S_{N\varepsilon}$ wees waar

(3.54) $S_{N\varepsilon} = \left\{x\colon H(1-H) > \eta_\varepsilon\nu_o/N\right\}$.

η_ε kan, onafhanklik van F_i en ν_{Ni}, so gekies word dat $P[\underline{x}_{ij} \in S_{N\varepsilon},\ j=1,2,\ldots,n_i;\ i=1,2,\ldots,k] \geq 1-\varepsilon$.

Uit vergelykings (3.36) tot (3.40) volg agtereenvolgens:

8). Voortaan skryf ons kortweg C+S(1958) in plaas van CHERNOFF en SAVAGE(1958).

$$C_{1N} = \sum_i c_i \int_I (H_N-H)J'(H)d(F_{in_i}-F_i)$$

$$= \sum_i c_i \int_I (\sum_g \nu_{Ng}F_{gn_g} - \sum_g \nu_{Ng}F_g)J'(H)d(F_{in_i}-F_i) \quad \text{m.b.v. (3.6) en (3.7)}$$

$$= \sum_i c_i \int_I \sum_g \nu_{Ng}(F_{gn_g}-F_g)J'(H)d(F_{in_i}-F_i)$$

$$= \sum_i \sum_g c_i \nu_{Ng} \int_I (F_{gn_g}-F_g)J'(H)d(F_{in_i}-F_i)$$

$$= \sum_i c_i \nu_{Ni} \int_I (F_{in_i}-F_i)J'(H)d(F_{in_i}-F_i)$$
$$+ \sum_{i \neq g}\sum c_i \nu_{Ng} \int_I (F_{gn_g}-F_g)J'(H)d(F_{in_i}-F_i)$$

$= o_p(N^{-\frac{1}{2}})$ op analoë wyse as by C+S(1958) se C_{1N}- en C_{2N}-terme.

$$C_{2N} = \sum_i c_i \int_{I_N} \tfrac{1}{2}(H_N-H)^2 J''[\emptyset H_N+(1-\emptyset)H]dF_{in_i}, \quad 0<\emptyset<1$$

$= o_p(N^{-\frac{1}{2}})$ netsoos C+S(1958) se C_{3N}-term.

$$-C_{3N} = \sum_i c_i \int_{H_N=1} [J(H)+(H_N-H)J'(H)]dF_{in_i}$$

$= o_p(N^{-\frac{1}{2}})$ soos C+S(1958) se C_{4N}-term.

$$C_{4N} = \sum_i c_i \int_{I_N} [J_N(H_N)-J(H_N)]dF_{in_i}$$

$= o_p(N^{-\frac{1}{2}})$ uit (3.12).

$$C_{5N} = \sum_i c_i \int_{H_N=1} J_N(H_N)dF_{in_i}$$

$= o(N^{\frac{1}{2}})O(N^{-1})$ uit (3.13)

$= o(N^{-\frac{1}{2}})$.

Bewys van stelling 3.1.

Volgens stelling 3.2 is $\sqrt{N}(B_{1N}+B_{2N})$ asimptoties normaal verdeel as $N\to\infty$, en wel met positiewe eindige variansie (lemma 3.9 opmerking 1).

Omdat A eindig is (lemma 3.5) en die C-terme asimptoties in waarskynlikheid van kleiner orde as $N^{-\frac{1}{2}}$ is (lemma 3.10), is $N^{\frac{1}{2}}S_N$ ook asimptoties normaal verdeel met asimptotiese verwagtingswaarde en variansie soos gegee in lemmas 3.8 en 3.9, want $E(B_{1N}+B_{2N}) = 0$.

$[N^{\frac{1}{2}}\underline{S}_N - E(N^{\frac{1}{2}}\underline{S}_N)][\mathrm{var}(N^{\frac{1}{2}}\underline{S}_N)]^{-\frac{1}{2}}$ is dus asimptoties normaal (0,1) verdeel. Gevolglik is

$[S_N - E(S_N)][\mathrm{var}(S_N)]^{-\frac{1}{2}}$ asimptoties normaal (0,1) verdeel as $N \to \infty$.

Hiermee is stelling 3.1 bewys.

Opmerking.

Die bewys kan uitgebrei word na die geval waar die F_i en ν_{Ni} nie vas is nie. Onder die voorwaardes van stelling 3.1 geld die asimptotiese normaliteit gelykmatig met betrekking tot die F_i en ν_{Ni}, $0<\nu_0 \leq \nu_{Ni} \leq 1-\nu_0 < 1$.

Die bewys van hierdie bewering gaan analoog aan dié van „Corollary 1" van C+S(1958). Soos in C+S(1958) kan bewys word dat $E|B_i(x)|^{2+\delta} < \infty$ en met behulp van lemma 3.11 hieronder kan verder bewys word dat $E|\xi_{ij}|^{2+\delta} < \infty$.

Lemma 3.11. Indien vir 'n ry funksies $G_g(x)$, $g=1,2,\ldots,k$ geld $E|G_g(x)|^{2+\delta} = O(1)$ vir $\delta>0$ en k eindig, dan volg uit 'n stelling van Taylor (TAYLOR, A. E. (1958) „Introduction to Functional Analysis" p. 16) dat

$$E|\sum_g G_g(x)|^{2+\delta} = O(1).$$

Lemma 3.12. (C+S(1958)): As $J_N(\frac{h}{N})$ die verwagtingswaarde is van die h^{de} waarde in rang van 'n ewekansige steekproef van grootte N uit 'n populasie met verdelingsfunksie die inverse funksie van J en

$$|J^{(m)}(U)| \leq K[U(1-U)]^{-m-\frac{1}{2}+\delta} \quad \text{vir } \delta>0 \text{ en } m=0,1,2, \quad \text{dan is}$$

$$\lim_{N\to\infty} J_N(H) = J(H) \qquad 0<H<1,$$

$$J_N(1) = o(N^{\frac{1}{2}}) \qquad \text{en}$$

$$\int_{I_N} [J_N(H_N) - J(H_N)]dF_{in_i} = o_p(N^{-\frac{1}{2}}).$$

Bewys: Die lemma volg regstreeks uit C+S(1958) se teorema 2.

3.4. DIE TOETSINGSGROOTHEID T_k.

Stel

$$(3.55)\quad t_i = \int_{-\infty}^{\infty} J_N[H_N(x)]dF_{in_i}(x)$$

waarby $F_{in_i}(x)$ en $H_N(x)$ in (3.5) en (3.6) gedefinieer is en J_N 'n willekeurige funksie is wat voldoen aan voorwaardes (3.10) tot (3.14).

Laat

$$(3.56)\quad \hat{t}_i = (N-n_i)^{\frac{1}{2}}[t_i-E(t_i|H)][N\mathrm{var}(t_i|H)]^{-\frac{1}{2}} .$$

Dan definieer ons as toetsingsgrootheid:

$$(3.57)\quad T_k = \sum_{i=1}^{k} \hat{t}_i^2 .$$

Opmerkings.

1). Die groothede t_i, $\hat{t}_i$ en T_k is almal afhanklik van N.

2). Uit (3.9) volg dat $S_N = t_i$ as

$$(3.58)\quad c_{i'} = \begin{cases} 1 \text{ indien } i' = i \\ 0 \text{ indien } i' \neq i \end{cases} .$$

3). t_i kan geskryf word in die vorm

$$(3.59)\quad t_i = A^{(i)} + B_{1N}^{(i)} + B_{2N}^{(i)} + C^{(i)}$$

netsoos S_N. Omdat die $A^{(i)}$-terme in hierdie geval ook konstant en $N^{\frac{1}{2}}$ x die $C^{(i)}$-terme asimptoties in waarskynlikheid gelyk aan nul is (sien byvoorbeeld lemma 3.10), is alleen die $B^{(i)}$-terme van belang by die bepaling van die asimptotiese variansies en kovariansies van die t_i 's.

Uit (3.45) volg m.b.v. (3.58) dat

$$(3.60)\quad B_{1N}^{(i)}+B_{2N}^{(i)} = \sum_g \nu_{Ng}[n_i^{-1} \sum_{j=1}^{n_i} \tilde{B}_g(x_{ij})-n_g^{-1} \sum_{j=1}^{n_g} \tilde{B}_i(x_{gj})]$$

$$= t_i'' \text{ , sê.}$$

4). Met behulp van (3.58) volg uit (3.33) dat

$$(3.61)\quad A^{(i)} = \int_I J(H)dF_i .$$

3.5. LIMIETVERDELING VAN T_k ONDER H.

Lemma 3.13. Die verwagtingswaarde van t''_i word gegee deur

(3.62) $E(t''_i) = 0.$

Bewys: Uit (3.60) volg

$$E(t''_i) = 0 \text{ omdat } E\tilde{B}_g(x_{ij}) = 0 \text{ vir alle } i,\ g \text{ en } j.$$

Lemma 3.14. Die asimptotiese waarde van die verwagtingswaarde van t_i word gegee deur

(3.63) $\lim\limits_{N\to\infty} E(t_i) = \int_I J(H)dF_i < \infty.$

Bewys: Met behulp van (3.58) volg uit (3.29) dat

$$\lim_{N\to\infty} E(t_i) = \int_I J(H)dF_i \quad \text{(sien §3.4 opmerking 1)}$$

$$= K < \infty \text{ soos in lemma 3.5.}$$

Lemma 3.15. Die variansie van t''_i word gegee deur

(3.64) $N\mathrm{var}(t''_i) = \nu_{Ni}^{-1} \sum\limits_g \sum\limits_{g'} \nu_{Ng}\nu_{Ng'}\sigma_{B_{gi}B_{g'i}} - 2\sum\limits_g \nu_{Ng}\sigma_{B_{gi}B_{ii}} + \sum\limits_g \nu_{Ng}\sigma^2_{B_{ig}}$

en die kovariansie tussen t''_i en $t''_{i'}$ vir $i' \neq i$ deur

(3.65) $N\mathrm{kov}(t''_i, t''_{i'}) = \sum\limits_g \nu_{Ng}\sigma_{B_{ig}B_{i'g}} - \sum\limits_g \nu_{Ng}\sigma_{B_{gi}B_{i'i}} - \sum\limits_g \nu_{Ng}\sigma_{B_{gi'}B_{ii'}}$

waar $\sigma_{B_{gi}B_{g'i}}$ gegee is in (3.22) en $\sigma^2_{B_{ig}} \equiv \sigma_{B_{ig}B_{ig}}$.

Bewys: $N\mathrm{kov}(t''_i, t''_{i'}) = NE(t''_i, t''_{i'})$ vir $i, i' = 1, 2, \ldots, k$

$$= N\,E\Big\{\Big[\sum_g \nu_{Ng} n_i^{-1} \sum_{j=1}^{n_i} \tilde{B}_g(x_{ij}) - \sum_g \nu_{Ng} n_g^{-1} \sum_{j=1}^{n_g} \tilde{B}_i(x_{gj})\Big] \cdot$$

$$\cdot \Big[\sum_{g'} \nu_{Ng'} n_{i'}^{-1} \sum_{j'=1}^{n_{i'}} \tilde{B}_{g'}(x_{i'j'}) - \sum_{g'} \nu_{Ng'} n_{g'}^{-1} \sum_{j'=1}^{n_{g'}} \tilde{B}_{i'}(x_{g'j'})\Big]\Big\}$$

$$= E\Big[\sum_g \nu_{Ng} n_i^{-1} \sum_{j=1}^{n_i} \tilde{B}_g(x_{ij}) \nu_{Ni'}^{-1} \sum_{g'} \nu_{Ng'} \sum_{j'=1}^{n_{i'}} \tilde{B}_{g'}(x_{i'j'})\Big] \qquad (= A' \text{ sê})$$

$$- E\Big[\sum_g \nu_{Ng} n_i^{-1} \sum_{j=1}^{n_i} \tilde{B}_g(x_{ij}) \sum_{g'} \sum_{j'=1}^{n_{g'}} \tilde{B}_{i'}(x_{g'j'})\Big] \qquad (= B' \text{ sê})$$

$$- E\Big[\sum_g \sum_{j=1}^{n_g} \tilde{B}_i(x_{gj}) n_{i'}^{-1} \sum_{g'} \nu_{Ng'} \sum_{j'=1}^{n_{i'}} \tilde{B}_{g'}(x_{i'j'})\Big] \qquad (= C' \text{ sê})$$

$$+ E\Big[\sum_g \nu_{Ng} n_g^{-1} \sum_{j=1}^{n_g} \tilde{B}_i(x_{gj}) \sum_{g'} \sum_{j'=1}^{n_{g'}} \tilde{B}_{i'}(x_{g'j'})\Big]. \qquad (= D' \text{ sê})$$

Neem die terme afsonderlik.

$$A' = E[\sum_g \sum_{g'} \nu_{Ng}\nu_{Ng'}\, n_i^{-1}\nu_{Ni'}^{-1} \sum_{j=1}^{n_i} \sum_{j'=1}^{n_{i'}} \widetilde{B}_g(x_{ij})\widetilde{B}_{g'}(x_{i'j'})]$$

$= 0$ indien $i' \neq i$ want dan is $\widetilde{B}_g(x_{ij})$ en $\widetilde{B}_{g'}(x_{i'j'})$ onafhanklik vanmekaar en $E\widetilde{B}_g(x_{ij}) = 0$.

Indien $i' = i$, is

$$A' = n_i^{-1}\nu_{Ni}^{-1} \sum_g \sum_{g'} \nu_{Ng}\nu_{Ng'}\, E[\sum_j \widetilde{B}_g(x_{ij})\widetilde{B}_{g'}(x_{ij}) + \sum_{j \neq j'}\sum \widetilde{B}_g(x_{ij})\widetilde{B}_{g'}(x_{ij'})]$$

$$= n_i^{-1}\nu_{Ni}^{-1} \sum_g \sum_{g'} \nu_{Ng}\nu_{Ng'}[\sum_j E\widetilde{B}_g(x_{ij})\widetilde{B}_{g'}(x_{ij}) + \sum_{j \neq j'}\sum E\widetilde{B}_g(x_{ij})\widetilde{B}_{g'}(x_{ij'})]$$

$= n_i^{-1}\nu_{Ni}^{-1} \sum_g \sum_{g'} \nu_{Ng}\nu_{Ng'}\, n_i \sigma_{B_{gi}B_{g'i}} + 0$ omdat x_{ij} en $x_{ij'}$ onafhanklik is en met behulp van (3.27)

$$= \nu_{Ni}^{-1} \sum_g \sum_{g'} \nu_{Ng}\nu_{Ng'}\, \sigma_{B_{gi}B_{g'i}} .$$

$B' = - \sum_g \sum_{g'} \nu_{Ng} n_i^{-1}\, E[\sum_{j=1}^{n_i} \sum_{j'=1}^{n_{g'}} \widetilde{B}_g(x_{ij})\widetilde{B}_{i'}(x_{g'j'})]$ wat nul is behalwe as $g' = i$ en $j' = j$, in welke geval

$$B' = - \sum_g \nu_{Ng} n_i^{-1} \sum_{j=1}^{n_i} E\widetilde{B}_g(x_{ij})\widetilde{B}_{i'}(x_{ij})$$

$$= - \sum_g \nu_{Ng}\sigma_{B_{gi}B_{i'i}} .$$

Netsoos B', is

$$C' = - \sum_g \nu_{Ng}\sigma_{B_{gi'}B_{ii'}} .$$

$D' = \sum_g \sum_{g'} \nu_{Ng} n_g^{-1}\, E[\sum_{j=1}^{n_g} \sum_{j'=1}^{n_{g'}} \widetilde{B}_i(x_{gj})\widetilde{B}_{i'}(x_{g'j'})]$ wat nul is behalwe as $g' = g$ en $j' = j$

$$= \sum_g \nu_{Ng} n_g^{-1} \sum_{j=1}^{n_g} E\widetilde{B}_i(x_{gj})\widetilde{B}_{i'}(x_{gj})$$

$$= \sum_g \nu_{Ng}\sigma_{B_{ig}B_{i'g}} .$$

Indien $i' = i$, volg dat

$$N\,\mathrm{var}(t_i'') = \nu_{Ni}^{-1} \sum_g \sum_{g'} \nu_{Ng}\nu_{Ng'}\, \sigma_{B_{gi}B_{g'i}} - 2\sum_g \nu_{Ng}\sigma_{B_{gi}B_{ii}} + \sum_g \nu_{Ng}\sigma^2_{B_{ig}} .$$

Indien $i' \neq i$, volg dat

$$N.\text{kov}(t_i'', t_{i'}'') = \sum_g \nu_{Ng}\sigma_{B_{ig}B_{i'g}} - \sum_g \nu_{Ng}\sigma_{B_{gi}B_{i'i}} - \sum_g \nu_{Ng}\sigma_{B_{gi'}B_{ii'}}$$

Hiermee is lemma 3.15 bewys.

Opmerkings.

1). Omdat $A^{(i)}$ konstant en die $C^{(i)}$-terme asimptoties in waarskynlikheid van kleiner orde as $N^{-\frac{1}{2}}$ is, volg:

(3.66) $\lim\limits_{N\to\infty} N\ \text{var}(t_i) = \lim\limits_{N\to\infty} N\ \text{var}(t_i'')$ en

(3.67) $\lim\limits_{N\to\infty} N\ \text{kov}(t_i, t_{i'}) = \lim\limits_{N\to\infty} N\ \text{kov}(t_i'', t_{i'}'')$, $i' \neq i$.

2). Substitusie van (3.58) in (3.30) lewer dieselfde uitdrukking vir $\lim\limits_{N\to\infty} N\ \text{var}(t_i)$ as dié in (3.64).

3). Met behulp van (3.58) volg uit (3.49) dat

(3.68) $0 < \lim\limits_{N\to\infty} N\ \text{var}(t_i) < \infty$. (Sien ook (3.23a) en lemma 3.7).

Lemma 3.16. Die verwagtingswaarde van T_k onder H word gegee deur $E(T_k) = (k-1)$.

Bewys: $E(T_k) = E\left\{\sum\limits_i (N-n_i)[t_i - E(t_i)]^2 [N\ \text{var}(t_i)]^{-1}\right\}$

$= \sum\limits_i (1-n_i/N)\ E[t_i - E(t_i)]^2/\text{var}(t_i)$

$= (k-1)$.

Lemma 3.17. Laat

(3.69) $e_i = Nc_i(N-n_i)^{-\frac{1}{2}}[\text{var}(t_i)]^{\frac{1}{2}}$ vir $i=1,2,\ldots,k$,

dan besit

(3.70) $\hat{\underline{S}} = \sum\limits_i e_i \hat{t}_i$

asimptoties 'n normaalverdeling as $N\to\infty$.

Bewys: $[S_N - E(S_N)]/\sigma_{S_N}$ is asimptoties normaal (0,1) verdeel as $N\to\infty$ (stelling 3.1). Gevolglik is $N^{\frac{1}{2}}(S_N - ES_N)$ asimptoties normaal verdeel as $N\to\infty$ omdat

(3.49) $0 < \lim\limits_{N\to\infty} N\sigma^2_{S_N} < \infty$.

Maar uit (3.9) en (3.55) volg

$$(3.71)\quad N^{\frac{1}{2}}(S_N-ES_N) = N^{\frac{1}{2}} \sum_i c_i[t_i-E(t_i)]$$

$$= N^{\frac{1}{2}} \sum_i (N-n_i)^{\frac{1}{2}} N^{-1}[\mathrm{var}(t_i)]^{-\frac{1}{2}} e_i[t_i-E(t_i)]$$

(sien lemma 3.15 opmerking 3)

$$= \sum_i e_i (N-n_i)^{\frac{1}{2}}[t_i-E(t_i)][N\ \mathrm{var}(t_i)]^{-\frac{1}{2}}$$

$$= \sum_i e_i \hat{t}_i \text{ uit (3.56)}$$

$$= \hat{S}, \text{ waarmee die lemma bewys is.}$$

Opmerking.

Deurdat die c_i 's willekeurige eindige getalle is, kan die e_i 's ook beskou word as willekeurige eindige getalle (uit (3.69) en lemma 3.15 opmerking 3).

STELLING 3.3.

Indien

(3.72) Enige twee frekwensiefunksies wat mekaar nie reeds oorvleuel nie, albei deur 'n gemeenskaplike derde frekwensiefunksie oorvleuel word (oorvleuelingsgebiede van maat, m.b.t. elk van die oorvleuelende verdelings, > 0), dan is die momentematriks M_3 van die variante $N^{\frac{1}{2}}t_i$ onder H asimptoties van rang (k-1) as $N \to \infty$.

Bewys: Die bewys gaan aan die hand van enkele hulpteoremas.

Opmerking.

Onder (3.58) impliseer voorwaarde (3.72) die geldigheid van (3.28).

Lemma 3.18. Stelling 3.3 geld onder H_o .

Bewys: Onder H_o is $\sigma_{B_{gi}B_{g'i'}}$ = konstant, sê $= \sigma^2$, vir alle i, i', g en g'.

Vergelyking (3.64) word dan:

$$N\ \mathrm{var}(t_i'') = \sigma^2(\nu_{Ni}^{-1} - 2 + 1) \text{ omdat } \sum_g \nu_{Ng} = 1$$

$$= \sigma^2(\nu_{Ni}^{-1} - 1)$$

$$= (1-\nu_{Ni})\nu_{Ni}^{-1}\ \sigma^2.$$

Vergelyking (3.65) word:

$$N \text{ kov}(t_i'', t_{i'}'') = \sigma^2(1-1-1) = -\sigma^2.$$

Dus:

$$\sum_i \nu_{Ni} \text{kov}(N^{\frac{1}{2}}t_i'', N^{\frac{1}{2}}t_{i'}'') = \nu_{Ni'} \text{var}(N^{\frac{1}{2}}t_{i'}'') + \sum_{i(\neq i')} \nu_{Ni} \text{kov}(N^{\frac{1}{2}}t_i'', N^{\frac{1}{2}}t_{i'}'')$$

$$= (1-\nu_{Ni'})\sigma^2 + \sum_{i(\neq i')} \nu_{Ni}(-\sigma^2)$$

$$= (1-\sum_i \nu_{Ni})\sigma^2$$

$$= 0.$$

M_3 is dus hoogstens van rang (k-1).

Vir alle $\nu_{Ni} \geq \nu_o > 0$ volg onmiddellik dat $\sum_{i(\neq r)} \nu_{Ni} \text{kov}(N^{\frac{1}{2}}t_i'', N^{\frac{1}{2}}t_{i'}'') \neq 0$ en verder bestaan daar geen getalle a_i sodanig dat

$$\sum_{i(\neq r)} a_i \text{kov}(N^{\frac{1}{2}}t_i'', N^{\frac{1}{2}}t_{i'}'') = 0 \text{ vir alle } i' \text{ nie.}$$

Hiermee is lemma 3.18 bewys.

Lemma 3.19. Onder die hipotese H_a, naamlik dat die populasies met verdelingsfunksies F_i nie almal identies is nie, is M_3 asimptoties hoogstens van rang (k-1).

Bewys: Die lemma word bewys deur aan te toon dat die determinant van M_3 asimptoties gelyk is aan nul. Nodig en voldoende hiervoor is dat daar asimptoties 'n lineêre verband tussen die k kolomme van die matriks bestaan, naamlik:

$$(3.73) \quad \lim_{N\to\infty} \sum_i n_i \text{kov}(t_i, t_{i'}) = 0 \text{ vir } i' = 1,2,\ldots,k.$$

Maar

$$\lim_{N\to\infty} n_i \text{kov}(t_i, t_{i'}) = \lim_{N\to\infty} \nu_{Ni} N \text{ kov}(t_i, t_{i'})$$

$$= \lim_{N\to\infty} \nu_{Ni} N \text{ kov}(t_i'', t_{i'}'') \quad \text{uit (3.67)}$$

$$= \lim_{N\to\infty} \nu_{Ni} \text{kov}(N^{\frac{1}{2}}t_i'', N^{\frac{1}{2}}t_{i'}'').$$

Nodig en voldoende vir die geldigheid van (3.73) is

$$(3.74) \quad \lim_{N\to\infty} \sum_i \nu_{Ni} \text{kov}(N^{\frac{1}{2}}t_i'', N^{\frac{1}{2}}t_{i'}'') = 0 \text{ vir } i' = 1,2,\ldots,k.$$

Met behulp van (3.64) en (3.65) kry ons:

$$\sum_i \nu_{Ni}\,\text{kov}(N^{\frac{1}{2}}t''_i, N^{\frac{1}{2}}t''_{i'}) = \nu_{Ni'}\text{var}(N^{\frac{1}{2}}t''_{i'}) + \sum_{i(\neq i')} \nu_{Ni}\,\text{kov}(N^{\frac{1}{2}}t''_i, N^{\frac{1}{2}}t''_{i'})$$

$$= \sum_g \sum_{g'} \nu_{Ng}\nu_{Ng'}\,\sigma_{B_{gi'}B_{g'i'}} - 2\nu_{Ni'} \sum_g \nu_{Ng}\sigma_{B_{gi'}B_{i'i'}} + \nu_{Ni'} \sum_g \nu_{Ng}\sigma^2_{B_{i'g}} +$$

$$+ \sum_{i(\neq i')} \sum_g \nu_{Ni}\nu_{Ng}\sigma_{B_{ig}B_{i'g}} - \sum_{i(\neq i')} \sum_g \nu_{Ni}\nu_{Ng}\sigma_{B_{gi}B_{i'i}} -$$

$$- \sum_{i(\neq i')} \sum_g \nu_{Ni}\nu_{Ng}\sigma_{B_{gi'}B_{ii'}}$$

$$= \sum_g \sum_{g'} \nu_{Ng}\nu_{Ng'}\,\sigma_{B_{gi'}B_{g'i'}} + \sum_i \sum_g \nu_{Ni}\nu_{Ng}\sigma_{B_{ig}B_{i'g}} -$$

$$- \sum_i \sum_g \nu_{Ni}\nu_{Ng}\sigma_{B_{gi}B_{i'i}} - \sum_i \sum_g \nu_{Ni}\nu_{Ng}\sigma_{B_{gi'}B_{ii'}}$$

$$= \sum_g \sum_{g'} \nu_{Ng}\nu_{Ng'}\,\sigma_{B_{gi'}B_{g'i'}} + \sum_g \sum_{g'} \nu_{Ng}\nu_{Ng'}\,\sigma_{B_{g'g}B_{i'g}} -$$

$$- \sum_g \sum_{g'} \nu_{Ng}\nu_{Ng'}\,\sigma_{B_{gg'}B_{i'g'}} - \sum_g \sum_{g'} \nu_{Ng'}\nu_{Ng}\sigma_{B_{gi'}B_{g'i'}}$$

$$= \sum_g \sum_{g'} \nu_{Ng}\nu_{Ng'}\,(\sigma_{B_{g'g}B_{i'g}} - \sigma_{B_{gg'}B_{i'g'}})$$

$$= 0.$$

Dit geld vir enige waarde van i' .

Aan voorwaarde (3.73) word dus voldoen. Daar bestaan dus asimptoties 'n lineêre verband tussen die k kolomme van die matriks M_3.

Hiermee is lemma 3.19 bewys.

Bewys van stelling 3.3.

In lemma 3.18 is bewys dat M_3 van rang (k-1) is onder H_o en in lemma 3.19 dat M_3 asimptoties hoogstens van rang (k-1) is onder H_a.

Ons bewys nou dat M_3 onder H_a ook asimptoties, vir $N \to \infty$, van rang (k-1) is.

Dui die momentematriks van die variante $N^{\frac{1}{2}}t''_i$ aan deur M_4. Indien die r^{de} ry en kolom van M_4 weggelaat word, ontstaan 'n matriks aangedui deur $M_4^{(r)}$. Ons moet aantoon dat daar geen lineêre verband tussen die kolomme van $M_4^{(r)}$ bestaan nie.

Beskou die $i'^{\,de}$ ry. In lemma 3.19 is bewys dat

(3.74a) $\sum_{i}\nu_{Ni}\mathrm{kov}(N^{\frac{1}{2}}t_i'',N^{\frac{1}{2}}t_{i'}'') = 0.$

Ons wil nou bewys dat

(3.75) $\sum_{i(\neq r)} a_i\mathrm{kov}(N^{\frac{1}{2}}t_i'',N^{\frac{1}{2}}t_{i'}'') \neq 0$

waar die a_i 's willekeurige koëffisiënte is (nie almal nul nie).

$$\sum_{i(\neq r)} a_i\mathrm{kov}(N^{\frac{1}{2}}t_i'',N^{\frac{1}{2}}t_{i'}'') = \sum_{i(\neq r)}(a_i-\nu_{Ni})\mathrm{kov}(N^{\frac{1}{2}}t_i'',N^{\frac{1}{2}}t_{i'}'') +$$
$$+\sum_{i}\nu_{Ni}\mathrm{kov}(N^{\frac{1}{2}}t_i'',N^{\frac{1}{2}}t_{i'}'')-\nu_{Nr}\mathrm{kov}(N^{\frac{1}{2}}t_{i'}'',N^{\frac{1}{2}}t_r'')$$
$$= \sum_{i(\neq r)}(a_i-\nu_{Ni})\mathrm{kov}(N^{\frac{1}{2}}t_i'',N^{\frac{1}{2}}t_{i'}'')-\nu_{Nr}\mathrm{kov}(N^{\frac{1}{2}}t_{i'}'',N^{\frac{1}{2}}t_r'')$$

met behulp van (3.74a).

Uit vergelyking (3.65) volg dat

$$\mathrm{kov}(N^{\frac{1}{2}}t_{i'}'',N^{\frac{1}{2}}t_r'') = \sum_{g}\nu_{Ng}\sigma_{B_{i'g}B_{rg}} - \sum_{g}\nu_{Ng}\sigma_{B_{i'r}B_{gr}} - \sum_{g}\nu_{Ng}\sigma_{B_{gi'}B_{ri'}}\,.$$

Hierin word terme soos $\sigma_{B_{i'r}B_{gr}}$, $\sigma_{B_{gi'}B_{ri'}}$ en $\sigma_{B_{i'g}B_{rg}}$ $(g=1,2,\ldots,k)$ bevat wat nie almal in $\sum_{i(\neq r)}(a_i-\nu_{Ni})\mathrm{kov}(N^{\frac{1}{2}}t_{i'}'',N^{\frac{1}{2}}t_i'')$ voorkom nie, waaronder byvoorbeeld $\sigma_{B_{i'r}B_{rr}}$ $(r\neq i')$ en $\sigma^2_{B_{ri'}}$. Aangesien die a_i 's nie funksies van die F_i 's is nie en omdat die indekse in B_{ig} en $B_{i'g}$ van $\sigma_{B_{ig}B_{i'g}}$ (sien vergelykings (3.16), (3.17) en (3.27)) nie verwissel kan word nie, behalwe onder H_0 in welke geval die rang van die matriks M_3 reeds as (k-1) bewys is (lemma 3.18), volg dat

$\sum_{i(\neq r)} a_i\mathrm{kov}(N^{\frac{1}{2}}t_{i'}'',N^{\frac{1}{2}}t_i'') \neq 0$ mits nie alle $\sigma_{B_{i'r}B_{gr}}$, $\sigma_{B_{gi'}B_{ri'}}$ en $\sigma_{B_{i'g}B_{rg}}$ vir vaste i' en r nul is nie. Voldoende hiervoor is dat $f_{i'}$ en f_r mekaar oorvleuel, of dat daar tenminste één f_g $(g \neq i',r)$ bestaan wat sowel $f_{i'}$ en f_r oorvleuel. Voorsiening hiervoor word deur voorwaarde (3.72) gemaak.

Daar bestaan dus in die algemeen geen lineêre verband tussen die (k-1) kolomme van $M_4^{(r)}$ nie.

Gevolglik is $|M_4^{(r)}| \neq 0$.

M_4 is dus van rang (k-1).

Omdat $\lim\limits_{N\to\infty} N \operatorname{kov}(t_i, t_{i'}) = \lim\limits_{N\to\infty} N \operatorname{kov}(t_i'', t_{i'}'')$

(sien (3.67)), is M_3 dus asimptoties, vir $N\to\infty$, van rang (k-1).

Hiermee is stelling 3.3 bewys.

Opmerking.

Indien aan voorwaarde (3.72) nie voldoen word nie, verminder die rang van M_3. Word $f_{i'}$ en f_r nie albei deur enige ander frekwensiefunksie f_g gesny nie, dan is alle $\sigma_{B_{i'g}B_{rg}}$, $\sigma_{B_{i'r}B_{gr}}$ en $\sigma_{B_{gi'}B_{ri'}}$ nul, sodat $\nu_{Nr}\operatorname{kov}(N^{\frac{1}{2}}t_{i'}'', N^{\frac{1}{2}}t_r'') = 0$. Kies ons $a_i = \nu_{Ni}$ vir alle i, dan is $\sum\limits_{i(\neq r)} a_i \operatorname{kov}(N^{\frac{1}{2}}t_{i'}'', N^{\frac{1}{2}}t_i'') = 0$ en is die rang van M_3 hoogstens (k-2).

STELLING 3.4.

Indien

(3.72) Enige twee frekwensiefunksies wat mekaar nie reeds oorvleuel nie, albei deur 'n gemeenskaplike derde frekwensiefunksie oorvleuel word (oorvleuelingsgebiede van maat, m.b.t. elk van die oorvleuelende verdelings, >0),

dan is die momentematriks M_5 van die variante $\hat{t}_i$ onder H asimptoties van rang (k-1) as $N\to\infty$.

Bewys: Die transformasie van $N^{\frac{1}{2}}t_i$ na $\hat{t}_i$ is lineêr en nie-singulier. Gevolglik is die rang van die matriks M_5 asimptoties dieselfde as dié van matriks M_3, naamlik (k-1) (stelling 3.3).

Lemma 3.20. Onder voorwaardes (3.10) tot (3.14') en (3.28) (of (3.72)) voldoen die variante $\hat{t}_i$ aan die voorwaarde

(3.76) $\lim\limits_{N\to\infty} E|\hat{t}_i|^{2+\delta} < \infty$ vir i=1,2,...,k en $\delta>0$.

Opmèrking.

Voorwaarde (3.14) word hier vervang deur

$$(3.14') \quad \begin{aligned} |J(H)| &\leq K[H(1-H)]^{-\frac{1}{4}+\delta} \\ |J^{(m)}(H)| &\leq K[H(1-H)]^{-m-\frac{1}{2}+\delta}, \quad m=1,2 \quad \text{en } \delta > 0. \end{aligned}$$

In alle verdere verwysings na voorwaarde (3.14) word implisiet bedoel (3.14').

Bewys: Sien bylaag B.

Lemma 3.21. Onder voorwaardes (3.10) tot (3.14) en (3.28) (of (3.72)) is die gesamentlike verdeling van $\hat{t}_1, \hat{t}_2, \ldots, \hat{t}_k$ asimptoties normaal as $N \to \infty$.

Bewys: Omdat

$$(3.77) \quad \text{var}(\hat{t}_i) = (N-n_i)N^{-1}\text{var}\left\{[t_i - E(t_i)][\text{var}(t_i)]^{-\frac{1}{2}}\right\} \quad \text{(sien (3.68))}$$

$$= 1 - n_i/N \text{ , is}$$

$$(3.78) \quad \lim_{N\to\infty} \text{var}(\hat{t}_i) > 0 \qquad \text{m.b.v. (3.4).}$$

Omdat ons verder ook bewys het dat

$$(3.76) \quad \lim_{N\to\infty} E|\hat{t}_i|^{2+\delta} < \infty \text{ vir } i=1,2,\ldots,k \text{ en } \delta>0$$

(lemma 3.20) en dat, vir willekeurige eindige konstantes e_i,

$$(3.70) \quad \hat{S} = \sum_i e_i \hat{t}_i$$

asimptoties normaal verdeel is as $N \to \infty$ (lemma 3.17), volg uit lemma 2.5 dat die gesamentlike verdeling van $\hat{t}_1, \hat{t}_2, \ldots, \hat{t}_k$ asimptoties 'n normaalverdeling is as $N \to \infty$.

STELLING 3.5.

Onder die voorwaardes (3.10) - (3.14) en (3.72) besit

$$(3.79) \quad T = \sum_i \sum_{i'} \sum_g c_{ig} c_{i'g} \gamma_g \hat{t}_i \hat{t}_{i'}$$

waar die c_{ig} 's en γ_g 's deur M_5 bepaal word (sien lemma 1.1), onder H asimptoties 'n χ^2-verdeling met (k-1) grade van vryheid.

Bewys: Die variante $\hat{t}_1, \hat{t}_2, \ldots, \hat{t}_k$ met $E(\hat{t}_i) = 0$ vir alle i, momentematriks Λ ($\equiv M_5$) wat asimptoties van rang (k-1) is (stelling 3.4) en karakteristieke wortels $\varkappa_1, \varkappa_2, \ldots, \varkappa_{k-1}$ en 0 besit asimptoties gesamentlik 'n normaalverdeling (lemma 3.21) met karakteristieke funksie

$e^{-\frac{1}{2}\vec{t}'\Lambda\vec{t}}$ waar $\vec{t}' = (t_1, t_2, \ldots, t_k)$ en $\vec{t}$ die ooreenkomstige kolomvektor.

Uit lemma 1.1 volg dat T asimptoties 'n χ^2-verdeling met (k-1) g.v.v. besit as $N \to \infty$.

Opmerkings.

1). Vir elke spesifieke alternatiewe hipotese is die momentematriks Λ bepaald en kan die karakteristieke wortels bepaal word - sien lemma 1.1 opmerking 3.

2). Die toetsingsgrootheid T is baie algemeen, maar die bepaling van die variansies en kovariansies en uit Λ die karakteristieke wortels $\varkappa_i$ is nie so maklik nie, indien dit wel moontlik is. Aangesien ons slegs in asimptotiese relatiewe doeltreffendheids-alternatiewe belangstel (d.i. alternatiewe hipoteses H_a waarby

(3.80) $H_a \xrightarrow[\infty]{N} H_o$),

is dit voldoende om 'n toetsingsgrootheid soortgelyk aan dié van hoofstuk II te bestudeer (sien (3.57)).

3). 'n Alternatiewe prosedure kan ook gevolg word, naamlik: Enige (k-1) $\hat{t}_i$'s besit asimptoties gesamentlik die nie-singuliere normaalverdeling (vergelyk lemma 3.17 met $c_k = 0 = e_k$ en stelling 3.4). Dui die momentematriks van $\hat{t}_1, \hat{t}_2, \ldots, \hat{t}_{k-1}$ aan deur $M_{(k)}$ en laat $\vec{t}'_{(k)}$ die ryvektor $(\hat{t}_1, \hat{t}_2, \ldots, \hat{t}_{k-1})$ voorstel met $\vec{t}_{(k)}$ die ooreenkomstige kolomvektor.

Dan besit $\vec{t}'_{(k)} M^{-1}_{(k)} \vec{t}_{(k)}$ asimptoties, vir $N \to \infty$, 'n χ^2-verdeling met (k-1) g.v.v. waar $M^{-1}_{(k)}$ die inverse matriks is van $M_{(k)}$ (sien stelling 1.1 opmerking 2).

STELLING 3.6.

Onder die voorwaardes (3.10) - (3.14) en onder hipoteses H_a van die vorm

(3.81) H_a: $F_i(x) = F(\delta_{iN}x + \varepsilon_{iN})$

waar die δ 's en ε 's willekeurige eindige reële getalle is sodanig dat $\delta_{iN} \xrightarrow[\infty]{N} 1$ en $\varepsilon_{iN} \xrightarrow[\infty]{N} 0$, $i=1,2,\ldots,k$, besit

(3.82) $T_k = \sum_i (N-n_i)[t_i - E(t_i|H_a)]^2 [N \text{ var}(t_i|H_a)]^{-1}$

asimptoties, vir $N\to\infty$, 'n χ^2-verdeling met (k-1) g.v.v.

Bewys: Die momentematriks van die $\hat{t}_i$'s word soos volg verkry indien $N\to\infty$:

Uit (3.64) en (3.66) volg

$$\lim_{N\to\infty} N\text{var}(t_i|H_a) = \lim_{N\to\infty} [\nu_{Ni}^{-1} \sum_g \sum_{g'} \nu_{Ng}\nu_{Ng'}\sigma_{B_{gi}B_{g'i}} - 2\sum_g \nu_{Ng}\sigma_{B_{gi}B_{ii}} + \sum_g \nu_{Ng}\sigma^2_{B_{ig}}]$$

waar

$$(3.22)\quad \sigma_{B_{gi}B_{g'i}} = \iint_{-\infty<x<y<\infty} F_i(x)[1-F_i(y)]J'[H(x)]J'[H(y)]\cdot [dF_g(x)dF_{g'}(y) + dF_{g'}(x)dF_g(y)]$$

$$= \iint F(\delta_{iN}x+\varepsilon_{iN})[1-F(\delta_{iN}y+\varepsilon_{iN})]J'[\sum_{i'}\nu_{Ni'}F(\delta_{i'N}x+\varepsilon_{i'N})]\cdot J'[\sum_{i'}\nu_{Ni'}F(\delta_{i'N}y+\varepsilon_{i'N})]\{dF(\delta_{gN}x+\varepsilon_{gN})dF(\delta_{g'N}y+\varepsilon_{g'N}) + dF(\delta_{g'N}x+\varepsilon_{g'N})dF(\delta_{gN}y+\varepsilon_{gN})\}$$

en omdat J' en F oral kontinu is, volg(sien GIBSON(1954) p.441):

$$(3.83)\quad \lim_{N\to\infty} \sigma_{B_{gi}B_{g'i}} = 2\iint_{-\infty<x<y<\infty} F(x)[1-F(y)]J'[F(x)]J'[F(y)]\cdot dF(x)dF(y)$$

vir alle i, g en g'

$= \sigma^2$ (sê) ($<\infty$ uit lemma 3.7. Sien ook lemma 3.18).

Gevolglik is

$$(3.84)\quad \lim_{N\to\infty} N \text{ var}(t_i|H_a) = [\nu_i^{-1}\sum_g \sum_{g'} \nu_g\nu_{g'} - 2\sum_g \nu_g + \sum_g \nu_g]\sigma^2$$

$$= [\nu_i^{-1} - 1]\sigma^2$$

$$= (1-\nu_i)\nu_i^{-1}\,\sigma^2.$$

Soortgelyk volg:

$$(3.85)\quad \lim_{N\to\infty} N\text{kov}(t_i, t_{i'}) = \lim_{N\to\infty} [\sum_g \nu_{Ng}\sigma_{B_{ig}B_{i'g}} - \sum_g \nu_{Ng}\sigma_{B_{gi}B_{i'i}} - \sum_g \nu_{Ng}\sigma_{B_{gi'}\ B_{ii'}}]$$

$$= [\sum_g \nu_g - \sum_g \nu_g - \sum_g \nu_g]\ \sigma^2$$

$$= -\ \sigma^2.$$

Nou is m.b.v. (3.77)

$$(3.86)\quad \lim_{N\to\infty} \text{var}(\hat{t}_i) = 1-\nu_i \quad \text{en}$$

$$(3.87)\quad \lim_{N\to\infty} \text{kov}(\hat{t}_i, \hat{t}_{i'}) = \lim_{N\to\infty} \frac{[(1-\nu_{Ni})(1-\nu_{Ni'})]^{\frac{1}{2}} N\text{kov}(t_i, t_{i'})}{[N\ \text{var}(t_i)\cdot N\ \text{var}(t_{i'})]^{\frac{1}{2}}}$$

$$= \frac{[(1-\nu_i)(1-\nu_{i'})]^{\frac{1}{2}}(-\sigma^2)}{[(1-\nu_i)\nu_i^{-1}\sigma^2(1-\nu_{i'})\nu_{i'}^{-1}\sigma^2]^{\frac{1}{2}}}$$

$$= -\sqrt{\nu_i\nu_{i'}}\ .$$

In die limiet, as $N\to\infty$, is die momentematriks van die $\hat{t}_i$'s dus

$$\Lambda = \begin{Bmatrix} 1-\nu_1 & ,-\sqrt{\nu_1\nu_2}, & \ldots, & -\sqrt{\nu_1\nu_k} \\ -\sqrt{\nu_1\nu_2}, & 1-\nu_2 & ,\ldots, & -\sqrt{\nu_2\nu_k} \\ \cdot & \cdot & \cdot & \cdot \\ -\sqrt{\nu_1\nu_k}, & -\sqrt{\nu_2\nu_k}, & \ldots, & 1-\nu_k \end{Bmatrix}$$

$$= I - \vec{\nu}\vec{\nu}'$$

waar I die eenheidsmatriks is en $\vec{\nu}' = (\sqrt{\nu_1}, \sqrt{\nu_2}, \ldots, \sqrt{\nu_k}\ )$.

Omdat $\sum_i \nu_i = 1$ en alle $\nu_i > 0$, volg uit die bewys van lemma 1.2 dat alle karakteristieke wortels van Λ gelyk is aan 1 behalwe één wortel wat 0 is.

Die toetsingsgrootheid T van stelling 3.5 herlei dus m.b.v. (1.13) na:

$$(3.88)\quad T = \sum_{i=1}^{k} \hat{t}_i^2$$

$$= \sum_i (N-n_i)[t_i - E(t_i)]^2 [N\ \text{var}(t_i)]^{-1} \quad \text{m.b.v. } (3.56)$$

$$= T_k.$$

Hiermee is die stelling bewys.

Opmerkings.

1). Voorwaarde (3.81) impliseer die geldigheid van (3.72) vir N voldoende groot.

2). M. L. PURI het in dieselfde rigting gewerk en skynbaar soortgelyke resultate gevind - sien Ann. Math. Stat. (1961), abstract 6, p. 1350. Sover dit aan ons bekend is, is die volledige artikel nog nie gepubliseer nie.*

Lemma 3.22. Vir alternatiewe hipoteses van die vorm

(3.81) H_a: $F_i(x) = F(\delta_{iN}x + \varepsilon_{iN})$ waar $\delta_{iN} \xrightarrow[\infty]{N} 1$ en $\varepsilon_{iN} \xrightarrow[\infty]{N} 0$,

$i=1,2,\ldots,k$, volg dat

$$(3.89)\quad \lim_{N\to\infty} \frac{N\ \mathrm{var}(t_i\,|H_a)}{N\ \mathrm{var}(t_i\,|H_o)} = 1.$$

Bewys: Soos in stelling 3.6 volg

$$(3.84)\quad \lim_{N\to\infty} N\ \mathrm{var}(t_i\,|H_a) = (\nu_i^{-1} - 1)\sigma^2$$

waar

$$(3.83)\quad \sigma^2 = 2 \iint_{-\infty<x<y<\infty} F(x)[1-F(y)]J'[F(x)]J'[F(y)]dF(x)dF(y).$$

Uit

$$(3.22)\quad \sigma_{B_{gi}B_{g'i}} = \iint_{-\infty<x<y<\infty} F_i(x)[1-F_i(y)]J'[H(x)]J'[H(y)]\cdot$$
$$\cdot[dF_g(x)dF_{g'}(y) + dF_{g'}(x)dF_g(y)]$$

volg onder H_o dat

$$\sigma_{B_{gi}B_{g'i}} = \sigma^2. \quad \text{Dus}$$

$$(3.90)\quad \lim_{N\to\infty} N\ \mathrm{var}(t_i\,|H_o) = [\nu_i^{-1} \sum_g \sum_{g'} \nu_g\nu_{g'} - 2\sum_g \nu_g + \sum_g \nu_g]\sigma^2$$
$$= (\nu_i^{-1} - 1)\sigma^2$$

sodat m.b.v. (3.84) volg

$$(3.89)\quad \lim_{N\to\infty} \frac{N\ \mathrm{var}(t_i\,|H_a)}{N\ \mathrm{var}(t_i\,|H_o)} = 1.$$

---

* Hierdie artikel het verskyn nadat die proefskrif reeds getik was, nl. in Ann. Math. Stat. <u>35</u>, Maart 1964, pp. 102-121 Puri beskou slegs verskuiwingsalternatiewe by die a.r.d.-bepaling.

3.6. ASIMPTOTIESE RELATIEWE DOELTREFFENDHEID.

3.6.1. Inleiding.

In hierdie paragraaf word die asimptotiese relatiewe doeltreffendheid (hierna genoem a.r.d.) van enkele spesiale gevalle van T_k met betrekking tot die F-toets of Bartlett se toets vir die homogeniteit van variansies bereken.

Beskou

(3.55) $t_i = \int_{-\infty}^{\infty} J_N[H_N(x)]dF_{in_i}(x)$ en

(3.91) $t_i' = \sum_{j=1}^{n_i} \Psi_N(\delta_{ij})$ soos in (2.1) gedefinieer, waar $\delta_{ij} = r_{ij}/(N+1)$ en r_{ij} die rangnommer is van x_{ij} in die gesamentlike rangskikking van die N waarnemings x_{ij}; $j=1,2,\ldots,n_i$; $i=1,2,\ldots,k$.

Lemma 3.23. Indien

(3.92) $\Psi_N[N/(N+1)] = o(N^{\frac{1}{2}})$ en

(3.93) $\int_{I_N} \left|J_N[H_N(x)]-\Psi_N\left[\frac{NH_N(x)}{N+1}\right]\right| dF_{in_i}(x)=o_p(N^{-\frac{1}{2}})$, $i=1,2,..,k$,

dan is, onder 'n algemene hipotese H,

(3.94) $E(t_i'/n_i|H) = \int_{-\infty}^{\infty} J[H(x)]dF_i(x) + o(N^{-\frac{1}{2}})$.

Bewys: Vir $\varepsilon>0$ geld uit die stelling van Tchebycheff:

$$P[|N^{\frac{1}{2}}(t_i-t_i'/n_i)|\geq\varepsilon] \leq E[|N^{\frac{1}{2}}(t_i-t_i'/n_i)|]/\varepsilon$$

$$= \tfrac{1}{\varepsilon}E[N^{\frac{1}{2}}|\int_{-\infty}^{\infty}\left\{J_N[H_N(x)]-\Psi_N[NH_N(x)/(N+1)]\right\}dF_{in_i}(x)|]$$

$$\leq \tfrac{1}{\varepsilon}E[N^{\frac{1}{2}}\int_{-\infty}^{\infty}|J_N[H_N(x)]-\Psi_N[NH_N(x)/(N+1)]|dF_{in_i}(x)]$$

$$= \tfrac{1}{\varepsilon}E[N^{\frac{1}{2}}\int_{I_N}|J_N[H_N(x)]-\Psi_N[NH_N(x)/(N+1)]|dF_{in_i}(x)$$

$$+N^{\frac{1}{2}}\int_{H_N=1}|J_N[H_N(x)]-\Psi_N[NH_N(x)/(N+1)]|dF_{in_i}(x)]$$

$= \tfrac{1}{\varepsilon}E[N^{\frac{1}{2}}o_p(N^{-\frac{1}{2}})+N^{\frac{1}{2}}o(N^{-\frac{1}{2}})]$ m.b.v. (3.13), (3.92) en (3.93)

$= o(1)$.

Opmerkings:

1). Voorwaarde (3.143) kan soos volg omskryf word:

(3.149) $|f(x+cxN^{-\frac{1}{2}})J'[\lambda_0 F(x+cxN^{-\frac{1}{2}})+(1-\lambda_0)F(x+c''xN^{-\frac{1}{2}})]|<M(x)$

vir N voldoende groot met $\int_{-\infty}^{\infty} x\, M(x)\, dF(x) < \infty$ waar $\lambda_0>0$ en die c 's willekeurige eindige reële getalle is.

2). 'n Voldoende voorwaarde vir die geldigheid van (3.143), is

(3.150) $|f(x+cxN^{-\frac{1}{2}})J'[F(x+c'xN^{-\frac{1}{2}})]| \leq M < \infty$ gelykmatig in x vir N voldoende groot waarby $\int_{-\infty}^{\infty} x\, dF(x) < \infty$ met c en c' willekeurige eindige reële getalle.

STELLING 3.11.

Onder die voorwaardes (3.10) - (3.14), (3.92), (3.93) en (3.143) besit T_k' (sien (3.110)) onder H_a asimptoties, vir $N\to\infty$, 'n nie-sentrale χ^2-verdeling met (k-1) g.v.v. en nie-sentraliteitsparameter

(3.151) $$\lambda_k^2 = [\int_{-\infty}^{\infty} xf(x)J'[F(x)]dF(x)]^2 \sum_i \nu_i(d.-d_i)^2 \cdot [\lim_{N\to\infty} N^{-1}\sum_h(\Psi_{Nh}-\bar{\Psi}_N)^2]^{-1}$$

Bewys: Soos in stelling 3.8 volg dat T_k' asimptoties, vir $N\to\infty$, 'n nie-sentrale χ^2-verdeling besit met (k-1) g.v.v. en nie-sentraliteitsparameter

(3.152) $$\lambda_k^2 = \lim_{N\to\infty} \sum_i \frac{(N-n_i)N[E(t_i'/n_i|H_a)-E(t_i'/n_i|H_o)]^2}{N^2 var(t_i'/n_i|H_o)}$$

uit (3.123)

$$= [\int_{-\infty}^{\infty} xf(x)J'[F(x)]dF(x)]^2 \lim_{N\to\infty} \sum_i \frac{(N-n_i)(d.-d_i)^2 n_i(N-1)}{N(N-n_i)\sum_h(\Psi_{Nh}-\bar{\Psi}_N)^2}$$

uit (2.20), (2.21) en (3.144).

$$= [\int_{-\infty}^{\infty} xf(x)J'[F(x)]dF(x)]^2 \sum_i \nu_i(d.-d_i)^2 / \lim_{N\to\infty} N^{-1}\sum_h(\Psi_{Nh}-\bar{\Psi}_N)^2.$$

Opmerking.

Soos in opmerking 2 na stelling 3.8, kan λ_k^2 onder voorwaarde (3.125) ook nog in 'n ander vorm geskryf word, naamlik

$$(3.153)\quad \lambda_k^2 = \frac{\left[\int_{-\infty}^{\infty} xf(x)J'[F(x)]dF(x)\right]^2 \sum_i \nu_i(d.-d_i)^2}{\lim_{N\to\infty} \int_{-\infty}^{\infty}\left\{J_N[H_N(x)] - \int_{-\infty}^{\infty} J_N[H_N(x)]dH_N(x)\right\}^2 dH_N(x)}$$

STELLING 3.12.

Die asimptotiese relatiewe doeltreffendheid van T_k' met betrekking tot Bartlett se toets B vir die homogeniteit van variansies, is

$$(3.154)\quad E_{T_k',B} = \left[\int_{-\infty}^{\infty} xf(x)J'[F(x)]dF(x)\right]^2(\beta_2-1)\cdot$$
$$\cdot\left[4 \lim_{N\to\infty} N^{-1} \sum_h(\Psi_{Nh}-\bar{\Psi}_N)^2\right]^{-1}$$

waar $\beta_2 = \mu_4/\mu_2^2$ en μ_r die r^{de} orde sentrale moment is van die verdeling met verdelingsfunksie $F(x)$.

Bewys: Bartlett se wysiging van die Neyman-Pearson L_1 - toets word gegee deur

$$(3.155)\quad B = N \log_e(N^{-1}\sum_i n_i s_i^2) - \sum_i n_i \log_e s_i^2$$

waar s_i^2 die variansie is van die i^{de} steekproef.

Volgens CROUSE(1960) §4.11 besit B onder H_a asimptoties, vir $N\to\infty$, 'n nie-sentrale χ^2-verdeling met (k-1) g.v.v. en nie-sentraliteitsparameter

$$(3.156)\quad \lambda_B^2 = 4(\beta_2-1)^{-1} \lim_{N\to\infty} \sum_i \nu_{Ni}(d.-d_i)^2$$
$$= 4(\beta_2-1)^{-1} \sum_i \nu_i(d.-d_i)^2.$$

Die a.r.d. van T_k' m.b.t. B is dan (sien (3.131)):

$$(3.157)\quad E_{T_k',B} = \lambda_k^2/\lambda_B^2$$
$$= (\beta_2-1)\left[\int_{-\infty}^{\infty} xf(x)J'[F(x)]dF(x)\right]^2\cdot$$
$$\cdot\left[4 \lim_{N\to\infty} N^{-1} \sum_h(\Psi_{Nh}-\bar{\Psi}_N)^2\right]^{-1}.$$

Definieer χ_α^2 deur $P[\chi^2 \geq \chi_\alpha^2 | H_o] = \alpha$.

Dan is T_k' asimptoties onderskeidend met betrekking tot die klas van hipoteses H waarvoor

(3.186) $\lim_{N\to\infty} P[T_k' \geq \chi_\alpha^2 | H] = 1$ vir elke alternatiewe hipotese H in die klas – sien §1.3 en CROUSE(1960).

3.7.2. 'n Belangrike Stelling.

STELLING 3.13.

Onder voorwaardes (3.10)-(3.14), (3.92) en (3.93) is T_k' asimptoties onderskeidend met betrekking tot alle H waarvoor minstens een $\gamma_i = \infty$ solank net (3.72) geld.

Bewys: Neem eerstens

(3.187) $\lim_{N\to\infty} (E_{ia}-E_{io})/\sigma_{io} = +\infty$ vir enige bepaalde i uit 1,2,...,k.

Dan is

(3.188) $P[T_k' < \chi_\alpha^2] = P[\sum_{i'} \hat{t}_{i'}'^2 < \chi_\alpha^2]$

$\leq P[\hat{t}_i'^2 < \chi_\alpha]$

$= P[|\hat{t}_i'| < \chi_\alpha]$

$= P[|(N-n_i)^{\frac{1}{2}}(t_i'/n_i - E_{io})/(N^{\frac{1}{2}}\sigma_{io})| < \chi_\alpha]$

$= P[|t_i'/n_i - E_{io}| < \chi_\alpha \sigma_{io}(1-\nu_{Ni})^{-\frac{1}{2}}]$

$\leq P[\ t_i'/n_i - E_{io} \ < \chi_\alpha \sigma_{io}(1-\nu_{Ni})^{-\frac{1}{2}}]$

$= P[(t_i'/n_i - E_{ia})/\sigma_{ia} < \sigma_{io}\sigma_{ia}^{-1}\{\chi_\alpha(1-\nu_{Ni})^{-\frac{1}{2}} - (E_{ia}-E_{io})/\sigma_{io}\}]$.

Omdat $\lim_{N\to\infty} \chi_\alpha(1-\nu_{Ni})^{-\frac{1}{2}}$ = konstant $< \infty$ (want $\lim_{N\to\infty} \nu_{Ni} < 1$ uit (3.4)), volg uit (3.187) dat

$\chi_\alpha(1-\nu_{Ni})^{-\frac{1}{2}} - (E_{ia}-E_{io})/\sigma_{io}$ negatief is vir voldoende groot N, sê vir $N > N_o$.

Derhalwe geld vir $N > N_o$:

(3.189) $P[(t_i'/n_i - E_{ia})/\sigma_{ia} < \sigma_{io}\sigma_{ia}^{-1}\{\chi_\alpha(1-\nu_{Ni})^{-\frac{1}{2}} - (E_{ia}-E_{io})/\sigma_{io}\}]$

$\leq P[|(t_i'/n_i - E_{ia})/\sigma_{ia}| \geq \sigma_{io}\sigma_{ia}^{-1}|\chi_\alpha(1-\nu_{Ni})^{-\frac{1}{2}} - (E_{ia}-E_{io})/\sigma_{io}|]$

$= P[(t_i'/n_i - E_{ia})^2/\sigma_{ia}^2 \geq \sigma_{io}^2\sigma_{ia}^{-2}\{\chi_\alpha(1-\nu_{Ni})^{-\frac{1}{2}} - (E_{ia}-E_{io})/\sigma_{io}\}^2]$

Dui die rangnommer van z_h aan deur r_h $(z_1 \leq z_2 \leq \ldots \leq z_N)$.

Omdat $r_h/(N+1)$ nie strengstygend in h is nie en die verdelingsfunksie $F(x)$ wel monotoon strengstygend is in x, is 'n aanpassing in die definisie van $J_N[F(x)]$ nodig sodat aan voorwaarde (3.93) voldoen word. Stel naamlik

$$(3.158) \quad J_N[F(x)] = \begin{cases} NF(x)/(N+1) & \text{vir } F(x) < \frac{1}{2} \\ 1-NF(x)/(N+1) & \text{vir } F(x) > \frac{1}{2} . \end{cases}$$

Dan is

$$J[F(x)] = \begin{cases} F(x) & \text{vir } F(x) < \frac{1}{2} \\ 1-F(x) & \text{vir } F(x) > \frac{1}{2} . \end{cases}$$

Nou is

$$J'[F(x)] = \begin{cases} 1 & \text{vir } x < x_0 \\ -1 & \text{vir } x > x_0 \end{cases}$$

waarby $F(x_0) = \frac{1}{2}$ (sien §3.3, net na vergelyking (3.17)). D.w.s. x_0 is die onbekende gemeenskaplike mediaan van die verdelings F_i .

Maar in die punt $x = x_0$ bestaan $J'[F(x)]$ nie. Aan voorwaardes (3.10) - (3.14) word voldoen (sien ANSARI en BRADLEY(1960) p. 1183), behalwe aan voorwaarde (3.14) in die punt $x = x_0$. Die bewys slaag egter nog omdat bogenoemde 'n punt van maat nul is. Voorwaardes (3.92) en (3.93) word bevredig. Verder is

$$f(x)\,|J'[F(x)]| = f(x) \leq K < \infty,$$ sodat (3.143) bevredig word (sien stelling 3.10 opmerking 1).

$$(3.159) \quad \int_{-\infty}^{\infty} xf(x)J'[F(x)]dF(x) = -\int_{-\infty}^{x_0} xf(x)dF(x) + \int_{x_0}^{\infty} xf(x)dF(x).$$

Uit (2.80) en (2.90) volg m.b.v. (2.20) dat

$$(3.160) \quad \lim_{N\to\infty} N^{-1} \sum_h (\Psi_{Nh} - \bar{\Psi}_N)^2 = \frac{1}{48} \quad .$$

Substitusie van (3.159) en (3.160) in (3.152) lewer

$$(3.161) \quad \lambda_\delta^2 = 48\left[\int_{x_0}^{\infty} xf(x)dF(x) - \int_{-\infty}^{x_0} xf(x)dF(x)\right]^2 \sum_i \nu_i (d.-d_i)^2$$

sodat met behulp van (3.156) en (3.157) volg

$$(3.162)\quad E_{T_\delta,B} = 12(\beta_2-1)\left[\int_{x_0}^{\infty} xf^2(x)dx - \int_{-\infty}^{x_0} xf^2(x)dx\right]^2.$$

Opmerkings.

1). Bostaande uitdrukking is dieselfde as dié verkry deur ANSARI en BRADLEY(1960) vir die twee-steekproef geval en is ook dieselfde as die a.r.d. van SUKHATME(1957) se T-toets m.b.t. die F-toets (in die twee-steekproef geval). Sien ook KLOTZ(1962) vergelyking (3.3).

2). Vir die normaal (0,1)-verdeling is

$$E_{T_\delta,B} = 6/\pi^2 = 0.61$$

Vir die homogene verdeling is $E_{T_\delta,B} = 0.60$.

Vir die Laplace-verdeling met $f(x) = \frac{1}{2}e^{-|x|}$ is

$E_{T_\delta,B} = 0.94$ (sien ANSARI en BRADLEY(1960)).

3.6.3.3. $\Psi_N(\delta_{ij}) = (\delta_{ij}-\frac{1}{2})^2$ (gewone toekenning van rangnommers).

Nou is

$$(2.101)\quad t_i' = \sum_j(\delta_{ij}-\tfrac{1}{2})^2$$ met toetsingsgrootheid T_M gegee in (2.109).

Stel $J_N[F(x)] = \Psi_N[NF(x)/(N+1)]$ vir $0<F(x)<1$

$$= [NF(x)/(N+1) - \tfrac{1}{2}]^2.$$

Dan is $J[F(x)] = [F(x) - \frac{1}{2}]^2$ met

$$J'[F(x)] = 2[F(x) - \tfrac{1}{2}]$$
$$= 2F(x) - 1 \quad \text{en}$$

$$f(x)J'[F(x)] = f(x)[2F(x) - 1] \leq K < \infty$$ sodat

(3.143) bevredig word. Dit volg maklik dat voorwaardes (3.10) - (3.14), (3.92) en (3.93) bevredig word.

Met behulp van (2.20) volg uit (2.103) dat

$$\lim_{N\to\infty} N^{-1} \sum_h(\Psi_{Nh}-\bar{\Psi}_N)^2 = \frac{1}{180} .$$

Uit (3.152) volg nou:

(3.163) $\lambda_M^2 = 180[\int_{-\infty}^{\infty} xf(x)[2F(x)-1]dF(x)]^2 \sum_i \nu_i(d.-d_i)^2$

sodat m.b.v. (3.156) en (3.157) volg:

(3.164) $E_{T_M,B} = 45(\beta_2-1)[2\int_{-\infty}^{\infty} xf^2(x)F(x)dx - \int_{-\infty}^{\infty} xf^2(x)dx]^2.$

Opmerkings.

1). Hierdie uitdrukking is dieselfde as die a.r.d. van MOOD(1954) se toets m.b.t. die F-toets (sien SUKHATME(1957)). Dit is in ooreenstemming met wat ons verwag, want T_M is 'n uitbreiding van MOOD(1954) se toetsingsgrootheid van twee na k steekproewe (sien §2.5.3.2). Sien ook KLOTZ(1962) vergelyking (3.2).

2). Vir die normaal (0,1)-verdeling is

$$E_{T_M,B} = \frac{15}{2\pi^2} = 0.76$$ (sien MOOD(1954) en DWASS(1956)).

Vir die homogene verdeling is $E_{T_M,B} = 1$.

Vir die verdeling met frekwensiefunksie

$$f(x) = \tfrac{1}{2}(1-\alpha)a^{\alpha-1}|x|^{-\alpha} \qquad -a\leq x\leq a,\ \alpha<1$$

is $E_{T_M,B} = 5(1-\alpha)(5-\alpha)^{-1}$ wat na 0 neig as $\alpha\to 1$ (SUKHATME(1957)).

Vir die verdeling met frekwensiefunksie

$$f(x) = \frac{(1+x^2)^{-m}}{B(\frac{1}{2},m-\frac{1}{2})}, \quad -\infty \leq x \leq \infty$$

geld $E_{T_M,B} \to\infty$ as $m\to 2.5$ (CROUSE(1960)).

Vir die Laplace-verdeling met $f(x) = \frac{1}{2}e^{-|x|}$ is $E_{T_M,B} = 1.08$ (ANSARI en BRADLEY(1960)).

3.6.3.4. $\Psi_N(\delta_{ij}) = (\delta_{ij}-\frac{1}{2})^2$ (rangnommers van weerskante af toegeken).

Hierdie geval stem presies ooreen met die vorige (sien §2.5.3.3 opmerking 2) en word nie verder bespreek nie.

3.6.3.5. $\Psi_N(\delta_{ij}) = \delta_{ij}^2$ (gewone toekenning van rangnommers).

Hier is

(2.121) $t_i' = \sum_j r_{ij}^2/(N+1)^2$ met toetsingsgrootheid T_R gegee in (2.129).

Stel $J_N[F(x)] = [NF(x)/(N+1)]^2$, dan is

$J[F(x)] = F^2(x)$ met

$J'[F(x)] = 2F(x)$ en word alle voorwaardes, soos in §3.6.3.3, bevredig.

Voorwaarde (3.143) word bevredig omdat

$f(x)J'[F(x)] = 2f(x)F(x) \leq K < \infty$.

Uit (2.127) volg: $\lim\limits_{N\to\infty} N^{-1}\sum\limits_h (\Psi_{Nh}-\bar{\Psi}_N)^2 = \frac{4}{45}$ sodat

(3.165) $\lambda_R^2 = 45[\int\limits_{-\infty}^{\infty} xf(x)F(x)dF(x)]^2 \sum\limits_i \nu_i(d.-d_i)^2$ uit (3.152).

Uit (3.156) en (3.157) volg:

(3.166) $E_{T_R,B} = \frac{45}{4}(\beta_2-1)[\int\limits_{-\infty}^{\infty} xf^2(x)F(x)dx]^2$.

Opmerkings.

1). Vir die normaal (0,1)-verdeling is

$$E_{T_R,B} = \frac{15}{32\pi^2} = 0.047$$

Vir die homogene verdeling is

$$E_{T_R,B} = \frac{1}{16} = 0.0625 \quad .$$

2). Bostaande toets word eintlik aanbeveel as toets teen verskuiwingsalternatiewe. Die a.r.d. van T_R m.b.t. die F-toets is uit (3.131):

$$E_{T_R,F} = 45[\int\limits_{-\infty}^{\infty} f^2(x)F(x)dx]^2\,\sigma_F^2 .$$

Vir die normaal (0,1)-verdeling is

$$E_{T_R,F} = \frac{45}{16\pi} = 0.895 \quad .$$

Vir die homogene verdeling is

$$E_{T_R,F} = \frac{15}{16} = 0.9375 \quad .$$

3.6.3.6. $\Psi_N(\delta_{ij}) = \delta_{ij}^2$ (rangnommers van weerskante af toegeken).

Hier is

(2.130) $t_i' = \sum_j r_{ij}^2/(N+1)^2$ met toetsingsgroothede T_G en T_H gegee in (2.138) en (2.145) namate N ewe of onewe is.

Laat $J_N[F(x)] = \begin{cases} [NF(x)/(N+1)]^2 & \text{vir } F(x) < \frac{1}{2} \\ [1-NF(x)/(N+1)]^2 & \text{vir } F(x) > \frac{1}{2} \end{cases}$.

Dan is $J[F(x)] = \begin{cases} F^2(x) & \text{vir } F(x) < \frac{1}{2} \\ [1-F(x)]^2 & \text{vir } F(x) > \frac{1}{2} \end{cases}$.

Stel weer $F(x_o) = \frac{1}{2}$.

Gevolglik is

$$J'[F(x)] = \begin{cases} 2F(x) & \text{vir } x < x_o \\ -2[1-F(x)] & \text{vir } x > x_o \end{cases}.$$

Alle voorwaardes word ook in hierdie geval bevredig.

Uit (2.136) en (2.143) volg

(3.167) $\lim_{N\to\infty} N^{-1} \sum_h (\Psi_{Nh}-\bar{\Psi}_N)^2 = \frac{1}{180}$

(3.168) $\int_{-\infty}^{\infty} xf(x)J'[F(x)]dF(x)$

$$= \int_{-\infty}^{x_o} xf(x)2F(x)dF(x) - \int_{x_o}^{\infty} xf(x)2[1-F(x)]dF(x)$$

$$= 2[\int_{-\infty}^{x_o} xf(x)F(x)dF(x) - \int_{x_o}^{\infty} xf(x)dF(x) + $$
$$+ \int_{x_o}^{\infty} xf(x)F(x)dF(x)]$$

$$= 2[\int_{-\infty}^{\infty} xf(x)F(x)dF(x) - \int_{x_o}^{\infty} xf(x)dF(x)].$$

Met behulp van (3.151) volg nou vir sowel T_G as T_H:

(3.169) $\lambda_{GH}^2 = 720[\int_{-\infty}^{\infty} xf^2(x)F(x)dx - \int_{x_o}^{\infty} xf^2(x)dx]^2 \sum_i \nu_i(d.-d_i)^2$.

Ten slotte volg uit (3.156) en (3.157):

(3.170) $E_{T_{GH},B} = 180(\beta_2-1)[\int_{-\infty}^{\infty} xf^2(x)F(x)dx - \int_{x_o}^{\infty} xf^2(x)dx]^2$.

Opmerking.

Vir die normaal (0,1)-verdeling is

$$E_{T_{GH},B} = \frac{15}{2\pi^2}(\sqrt{3} - 1)^2 = 0.41 .$$

Vir die homogene verdeling is $E_{T_{GH},B} = 0.25$.

3.6.3.7. $\Psi_N(\delta_{ij}) = Q^2(\delta_{ij})$ – sien §2.5.3.6.

In dié geval is

(2.146) $t_i' = \sum_j Q^2(\delta_{ij})$ met toetsingsgrootheid T_{Q_2} gegee in (2.149).

Stel $J_N[F(x)] = [Q\{NF(x)/(N+1)\}]^2$.

Dan is $J[F(x)] = [Q\{F(x)\}]^2$

$$= [\Phi^{-1}\{F(x)\}]^2 \quad \text{en}$$

$$(3.171) \quad J'[F(x)] = \frac{d[\Phi^{-1}\{F(x)\}]^2}{dF(x)}$$

$$= \frac{d[\Phi^{-1}\{F(x)\}]^2}{d\Phi^{-1}[F(x)]} \cdot \frac{d\Phi^{-1}[F(x)]}{dF(x)}$$

$$= \frac{2\Phi^{-1}[F(x)]}{\phi[\Phi^{-1}\{F(x)\}]} \quad \text{uit (3.135), sodat}$$

$$f(x+cxN^{-\frac{1}{2}})J'[F(x+c'xN^{-\frac{1}{2}})] = f(x+cxN^{-\frac{1}{2}})\frac{2\Phi^{-1}[F(x+c'xN^{-\frac{1}{2}})]}{\phi[\Phi^{-1}\{F(x+c'xN^{-\frac{1}{2}})\}]}.$$

Vir byvoorbeeld die normaal (0,1)-verdeling is

$$f(x+cxN^{-\frac{1}{2}})J'[F(x+c'xN^{-\frac{1}{2}})] = \frac{2\phi(x+cxN^{-\frac{1}{2}})\ (x+c'xN^{-\frac{1}{2}})}{\phi(x+c'xN^{-\frac{1}{2}})} = O(x).$$

Omdat $\int_{-\infty}^{\infty} x^2 d\Phi(x) < \infty$, volg dat aan voorwaarde (3.149) voldoen word.

Alle ander voorwaardes word bevredig – sien CROUSE(1960) [14].

Met behulp van (3.171) is

$$(3.172) \quad \int_{-\infty}^{\infty} xf(x)J'[F(x)]dF(x) = 2\int_{-\infty}^{\infty} \frac{xf(x)\Phi^{-1}[F(x)]}{\phi[\Phi^{-1}\{F(x)\}]} dF(x).$$

Uit STOKER(1955) p. 57 volg

$$\lim_{N\to\infty} N^{-1} \sum_h Q_h^r = \int_{-\infty}^{\infty} \xi^r \, d\Phi(\xi).$$

Dus $\lim_{N\to\infty} N^{-1} \sum_h Q_h^2 = 1$ – sien §3.6.2.4 – en

$$\lim_{N\to\infty} N^{-1} \sum_h Q_h^4 = 3.$$

---

14). Sien H. S. SCHOEMAN (nog onvoltooide proefskrif).

In hierdie geval is

$$(3.173)\quad \lim_{N\to\infty} N^{-1}\sum_h(\Psi_{Nh}-\bar{\Psi}_N)^2 = \lim_{N\to\infty} N^{-1}\sum_h \Psi^2_{Nh} - \lim_{N\to\infty}\bar{\Psi}^2_N$$
$$= \lim_{N\to\infty} N^{-1}\sum_h Q_h^4 - \lim_{N\to\infty}(N^{-1}\sum_h Q_h^2)^2$$
$$= 3 - 1$$
$$= 2\ .$$

Met behulp van (3.172) en (3.173) volg uit (3.151):

$$(3.174)\quad \lambda^2_{Q_2} = 4\Big[\int_{-\infty}^{\infty}\frac{xf(x)\Phi^{-1}[F(x)]}{\phi[\Phi^{-1}\{F(x)\}]}\,dF(x)\Big]^2.\ \tfrac{1}{2}\sum_i \nu_i(d.-d_i)^2$$

sodat ten slotte

$$(3.175)\quad E_{T_{Q_2},B} = \tfrac{1}{2}(\beta_2-1)\Big[\int_{-\infty}^{\infty}\frac{xf^2(x)\Phi^{-1}[F(x)]}{\phi[\Phi^{-1}\{F(x)\}]}\,dx\ \Big]^2\ .$$

(Sien byvoorbeeld KLOTZ(1962) vergelyking (3.1)).

Opmerkings.

1). KLOTZ(1962) beweer dat $0 \leq E_{T_{Q_2},B} \leq \infty$. Hy kon egter slegs bewys dat $0.47 \leq E_{T_{Q_2},B} \leq \infty$, waarby $E_{T_{Q_2},B} = 0.47$ vir $F(x) = \Phi(x^\alpha)$ met $\alpha\to\infty$.

2). Vir die normaal (0,1)-verdeling is

$$E_{T_{Q_2},B} = \tfrac{1}{2}(3-1)\Big[\int_{-\infty}^{\infty}\frac{x\phi^2(x)x}{\phi(x)}\,dx\ \Big]^2$$
$$= \Big[\int_{-\infty}^{\infty}x^2\phi(x)\,dx\ \Big]^2$$
$$= \mu_2^2$$

$= 1$, wat ooreenstem met wat KLOTZ(1962) beweer vir die twee-steekproef geval.

3). Vir 'n vergelyking tussen die toetse van §e 3.6.3.2, 3.6.3.3 en 3.6.3.7 (in die twee-steekproef geval) vir ses spesifieke verdelings, sien KLOTZ(1962).

3.6.3.8. $\Psi_N(\delta_{ij}) = E(\xi^2_{r_{ij}})$ – sien §2.5.3.7, CROUSE(1960) §4.14 en CAPON(1961).

Hier is

$$(2.150)\quad t_i' = \sum_j E(\xi^2_{r_{ij}})$$ met toetsingsgrootheid T_{T_2} gegee in (2.153).

Met $J_N[F(x)] = E(\xi^2_{NF(x)})$ word voorwaarde (3.93) bevredig. Ook die ander nodige voorwaardes word in die algemeen bevredig - sien byvoorbeeld CAPON(1961).

Uit STOKER(1955) p. 55 volg

$$\lim_{N\to\infty} N^{-1} \sum_h \Xi_h^r = \int_{-\infty}^{\infty} \xi^r dF(\xi) \quad \text{mits} \quad \int_{-\infty}^{\infty} |\xi|^r dF(\xi) < \infty.$$

Vir $F(x) = \Phi(z)$, die normaal (0,1)-verdelingsfunksie, volg op soortgelyke wyse as in §3.6.3.7 dat

$$\lim_{N\to\infty} N^{-1} \sum_h (\Psi_{Nh} - \bar{\Psi}_N)^2 = 2.$$

Met behulp hiervan en vergelyking (3.157) volg:

$$(3.176) \quad E_{T_{T_2},B} = \tfrac{1}{8}(\beta_2 - 1)[\int_{-\infty}^{\infty} x f^2(x) J'[F(x)]dx]^2$$

$$= \tfrac{1}{2}(\beta_2 - 1)[\int_{-\infty}^{\infty} xz \frac{\phi(z)}{(\frac{dx}{dz})} dz]^2 \text{ vir } F(x) = \Phi(z) \quad -$$

sien CROUSE(1960) §4.14 .

Hierdie toets is bespreek deur CROUSE(1960) en CAPON(1961) in die twee-steekproef geval. KLOTZ(1962) beweer (sonder bewys) dat hierdie toets asimptoties ekwivalent is aan dié van §3.6.3.7 vir k = 2.

CROUSE(1960) het vir die twee-steekproef geval bewys dat die a.r.d. van T_{T_2} m.b.t. die F-toets gegee word deur

$$(3.177) \quad E_{T_{T_2},F} = \tfrac{1}{8}(\beta_2 - 1)[\int_{-\infty}^{\infty} x f^2(x) J'[F(x)]dx]^2$$

wat ekwivalent is aan $E_{T_{Q_2},B}$ in (3.175) en ook dieselfde as die ooreenkomstige uitdrukking (3.176) in die geval van k steekproewe. (Die aantal steekproewe k speel geen rol in die a.r.d.-uitdrukkings nie).

Opmerkings.

1). Vir die normaal $(0,\sigma^2)$-verdeling is $E_{T_{T_2},B} = 1$ en vir die homogene verdeling is $E_{T_{T_2},B} = \infty$ (CROUSE(1960)).

2). CROUSE(1960) se vermoede dat $E_{T_{T_2},B} \geq 1$ word weerspreek deur KLOTZ(1962) wat bewys het dat $E_{T_{T_2},B} = 0.47$ vir die verdeling met verdelingsfunksie

$F(x) = \Phi(x^{\alpha})$, $\alpha \to \infty$ (sien §3.6.3.7 opmerking 1).

3.7. ASIMPTOTIESE ONDERSKEIDENDHEID.

3.7.1. Inleiding.

Uit (3.110) volg

(3.178) $T_k' = \sum_i (N-n_i)[t_i'/n_i - E(t_i'/n_i \mid H_o)]^2 [N\mathrm{var}(t_i'/n_i \mid H_o)]^{-1}$.

Stel

(3.179) $E_{io} = E(t_i'/n_i \mid H_o)$,

(3.180) $E_{ia} = E(t_i'/n_i \mid H_a)$,

(3.181) $\sigma_{io}^2 = \mathrm{var}(t_i'/n_i \mid H_o)$ en

(3.182) $\sigma_{ia}^2 = \mathrm{var}(t_i'/n_i \mid H_a)$.

Uit stellings 3.8 en 3.11 en vergelyking (3.178) volg dat

(3.183) $T_k' = \sum_i (N-n_i)[t_i'/n_i - E_{io}]^2 [N\sigma_{io}^2]^{-1}$

onder voorwaardes (3.10)-(3.14), (3.92), (3.93) en

óf

(3.97) H_a: $F_i(x) = F(x+d_iN^{-\frac{1}{2}})$, $i=1,2,\ldots,k$ en (3.98)

óf

(3.142) H_a: $F_i(x) = F(x+d_ixN^{-\frac{1}{2}})$ met $\int_{-\infty}^{\infty} x\, dF_i(x) = 0$,

$i=1,2,\ldots,k$ en (3.143),

asimptoties, vir $N \to \infty$, 'n nie-sentrale χ^2-verdeling besit met $(k-1)$ g.v.v. en nie-sentraliteitsparameter

(3.184) $\lambda_k^2 = \lim_{N\to\infty} \sum_i (N-n_i)[E_{ia}-E_{io}]^2[N\sigma_{io}^2]^{-1}$

$= \sum_i (1-\nu_i)\gamma_i^2$ waar

(3.185) $\gamma_i \overset{\mathrm{def}}{=} \lim_{N\to\infty} |(E_{ia}-E_{io})/\sigma_{io}|$.

Definieer χ_α^2 deur $P[\chi^2 \geq \chi_\alpha^2 | H_o] = \alpha$.

Dan is T_k' asimptoties onderskeidend met betrekking tot die klas van hipoteses H waarvoor

(3.186) $\lim\limits_{N\to\infty} P[T_k' \geq \chi_\alpha^2 | H] = 1$ vir elke alternatiewe hipotese H in die klas - sien §1.3 en CROUSE(1960).

3.7.2. 'n Belangrike Stelling.

STELLING 3.13.

Onder voorwaardes (3.10)-(3.14), (3.92) en (3.93) is T_k' asimptoties onderskeidend met betrekking tot alle H waarvoor minstens een $\gamma_i = \infty$ solank net (3.72) geld.

Bewys: Neem eerstens

(3.187) $\lim\limits_{N\to\infty} (E_{ia}-E_{io})/\sigma_{io} = +\infty$ vir enige bepaalde i uit 1,2,...,k.

Dan is

(3.188) $P[T_k' < \chi_\alpha^2] = P[\sum\limits_{i'} \hat{t}_{i'}'^2 < \chi_\alpha^2]$

$\leq P[\hat{t}_i'^2 < \chi_\alpha]$

$= P[|\hat{t}_i'| < \chi_\alpha]$

$= P[|(N-n_i)^{\frac{1}{2}}(t_i'/n_i - E_{io})/(N^{\frac{1}{2}}\sigma_{io})| < \chi_\alpha]$

$= P[|t_i'/n_i - E_{io}| < \chi_\alpha \sigma_{io}(1-\nu_{Ni})^{-\frac{1}{2}}]$

$\leq P[t_i'/n_i - E_{io} < \chi_\alpha \sigma_{io}(1-\nu_{Ni})^{-\frac{1}{2}}]$

$= P[(t_i'/n_i - E_{ia})/\sigma_{ia} < \sigma_{io}\sigma_{ia}^{-1}\{\chi_\alpha(1-\nu_{Ni})^{-\frac{1}{2}} - (E_{ia}-E_{io})/\sigma_{io}\}]$.

Omdat $\lim\limits_{N\to\infty} \chi_\alpha(1-\nu_{Ni})^{-\frac{1}{2}} =$ konstant $< \infty$ (want $\lim\limits_{N\to\infty} \nu_{Ni} < 1$ uit (3.4)), volg uit (3.187) dat

$\chi_\alpha(1-\nu_{Ni})^{-\frac{1}{2}} - (E_{ia}-E_{io})/\sigma_{io}$ negatief is vir voldoende groot N, sê vir $N > N_o$.

Derhalwe geld vir $N > N_o$:

(3.189) $P[(t_i'/n_i - E_{ia})/\sigma_{ia} < \sigma_{io}\sigma_{ia}^{-1}\{\chi_\alpha(1-\nu_{Ni})^{-\frac{1}{2}} - (E_{ia}-E_{io})/\sigma_{io}\}]$

$\leq P[|(t_i'/n_i - E_{ia})/\sigma_{ia}| \geq \sigma_{io}\sigma_{ia}^{-1}|\chi_\alpha(1-\nu_{Ni})^{-\frac{1}{2}} - (E_{ia}-E_{io})/\sigma_{io}|]$

$= P[(t_i'/n_i - E_{ia})^2/\sigma_{ia}^2 \geq \sigma_{io}^2\sigma_{ia}^{-2}\{\chi_\alpha(1-\nu_{Ni})^{-\frac{1}{2}} - (E_{ia}-E_{io})/\sigma_{io}\}^2]$

$$\leq \frac{1}{\sigma_{io}^2\sigma_{ia}^{-2}[\chi_\alpha(1-\nu_{Ni})^{-\frac{1}{2}} - (E_{ia}-E_{io})/\sigma_{io}]^2}$$ met behulp van

Tchebycheff se ongelykheid.

Onder voorwaarde (3.72) volg m.b.v. (3.68) en lemma 3.23 dat

(3.190) $0 < \lim\limits_{N\to\infty} N\sigma_{ia}^2 < \infty$ en

(3.191) $0 < \lim\limits_{N\to\infty} N\sigma_{io}^2 < \infty$, sodat

(3.192) $0 < \lim\limits_{N\to\infty} \sigma_{io}^2/\sigma_{ia}^2 < \infty$.

Nou volg m.b.v. (3.188) en (3.189):

$\lim\limits_{N\to\infty} P[T_k' < \chi_\alpha^2] = 0$, d.w.s.

(3.193) $\lim\limits_{N\to\infty} P[T_k' \geq \chi_\alpha^2] = 1$.

T_k' is dus asimptoties onderskeidend indien (3.187) geld.

Beskou vervolgens

(3.194) $\lim\limits_{N\to\infty} (E_{ia}-E_{io})/\sigma_{io} = -\infty$ vir enige bepaalde i uit 1,2,...,k.

(3.195) $P[T_k' \geq \chi_\alpha^2] = P[\sum\limits_{i'} \hat{t}_{i'}'^2 \geq \chi_\alpha^2]$

$$\geq P[\hat{t}_i'^2 \geq \chi_\alpha^2]$$

$$= P[|\hat{t}_i'| \geq \chi_\alpha]$$

$$= P[|t_i'/n_i - E_{io}| \geq \chi_\alpha\sigma_{io}(1-\nu_{Ni})^{-\frac{1}{2}}]$$

$$= P[|(t_i'/n_i - E_{io})/\sigma_{ia}| \geq \sigma_{io}\sigma_{ia}^{-1}\chi_\alpha(1-\nu_{Ni})^{-\frac{1}{2}}] .$$

Omdat $\lim\limits_{N\to\infty} \chi_\alpha(1-\nu_{Ni})^{-\frac{1}{2}}$ = konstant $< \infty$, geld vir alle N voldoende groot:

(3.196) $P[|(t_i'/n_i - E_{io})/\sigma_{ia}| \geq \sigma_{io}\sigma_{ia}^{-1}\chi_\alpha(1-\nu_{Ni})^{-\frac{1}{2}}] \geq$

$$\geq P[(t_i'/n_i - E_{io})/\sigma_{ia} < -\sigma_{io}\sigma_{ia}^{-1}\chi_\alpha(1-\nu_{Ni})^{-\frac{1}{2}}]$$

$$= 1 - P[(t_i'/n_i - E_{ia})/\sigma_{ia} \geq -\sigma_{io}\sigma_{ia}^{-1}\{\chi_\alpha(1-\nu_{Ni})^{-\frac{1}{2}} + (E_{ia}-E_{io})/\sigma_{io}\}]$$

$$\geq 1 - [-\sigma_{io}\sigma_{ia}^{-1}\{\chi_\alpha(1-\nu_{Ni})^{-\frac{1}{2}} + (E_{ia}-E_{io})/\sigma_{io}\}]^{-2}$$ vir N voldoende groot m.b.v. Tchebycheff se ongelykheid, want

$- \sigma_{io}\sigma_{ia}^{-1}[\chi_\alpha(1-\nu_{Ni})^{-\frac{1}{2}} + (E_{ia}-E_{io})/\sigma_{io}] > 0$ vir N voldoende groot.

Dus

$$(3.197)\quad \lim_{N\to\infty} P[T_k' \geq \chi_\alpha^2] = 1$$

waarmee bewys is dat T_k' asimptoties onderskeidend is as (3.194) geld.

Hiermee is die stelling bewys.

Opmerkings.

1). Neem alternatiewe hipoteses van die vorm

$$(3.97)\quad H_a\colon F_i(x) = F(x+d_iN^{-\frac{1}{2}}),\quad i=1,2,\ldots,k\ ,$$

dan geld onder voorwaarde (3.98) dat

$$(3.198)\quad \gamma_i = \lim_{N\to\infty} |(E_{ia}-E_{io})/\sigma_{io}|$$

$$= \nu_i^{\frac{1}{2}}(1-\nu_i)^{-\frac{1}{2}}|(d.-d_i)\int_{-\infty}^{\infty} f(x)J'[F(x)]dF(x)\lim_{N\to\infty}[N^{-1}\sum_h(\Psi_{Nh}-\bar{\Psi}_N)^2]^{-\frac{1}{2}}|$$

uit (3.117)

= konstant $< \infty$ m.b.v. (2.41) omdat $F(x)$ kontinu veronderstel word, sodat

$$(3.199)\quad 0 \leq \gamma_i < \infty \text{ vir } i=1,2,\ldots,k.$$

Vir alle γ_i eindig besit T_k' onder voorwaardes (3.97), (3.98) en die voorwaardes van hierdie stelling asimptoties 'n nie-sentrale χ^2-verdeling met eindige nie-sentraliteitsparameter

$$(3.184)\quad \lambda_k^2 = \sum_i(1-\nu_i)\gamma_i^2 \qquad \text{(stelling 3.8).}$$

Gevolglik is T_k' nie asimptoties onderskeidend indien alle γ_i eindig is nie (sien ook KENDALL en STUART(1961) vol. II p. 231).

2). Vir alternatiewe hipoteses van die vorm

$$(3.142)\quad H_a\colon F_i(x) = F(x+d_ixN^{-\frac{1}{2}}) \text{ met } \int_{-\infty}^{\infty} x\, dF_i(x) = 0,$$

$i=1,2,..,k$ kan onder voorwaarde (3.143) en die voorwaardes van hierdie stelling analoë resultate bewys word as in opmerking 1 hierbo.

3). Uit (3.94) volg

$\lim\limits_{N\to\infty} |E_{ia}-E_{io}| \neq 0$ as

$$\int_{-\infty}^{\infty} J[H(x)]dF_i(x) \neq \int_{-\infty}^{\infty} J[F(x)]dF(x) \ . \tag{3.200}$$

Maar uit (3.191) is $\sigma_{io} = O(N^{-\frac{1}{2}})$.

Dus $\gamma_i = \lim\limits_{N\to\infty} |(E_{ia}-E_{io})/\sigma_{io}| = \infty$ onder (3.200).

T_k' is dus asimptoties onderskeidend met betrekking tot die klas van alternatiewe waarvoor (3.200) geld onderhewig aan (3.72).

HOOFSTUK IV.

DIE PROBLEEM VAN m-RANGSKIKKINGS.

4.1. ALGEMENE INLEIDING.

Die probleem van m-rangskikkings is voorgestel en bespreek deur FRIEDMAN(1937) en is breedvoerig behandel deur KENDALL(1948).

KENDALL(1948) beskou m „waarnemers". Elke waarnemer rangskik k gegewe „items" volgens 'n bepaalde eienskap en ken dan rangnommers toe waarby r_{ij} $(i=1,2,\ldots,m;\ j=1,2,\ldots,k)$ die rangnommer is van item x_{ij}. Omdat elke waarnemer opnuut, onafhanklik van die ander waarnemers, rangnommers toeken, het ons:

$$\sum_{j=1}^{k} r_{ij} = \tfrac{1}{2}k(k+1) \text{ vir alle } i=1,2,\ldots,m.$$

Skematies word bogenoemde soos volg voorgestel:

TABEL 4.1.

Skematiese voorstelling van m-rangskikkings.

		Items			
		1	2	. . .	k
Waarnemers:	1	r_{11}	r_{12}	. . .	r_{1k}
	2	r_{21}	r_{22}	. . .	r_{2k}
	.	. . .	. . .	. . .	. . .
	.				
	m	r_{m1}	r_{m2}	. . .	r_{mk}
Totaal		u_1^*	u_2^*		u_k^*

Die kolomtotale $u_j^* = \sum_{i=1}^{m} r_{ij}$; $j=1,2,\ldots,k$, word met mekaar vergelyk en gee 'n aanduiding van die ooreenstemming wat daar tussen die verskillende waarnemers bestaan. Indien alle rangskikkings identies is (volkome ooreenstemming tussen waarnemers), sal die totale u_j^* die waardes m, 2m,...,km aanneem (nie noodwendig in dié volgorde nie). Indien daar egter min verskil tussen die items bestaan, sal die waarnemers beswaarlik tussen hulle kan onderskei en mag die totale u_j^* betreklik min van mekaar verskil. ('n Item wat deur een waarnemer vroeg in die rangskikking geplaas word, kan deur 'n ander waarnemer baie later geplaas word. Sien byvoorbeeld KENDALL(1948) pp. 80-81).

Die nulhipotese H_o wat ondersoek word, veronderstel dat elke rangskikking, onafhanklik van die ander, ewekansig gekies is uit die versameling van alle moontlike permutasies van die getalle 1,2,...,k.

FRIEDMAN(1937) het as toetsingsgrootheid voorgestel

(4.1) $\chi^2 = 12S^*/[mk(k+1)]$ waar

(4.2) $S^* = \sum_{j=1}^{k} [u_j^* - \frac{1}{2}m(k+1)]^2$

en het bewys dat χ^2 onder H_o asimptoties 'n χ^2-verdeling met (k-1) g.v.v. besit as $m\to\infty$.

In die geval waar gelyke waarnemings voorkom, het BENARD en VAN ELTEREN(1953) 'n wysiging aan χ^2 aangebring, naamlik

(4.3) $\chi^2 = 12S^*/[mk(k+1) - T^*]$ met T^* gedefinieer in (4.73). (Sien §4.4.3.2 vir die voorwaardes waaronder χ^2 se asimptotiese verdeling afgelei is).

Bogenoemde probleem is deur BENARD en VAN ELTEREN(1953) uitgebrei na die geval waar die aantal waarnemings x_{ijh} van die item x_{ij} (d.w.s. die aantal waarnemings in die „sel" (i,j)), enige nie-negatiewe heelgetal k_{ij} is. Die i^{de} rangskikking bevat dan

(4.4) $k_i = \sum_{j=1}^{k} k_{ij}$ [1] waarnemings wat weer volgens

---

1). Tensy anders vermeld, deurloop i,i' in hierdie hoofstuk die waardes 1,2,...,m; j en j' die waardes 1,2,...,k; p_i die waardes $1,2,\ldots,k_i$; h die waardes $1,2,\ldots,k_{ij}$ en ζ,ζ' die waardes 1,2,...,(k-1).

'n bepaalde eienskap gerangskik word. Daar is dus altesame

(4.5) $N = \sum_i k_i = \sum_i \sum_j k_{ij}$ waarnemings waarby alle k_i en k_{ij} eindig is, tensy anders gespesifiseer.

BENARD en VAN ELTEREN(1953) [2)] definieer

(4.6) $$\tilde{u}_{ij} = \begin{cases} \sum_h [r_{ijh} - \frac{1}{2}(k_i+1)] & \text{indien } k_{ij}>0 \\ 0 & \text{indien } k_{ij}=0 \end{cases}$$

waarby r_{ijh} die rangnommer is van die waarneming x_{ijh} in die i^{de} rangskikking. (Rangnommers word opnuut in elke ry toegeken, onafhanklik van die ander rye). B+v.E(1953) neem dan 'n funksie van die kolomtotale $\tilde{u}_j$ as toetsingsgrootheid, waar $\tilde{u}_j = \sum_i \tilde{u}_{ij}$ (sien §4.4.3.1).

In hierdie hoofstuk word 'n veralgemeende toetsingsgrootheid vir die probleem van m-rangskikkings onder H_0 voorgestel en 'n uitbreiding na 'n meer algemene geval behandel.

4.2. 'n MEER ALGEMENE BENADERING.

4.2.1. Inleiding.

Laat $\Psi_{k_i}(\delta)$, vir alle waardes van k_i, 'n willekeurige funksie wees wat gedefinieer en eindig is vir iedere δ op die ope interval (0,1) by alle waardes van k_i en monotoon stygend (of monotoon dalend) in sy argument δ by vaste k_i. (Sien ook hoofstuk II §2.2.1).

Beskou 'n funksie van die rangnommers, naamlik $\Psi_{k_i}[r_{ijh}/(k_i+1)] = \Psi_{k_i}(\delta_{ijh})$ waar

(4.7) $\delta_{ijh} = r_{ijh}/(k_i+1)$.

Let op dat $1 \leq r_{ijh} \leq k_i$ vir alle $j=1,2,\ldots,k$ en $h=1,2,\ldots,k_{ij}$.

---

2). Voortaan word na hierdie artikel verwys as B+v.E(1953).

Opmerking.

Voortaan skryf ons kortweg $\Psi_i(\delta)$ in plaas van $\Psi_{k_i}(\delta)$, behalwe in gevalle waar $\Psi_{k_i}(\delta)$ se afhanklikheid van k_i beklemtoon moet word.

4.2.2. Gelyke Waarnemings (Knope).

In die geval van gelyke waarnemings binne 'n bepaalde rangskikking (ry), word dieselfde prosedure gevolg as in hoofstuk II. Ten opsigte van notasie is die volgende aanpassing nodig (alleen die geval van gemiddelde range word bespreek):

Dui, vir die i^{de} ry, die gerangskikte stel x_{ijh}-waardes aan deur $z_{i1}, z_{i2}, \ldots, z_{ik_i}$ waarby $z_{i1} \leq z_{i2} \leq \ldots \leq z_{ik_i}$. Gestel daar kom 'n knoop van lengte g_{α_i} voor. Aan elkeen van die gelyke waarnemings z_{i,ν_i+1} , z_{i,ν_i+2} , ... , $z_{i,\nu_i+g_{\alpha_i}}$ word dieselfde rang toegeken, naamlik:

$$(4.8) \quad g_{\alpha_i}^{-1} \sum_{\mu_i} (\nu_i + \mu_i) = \nu_i + \tfrac{1}{2}(g_{\alpha_i} + 1) \quad ^{3)}.$$

Op ooreenkomstige wyse word die funksie $\Psi_i(\delta_{ip_i})$, waar

(4.9) $\delta_{ip_i} = r_{ip_i}/(k_i+1)$ en r_{ip_i} die rangnommer is van z_{ip_i} in die i^{de} rangskikking, gedefinieer deur

$$(4.10) \quad \Psi_i(\delta_{ip_i}) = g_{\alpha_i}^{-1} \sum_{\mu_i} \Psi_i[(\nu_i+\mu_i)/(k_i+1)]$$

vir alle p_i waarvoor $z_{ip_i} = z_{i,\nu_i+1}$.

Gestel nou daar kom onder die z_{ip_i} 's λ_i knope van lengtes $g_{i1}, g_{i2}, \ldots, g_{i\lambda_i}$ voor. Stel

$$(4.11) \quad G_{\alpha_i} = g_{i1} + g_{i2} + \ldots + g_{i\alpha_i} \quad \text{waar } \alpha_i \in \{1,2,\ldots \lambda_i\} \text{ en}$$

$G_o = 0$.

---

3). Tensy anders gespesifiseer, deurloop α_i die waardes $1,2,\ldots,\lambda_i$ en μ_i die waardes $1,2,\ldots,g_{\alpha_i}$.

Definieer

(4.12) $a_{ip_i} = g_{\alpha_i}^{-1} \sum_{\mu_i} \Psi_i[(G_{\alpha_i-1} + \mu_i)/(k_i+1)]$ wanneer

$G_{\alpha_i-1} + 1 \leq p_i \leq G_{\alpha_i}$.

Stel

(4.13) $\sigma_{ai}^2 = k_i^{-1} \sum_{p_i} (a_{ip_i} - \bar{\Psi}_i)^2$ waar

(4.14) $\bar{\Psi}_i = k_i^{-1} \sum_{p_i} a_{ip_i}$

$= k_i^{-1} \sum_j \sum_h \Psi_i(\delta_{ijh})$.

4.2.3. Definisies.

Definieer

(4.15) $a_{ijh} = a_{ip_i}$ indien $x_{ijh} = z_{ip_i}$.

Stel

(4.16) $u_{ij} = \begin{cases} \sum_h (a_{ijh} - \bar{\Psi}_i) & \text{indien } k_{ij}>0 \\ 0 & \text{indien } k_{ij}=0 \end{cases}$ en

(4.17) $u_j = m^{-\frac{1}{2}} \sum_i u_{ij}$

$= m^{-\frac{1}{2}} \sum_i \sum_h (a_{ijh} - \bar{\Psi}_i)$.

Stel verder

(4.18) $k_{\cdot j} = \sum_i k_{ij}$.

Opmerking.

In hierdie behandeling van m-rangskikkings word alle rye weggelaat waarin alle rangnommers dieselfde is (een enkele knoop in die ry) of waarin $k_{ij} = 0$ vir enige (k-1) waardes van j uit 1,2,...,k. Sodanige rye lewer geen bydrae tot die studie van bogenoemde probleem nie omdat dan geld $u_{ij} = 0$ vir alle j.

4.2.4. 'n Aantal Lemmas.

Lemma 4.1. Die verwagtingswaarde van u_j onder H_o word, vir vaste j, gegee deur

(4.19) $E(u_j|H_o) = 0$ [4].

Bewys: Op analoë wyse as in lemma 2.1 volg:

$$(4.20)\quad \begin{aligned} E(u_{ij}) &= E[\sum_h (a_{ijh} - \bar{\Psi}_i)] \\ &= E(\sum_h a_{ijh}) - E(\sum_h \bar{\Psi}_i) \\ &= k_{ij}k_i^{-1} \sum_{p_i} a_{ip_i} - k_{ij}\bar{\Psi}_i \\ &= 0 \text{ sodat} \\ E(u_j) &= E(m^{-\frac{1}{2}} \sum_i u_{ij}) \\ &= m^{-\frac{1}{2}} \sum_i E(u_{ij}) \\ &= 0. \end{aligned}$$

Lemma 4.2. Die variansie van u_j onder H_o word gegee deur

$$(4.21)\quad var(u_j) = m^{-1} \sum_i k_{ij}(k_i - k_{ij})K_i$$

waar

$$(4.22)\quad \begin{aligned} K_i &= k_i^{-1}(k_i-1)^{-1} \sum_{p_i}(a_{ip_i} - \bar{\Psi}_i)^2 \\ &= (k_i-1)^{-1}\sigma_{ai}^2 \end{aligned}$$

waar σ_{ai}^2 in (4.13) gedefinieer is oor alle knope.

Bewys: $$\begin{aligned} \sigma_{ij}^2 &= var(u_{ij}) \\ &= E(u_{ij}^2) \quad \text{omdat } E(u_{ij}) = 0 \\ &= E[\sum_h (a_{ijh} - \bar{\Psi}_i)]^2 \\ &= \binom{k_i}{k_{ij}}^{-1} \Sigma' [\sum_h (a_{ijh} - \bar{\Psi}_i)]^2 \end{aligned}$$

waar Σ' die sommasie oor alle moontlike kombinasies van k_{ij} groothede uit k_i aandui, wat almal ewekansig is onder H_o, sodat

---

4). In hierdie hoofstuk word, tensy anders vermeld, onder die hipotese H_o gewerk. $E(u_j)$ byvoorbeeld dui dus $E(u_j|H_o)$ aan. In die gevalle waar knope voorkom, word voorwaardelik onder die gegewe stel waarnemings gewerk.

(4.23) $\sigma^2_{ij} = k_{ij}(k_i-k_{ij})k_i^{-1}(k_i-1)^{-1}\sum_{p_i}(a_{ip_i} - \bar{\Psi}_i)^2$

netsoos in lemma 2.2

$= k_{ij}(k_i-k_{ij})K_i$ sodat

(4.24) $\sigma^2_j = \mathrm{var}(u_j)$

$= E(u_j^2)$ omdat $E(u_j) = 0$ (lemma 4.1)

$= E(m^{-\frac{1}{2}} \sum_i u_{ij})^2$

$= m^{-1} \sum_i E(u^2_{ij}) + m^{-1} \sum_{i\neq i'}\sum E(u_{ij}.u_{i'j})$

$= m^{-1} \sum_i \sigma^2_{ij} + 0$ omdat alle rangskikkings onderling onafhanklik is

$= m^{-1} \sum_i k_{ij}(k_i-k_{ij})K_i.$

Opmerking.

Uit $K_i = k_i^{-1}(k_i-1)^{-1}\sum_{p_i}(a_{ip_i} - \bar{\Psi}_i)^2$ volg dat

(4.25) $K_i \geq D > 0$ vir alle i omdat die rangnommers r_{ip_i} (vir vaste i) nie almal dieselfde is nie (opmerking §4.2.3); omdat $\Psi_i(\delta)$ monotoon stygend is in δ en omdat alle k_i eindig bly as $m\to\infty$.

Ook is

(4.26) $(k_i-k_{ij}) \geq 1$ vir enige i en j, want:

$k_i-k_{ij} = 0$ alleenlik indien

i) alle $k_{ij} = 0$

of ii) $k_{ij} > 0$, $k_{ij'} = 0$ vir alle $j' \neq j$.

Hierdie twee gevalle word deur die opmerking in §4.2.3 uitgesluit.

Dus

(4.27) $\sigma^2_{ij} \geq k_{ij}D$ en uit (4.24):

(4.28) $\sigma^2_j \geq Dm^{-1} \sum_i k_{ij}$

$= Dk_{.j}/m > 0.$

Lemma 4.3. Die kovariansie tussen u_j en $u_{j'}$ $(j' \neq j)$ onder H_o word gegee deur

(4.29) $\text{kov}(u_j, u_{j'}) = -m^{-1} \sum_i k_{ij} k_{ij'} K_i$.

Bewys: Netsoos in lemma 2.3 volg dat

(4.30) $$\begin{aligned} \text{kov}(u_{ij}, u_{ij'}) &= E(u_{ij}.u_{ij'}) \text{ omdat } E(u_{ij})=0 \text{ uit } (4.20) \\ &= -k_{ij}k_{ij'} k_i^{-1}(k_i-1)^{-1} \sum_{p_i} (a_{ip_i} - \bar{\Psi}_i)^2 \\ &= -k_{ij}k_{ij'} K_i \quad \text{sodat} \end{aligned}$$

$$\begin{aligned} \sigma_{jj'} &= \text{kov}(u_j, u_{j'}) \\ &= E(u_j . u_{j'}) \text{ omdat } E(u_j) = 0 \text{ vir alle } j \\ &= E(m^{-\frac{1}{2}} \sum_i u_{ij} . m^{-\frac{1}{2}} \sum_{i'} u_{i'j'}) \\ &= E(m^{-1} \sum_i u_{ij} u_{ij'} + m^{-1} \sum_{i \neq i'} \sum u_{ij} u_{i'j'}) \\ &= m^{-1} \sum_i E(u_{ij} u_{ij'}) + 0 \\ &= -m^{-1} \sum_i k_{ij} k_{ij'} K_i . \end{aligned}$$

Opmerking.

Die uitdrukkings vir $\text{var}(u_j)$ en $\text{kov}(u_j, u_{j'})$ is analoog aan dié wat afgelei is deur B+v.E(1953) vergelyking (2.3.1).

4.3. DIE TOETSINGSGROOTHEID T.

4.3.1. 'n Limietverdeling onder H_o.

Stel

(4.31) $v_i = \sum_j c_j u_{ij}$ waar die c_j 's willekeurige eindige reële getalle is, dan is

(4.32) $$\begin{aligned} E(v_i) &= E(\sum_j c_j u_{ij}) \\ &= \sum_j c_j E(u_{ij}) \\ &= 0 \text{ uit } (4.20) \quad \text{en} \end{aligned}$$

(4.33) $$\begin{aligned} \sigma_{i.}^2 &= \text{var}(v_i) \\ &= E(v_i^2) \\ &= E(\sum_j c_j u_{ij})^2 \end{aligned}$$

$$= E(\sum_j c_j^2 u_{ij}^2 + \sum_{j\neq j'}\sum c_j c_{j'}\, u_{ij}u_{ij'})$$

$$= \sum_j c_j^2 E(u_{ij}^2) + \sum_{j\neq j'}\sum c_j c_{j'}\, E(u_{ij}u_{ij'})$$

$$= \sum_j c_j^2 k_{ij}(k_i - k_{ij})K_i - \sum_{j\neq j'}\sum c_j c_{j'}\, k_{ij}k_{ij'}\, K_i$$

uit (4.23) en (4.30)

$$= K_i \sum_j c_j k_{ij}[c_j(k_i - k_{ij}) - \sum_{j'(\neq j)} c_{j'}\, k_{ij'}]$$

$$= K_i \sum_j c_j k_{ij} \sum_{j'} k_{ij'}(c_j - c_{j'}).$$

STELLING 4.1.

Indien

(4.34) $\sigma_{i.}^2 > 0$ vir $i=1,2,\ldots,m$,

dan besit $\sum_j c_j u_j$ asimptoties, vir $m\to\infty$, 'n normaalverdeling onder H_o.

Bewys: Omdat alle k_i eindig is, geld vir alle waardes van i, j en h dat

(4.35) $0 < \delta_{ijh} < 1$.

Gevolglik is $\Psi_i(\delta_{ijh})$ eindig vir enige waardes van i, j en h (uit die definisie van $\Psi_i(\delta)$).

Omdat alle k_{ij} eindig is, volg dus dat

(4.16) $u_{ij} = \sum_h (a_{ijh} - \overline{\Psi}_i)$ eindig is, sodat

(4.36) $E|u_{ij}|^{2+\delta} < \infty$ vir alle i en j en vir $\delta > 0$.

Vir eindige c_j geld

(4.37) $E|c_j u_{ij}|^{2+\delta} < \infty$.

Uit lemma 3.11 volg nou dat

(4.38) $E|\sum_j c_j u_{ij}|^{2+\delta} < \infty$ omdat k eindig is.

Dus

(4.39) $E|v_i|^{2+\delta} < \infty$ vir $i=1,2,\ldots,m$.

Die groothede v_i en $v_{i'}$ is onderling onafhanklik indien $i' \neq i$. Volgens die sentrale limietstelling (Liapounoff-voorwaarde) besit $m^{-\frac{1}{2}} \sum_{i=1}^{m} v_i$ onder voorwaarde (4.34) asimptoties, vir $m\to\infty$, 'n normaalverdeling.

Maar $m^{-\frac{1}{2}}\sum\limits_{i} v_i = m^{-\frac{1}{2}}\sum\limits_{i}\sum\limits_{j} c_j u_{ij}$

$= \sum\limits_{j} c_j u_j$ m.b.v. (4.17).

Gevolglik besit $\sum\limits_{j} c_j u_j$ asimptoties, vir $m \to \infty$, 'n normaalverdeling.

Opmerking.

'n Nodige en voldoende voorwaarde dat

(4.34) $\sigma_{i.}^{2} > 0$ vir $i=1,2,\ldots,m$,

is dat die c_j 's nie almal gelyk is nie en vir tenminste twee verskillende c_j 's die ooreenkomstige k_{ij} 's > 0.

Bewys: Uit (4.25) volg $K_i \geq D > 0$ vir alle i.

Gevolglik is $\sigma_{i.}^{2} > 0$ as en slegs as

(4.40) $\sum\limits_{j} c_j k_{ij} \sum\limits_{j'} k_{ij'}(c_j - c_{j'}) > 0$ (sien (4.33)).

Maar $\sum\limits_{j} c_j k_{ij} \sum\limits_{j'} k_{ij'}(c_j - c_{j'})$

$$= \sum_{j} c_j k_{ij} \sum_{j'(<j)} k_{ij'}(c_j - c_{j'}) + \sum_{j} c_j k_{ij} \sum_{j'(>j)} k_{ij'}(c_j - c_{j'})$$

$$= \sum_{j} \sum_{j'(<j)} k_{ij} k_{ij'}\, c_j(c_j - c_{j'}) + \sum_{j} \sum_{j'(<j)} k_{ij} k_{ij'}\, c_{j'}(c_{j'} - c_j)$$

$$= \sum_{j} \sum_{j'(<j)} k_{ij} k_{ij'}\, [c_j(c_j - c_{j'}) + c_{j'}(c_{j'} - c_j)]$$

$$= \sum_{j} \sum_{j'(<j)} k_{ij} k_{ij'}(c_j - c_{j'})^2$$

wat positief is as en slegs as bogenoemde voorwaardes geld.

Lemma 4.4. Indien

(4.41) $\lim\limits_{m \to \infty} m^{-1} k_{.j} > 0$ vir $j=1,2,\ldots,k$,

dan voldoen die u_j 's onder H_o aan die voorwaardes

(4.42) $\lim\limits_{m \to \infty} \mathrm{var}(u_j) > 0$ vir $j=1,2,\ldots,k$ en

(4.43) $\lim\limits_{m \to \infty} E|u_j|^{2+\delta} < \infty$ vir $j=1,2,\ldots,k$ en $\delta > 0$.

Bewys: $\lim\limits_{m \to \infty} \mathrm{var}(u_j) \geq D \lim\limits_{m \to \infty} m^{-1} k_{.j}$ uit (4.28)

> 0 m.b.v. (4.41), sodat (4.42) bevredig word.

Uit die bewys van stelling 4.1 volg, omdat alle k_i na bo begrens is, dat

(4.36') $E|u_{ij}|^{2+\delta} \leq M < \infty$ vir $i=1,2,\ldots,m$; $j=1,2,\ldots,k$,

vaste $\delta > 0$ en M 'n eindige positiewe getal.

Dit geld in die besonder ook vir $\delta = 2$.

Nou is

$$E(u_j^4) = E[m^{-\frac{1}{2}} \sum_i u_{ij}]^4$$

$$= E[m^{-2} \sum_i \sum_{i'} \sum_{i''} \sum_{i'''} u_{ij} u_{i'j} u_{i''j} u_{i'''j}]$$ waar alle i 's gaan van 1 tot m

$$= m^{-2} \sum_i \sum_{i'} E(u_{ij}^2) E(u_{i'j}^2) + m^{-2} \sum_i E(u_{ij}^4)$$ want u_{ij} en $u_{i'j}$ is onderling onafhanklik vir $i \neq i'$ en $E(u_{ij}) = 0$ uit (4.20)

$< \infty$ onafhanklik van m uit (4.36').

Hieruit volg dat (4.43) geld vir $\delta \leq 2$.

Lemma 4.5. Indien

(4.41) $\lim_{m \to \infty} m^{-1} k_{.j} > 0$ vir $j=1,2,\ldots,k$,

dan is die gesamentlike verdeling van die variante u_j onder H_o asimptoties normaal as $m \to \infty$.

Bewys: Omdat

(4.42) $\lim_{m \to \infty} \operatorname{var}(u_j) > 0$ vir $j=1,2,\ldots,k$,

(4.43) $\lim_{m \to \infty} E|u_j|^{2+\delta} < \infty$ vir $j=1,2,\ldots,k$ en $\delta > 0$

en omdat $\sum_j c_j u_j$ asimptoties normaal verdeel is as $m \to \infty$ (stelling 4.1) [5], volg die resultaat met behulp van lemma 2.5.

---

5). Aan voorwaarde (4.34) word in die algemeen voldoen - vergelyk stelling 4.1 (opmerking).

Opmerkings.

1). Die karakteristieke funksie van die gesamentlike verdeling van die u_j 's het dus die vorm

$e^{-\frac{1}{2}\vec{\theta}'\Sigma\vec{\theta}}$ waar Σ die momentematriks van die u_j 's voorstel, $\vec{\theta}' = \{\theta_1,\theta_2,\ldots,\theta_k\}$ en $\vec{\theta}$ die ooreenkomstige kolomvektor.

2). Neem as voorwaarde

(4.44) Σ is van rang $(k-1)$.

Omdat $\sum\limits_j u_j = 0$ (sien (4.17)), kan enige u_j uitgedruk word in terme van die ander $(k-1)$ u_j 's, naamlik

(4.45) $u_{j'} = -\sum\limits_{j(\neq j')} u_j$,

maar onder voorwaarde (4.44) kan $u_{j'}$ nie in 'n kleiner aantal as $(k-1)$ u_j 's uitgedruk word nie.

Dui die momentematriks van $u_1,u_2,\ldots,u_{k-1}$ aan deur $\Lambda = \{\lambda_{\zeta\zeta'}\}$ waar ζ en ζ' nou die waardes $1,2,\ldots,(k-1)$ deurloop. Onder voorwaarde (4.44) is Λ dus nie-singulier.

Dui die kofaktore van die element $\lambda_{\zeta\zeta'}$ van Λ aan deur $|\Lambda_{\zeta\zeta'}|$.

STELLING 4.2.

Indien

(4.41) $\lim\limits_{m\to\infty} m^{-1}k_{.j} > 0$ vir $j=1,2,\ldots,k$ en

(4.44) Σ van rang $(k-1)$ is,

dan besit die toetsingsgrootheid

(4.46) $$T = \sum_{\zeta=1}^{k-1}\sum_{\zeta'=1}^{k-1}\frac{|\Lambda_{\zeta\zeta'}|}{|\Lambda|}u_\zeta u_{\zeta'}$$

asimptoties, vir $m\to\infty$, onder H_o 'n χ^2-verdeling met $(k-1)$ grade van vryheid.

Bewys: Die bewys volg met behulp van lemma 4.5 en §1.4.

Opmerkings.

1). T kan meer kompak geskryf word in die vorm

$$T = \frac{|\Delta_u|}{|\Delta|}$$ waar die simbole soos volg gedefinieer word:

Beskou die matriks Δ_u verkry uit

$$V_u = \left\{\begin{matrix} \sigma_1^2, & \sigma_{12}, & \ldots\ldots, & \sigma_{1k}, & u_1 \\ \sigma_{21}, & \sigma_2^2, & \ldots\ldots, & \sigma_{2k}, & u_2 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \sigma_{k1}, & \sigma_{k2}, & \ldots\ldots, & \sigma_k^2, & u_k \\ u_1, & u_2, & \ldots\ldots, & u_k, & 0 \end{matrix}\right\}$$

deur weglating van 'n willekeurige ry en kolom behalwe die laaste ry en kolom, en bereken sy determinant $|\Delta_u|$.

Beskou ook die matriks Δ verkry uit $\Sigma = \{\sigma_{jj'}\}$ $(j, j' = 1, 2, \ldots, k)$ deur weglating van dieselfde ry en kolom as in die eerste geval, en bereken sy determinant $|\Delta|$.

2). Die toetsingsgrootheid T is 'n uitbreiding van die toetsingsgrootheid χ_r^2 van B+v.E(1953), naamlik na die geval waar gewerk word met funksies van rangnommers in plaas van slegs met rangnommers.

3). Stelling 4.2 kan uitgebrei word na die geval waar Σ van rang $\nu \leq (k-1)$ is. Hierop gaan ons nie verder in nie, behalwe dat ons hier 'n stelling van B+v.E(1953) vermeld: Indien 'n aantal k_{ij} 's = 0, kan dit gebeur dat ons in Σ twee of meer komplementêre groepe elemente het (vergelyk B+v.E(1953) p. 361), d.w.s. dat in elke ry (en elke kolom) van Σ slegs elemente van één van die groepe voorkom. Hierdie groepe word genoem „non-compared" groepe elemente, en die aantal sulke groepe word aangedui deur s.

Lemma 4.6. (B+v.E(1953) stelling 2): As en slegs as daar s „non-compared" groepe elemente in 'n k x k momentematriks van die u_j 's voorkom, is die rang van die matriks gelyk aan (k-s).

4.3.2. Die toetsingsgrootheid van BENARD en VAN ELTEREN(1953).

Stel ons $\Psi_i(\delta_{ijh}) = r_{ijh}$, dan herlei u_{ij} van vergelyking (4.16) na $\tilde{u}_{ij}$ van (4.6), d.w.s. tot die tipe grootheid wat deur B+v.E(1953) gebruik word.

Stel voorts: $\tilde{u}_j = m^{-\frac{1}{2}} \sum_i \tilde{u}_{ij}$ 6).

Definieer

$$\tilde{\chi}_r^2 = \frac{|\tilde{\Delta}_u|}{|\tilde{\Delta}|}$$

waar alle simbole soortgelyk is as in die voorgaande paragrawe, behalwe dat nou deurgaans geld $\Psi_i(\delta_{ijh}) = r_{ijh}$.

B+v.E(1953) het bewys dat die toetsingsgrootheid $\tilde{\chi}_r^2$ onder die voorwaardes

(4.47) die aantal kolomme k is eindig

(4.48) die aantal rye m neig na ∞

(4.49) $\lim\limits_{m\to\infty} \sum\limits_i \tilde{\sigma}_j^{-3} E|\tilde{u}_{ij}|^3 = 0$ vir $j=1,2,\ldots,k$ en

(4.50) die rang van die matriks $\tilde{R} = \{\tilde{\rho}_{jj'}\}$ waar

$\tilde{\rho}_{jj'} = \lim\limits_{m\to\infty} \tilde{\sigma}_{jj'}/\tilde{\sigma}_j\tilde{\sigma}_{j'}$, is gelyk aan (k-1),

asimptoties 'n χ^2-verdeling met (k-1) g.v.v. besit.

Hierby word aangeneem dat alle k_{ij} eindig bly as $m\to\infty$.

---

6). In §4.1 is genoem dat B+v.E(1953) met kolomtotale $\tilde{u}_j = \sum_i \tilde{u}_{ij}$ werk. Ons wysig egter die definisie van $\tilde{u}_j$ sodat $0 < \lim\limits_{m\to\infty} \text{var}(\tilde{u}_j) < \infty$. Hierdeur word die toetsingsgrootheid $\tilde{\chi}_r^2$ (sien die definisie hieronder) egter nie beïnvloed nie aangesien die addisionele m 'e in $\tilde{\chi}_r^2$ mekaar uitkanselleer.

B+v.E(1953) het ook bewys dat die voorwaardes

(4.51) $\lim\limits_{m\to\infty} m^{-1}\sum\limits_i k_{ij} > 0$ vir $j=1,2,\ldots,k$ [7] en

(4.52) die matriks

$$\tilde{C} = \begin{Bmatrix} \tilde{\varkappa}_{11}, \tilde{\varkappa}_{12}, \ldots, \tilde{\varkappa}_{1k} \\ \tilde{\varkappa}_{21}, \tilde{\varkappa}_{22}, \ldots, \tilde{\varkappa}_{2k} \\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot \\ \tilde{\varkappa}_{k1}, \tilde{\varkappa}_{k2}, \ldots, \tilde{\varkappa}_{kk} \end{Bmatrix}$$

met $\tilde{\varkappa}_{jj'} = \lim\limits_{m\to\infty} m^{-1}\sum\limits_i k_{ij}k_{ij'}$, is nie 'n matriks van die

tipe $\begin{pmatrix} P & O' \\ O & Q \end{pmatrix}$ nie waar P en Q vierkante matrikse is en O en O' uit nulle bestaan,

voldoende is vir die geldigheid van (4.49) en (4.50).

Omdat $E(u_{ij}.u_{ij'}) = 0$ as en slegs as $k_{ij}k_{ij'} = 0$ (sien vergelykings (4.25) en (4.30)), kan voorwaarde (4.44) deur voorwaarde (4.52) vervang word (want (4.52) impliseer dat (4.44) geld).

4.4. DIE TOETSINGSGROOTHEID T_k.

4.4.1. Inleiding.

Die toetsingsgrootheid T soos gedefinieer in vergelyking (4.46) is baie algemeen, maar dit is in 'n moeilik hanteerbare vorm.

Ons gaan ons vervolgens beperk tot 'n eenvoudiger toetsingsgrootheid T_k waartoe T reduseer indien alle $k_{ij} = b > 0$. Ten spyte van hierdie beperking, is T_k nog 'n uitbreiding op die probleem van m-rangskikkings soos deur FRIEDMAN(1937) en KENDALL(1948) behandel.

---

7). B+v.E(1953) beweer dat die voorwaarde $\lim\limits_{m\to\infty} m^{-1}\sum\limits_i k_i > 0$ voldoende is vir hulle stelling 5. Die voorwaarde moet egter wees $\lim\limits_{m\to\infty} m^{-1}\sum\limits_i k_{ij} > 0$ vir $j=1,2,\ldots,k$. Hierdie voorwaarde is dieselfde as voorwaarde

(4.41) $\lim\limits_{m\to\infty} m^{-1}k_{.j} > 0$ vir $j=1,2,\ldots,k$.

4.4.2. Definisie en Limietverdeling van T_k.

Stel $k_{ij} = b > 0$ vir alle i en j, dan is

$k_{i} = kb$ vir alle i,

$k_{.j} = mb$ vir alle j en

$N = mkb$.

Laat

(4.53) $u_j = m^{-\frac{1}{2}} \sum_i \sum_{\eta=1}^{b} (a_{ij\eta} - \bar{\Psi}_i)$ soos in (4.17) 8).

Verder is

(4.54) $\sigma_j^2 = m^{-1} b(kb-b) \sum_i K_i$ uit (4.24)

$= b(k-1)[mk(kb-1)]^{-1} \sum_i \sum_{p=1}^{kb} (a_{ip} - \bar{\Psi}_i)^2$ uit (4.22)

$= (k-1)K$ waar

(4.55) $K = b[mk(kb-1)]^{-1} \sum_i \sum_p (a_{ip} - \bar{\Psi}_i)^2$.

Stel

(4.56) $\sigma_j^2 = \sigma^2$ vir alle j.

Verder is

(4.57) $\sigma_{jj'} = - b^2 m^{-1} \sum_i K_i$ uit (4.29)

$= - K$.

$\Lambda = \{\lambda_{\varsigma\varsigma'}\}$ is nou die $(k-1)x(k-1)$ matriks met determinant

$$|\Lambda| = \begin{vmatrix} (k-1)K, & -K & ,\ldots\ldots, & -K \\ -K & ,(k-1)K, & \ldots\ldots, & -K \\ \cdot & \cdot & \cdot & \cdot \\ -K & ,-K & ,\ldots\ldots, & (k-1)K \end{vmatrix}$$

$$= K^{k-1} \begin{vmatrix} (k-1), & -1 & ,\ldots\ldots\ldots, & -1 \\ -1 & ,(k-1), & \ldots\ldots\ldots, & -1 \\ \cdot & \cdot & \cdot & \cdot \\ -1 & ,-1 & ,\ldots\ldots\ldots, & (k-1) \end{vmatrix}$$

$= K^{k-1} k^{k-2}$, want:

Neem $A = \begin{Bmatrix} a & , -1 \\ -1 & , a \end{Bmatrix}$, 'n 2x2 matriks.

Dan is $|A| = (a+1)(a-1)$.

---

8). η deurloop die waardes 1,2,...,b en p die waardes 1,2,...,kb.

Neem $A' = \begin{Bmatrix} a, & -1, & -1 \\ -1, & a, & -1 \\ -1, & -1, & a \end{Bmatrix}$, 'n 3x3 matriks, dan is

$|A'| = (a+1)^2(a-2).$

Neem $A'' = \begin{Bmatrix} a, & -1, & -1, & -1 \\ -1, & a, & -1, & -1 \\ -1, & -1, & a, & -1 \\ -1, & -1, & -1, & a \end{Bmatrix}$, 'n 4x4 matriks, dan is

$|A''| = (a+1)^3(a-3)$, ens.

Vir $A^* = \begin{Bmatrix} a, & -1, & \ldots, & -1 \\ -1, & a, & \ldots, & -1 \\ \cdot & \cdot & \cdot & \cdot \\ -1, & -1, & \ldots, & a \end{Bmatrix}$, 'n (k-1)x(k-1) matriks, is

(4.58) $|A^*| = (a+1)^{k-2}(a-k+2).$

Stel $a = (k-1)$ in (4.58), dan is die determinant van die (k-1)x(k-1) matriks

$$\begin{Bmatrix} (k-1), & -1, & \ldots, & -1 \\ -1, & (k-1), & \ldots, & -1 \\ \cdot & \cdot & \cdot & \cdot \\ -1, & -1, & \ldots, & (k-1) \end{Bmatrix} \text{ gelyk aan } k^{k-2}.$$

As kofaktore van Λ vind ons:

$$|\Lambda_{\zeta\zeta}| = \begin{vmatrix} (k-1)K, & -K, & \ldots, & -K \\ -K, & (k-1)K, & \ldots, & -K \\ \cdot & \cdot & \cdot & \cdot \\ -K, & -K, & \ldots, & (k-1)K \end{vmatrix}, \text{ 'n } (k-2)x(k-2) \text{ determinant}$$

$$= K^{k-2} \begin{vmatrix} (k-1), & -1, & \ldots, & -1 \\ -1, & (k-1), & \ldots, & -1 \\ \cdot & \cdot & \cdot & \cdot \\ -1, & -1, & \ldots, & (k-1) \end{vmatrix}$$

$= 2K^{k-2}k^{k-3}$ m.b.v. (4.58) met $a = (k-1)$.

Soortgelyk is

$|\Lambda_{\zeta\zeta'}| = K^{k-2}k^{k-3}$ vir $\zeta' \neq \zeta$.

Vervolgens is

(4.59) $\dfrac{|\Lambda_{\zeta\zeta}|}{|\Lambda|} = \dfrac{2}{Kk}$ en

(4.60) $\dfrac{|\Lambda_{\zeta\zeta'}|}{|\Lambda|} = \dfrac{1}{Kk}$ vir $\zeta' \neq \zeta$ sodat

$$(4.61) \quad T_k = \sum_{\zeta=1}^{k-1} \sum_{\zeta'=1}^{k-1} \frac{|\Lambda_{\zeta\zeta'}|}{|\Lambda|} u_\zeta u_{\zeta'} \qquad \text{- sien } (4.46)$$

$$= \frac{2}{Kk} \sum_\zeta u_\zeta^2 + \frac{1}{Kk} \sum_{\zeta \neq \zeta'}\sum u_\zeta u_{\zeta'}$$

$$= \frac{2}{Kk} [\sum_\zeta u_\zeta^2 + \sum_{\zeta<\zeta'}\sum u_\zeta u_{\zeta'}].$$

Omdat $\sum_{j=1}^{k} u_j = 0$ uit die definisie van u_j, is

$$u_k = -\sum_{\zeta=1}^{k-1} u_\zeta,$$

$$u_k^2 = (-\sum_\zeta u_\zeta)^2 = \sum_\zeta u_\zeta^2 + 2 \sum_{\zeta<\zeta'}\sum u_\zeta u_{\zeta'} \quad \text{sodat}$$

$$(4.62) \quad 2\sum_\zeta u_\zeta^2 + 2 \sum_{\zeta<\zeta'}\sum u_\zeta u_{\zeta'} = \sum_\zeta u_\zeta^2 + u_k^2$$

$$= \sum_{j=1}^{k} u_j^2 .$$

Uit (4.61) en (4.62) volg dan

$$(4.63) \quad T_k = \frac{1}{Kk} \sum_{j=1}^{k} u_j^2$$

$$= \frac{(k-1)}{k\sigma^2} \sum_j u_j^2 \text{ uit } (4.54) \text{ en } (4.56).$$

STELLING 4.3.

Die toetsingsgrootheid

$$T_k = \frac{(k-1)}{k\sigma^2} \sum_j u_j^2$$

besit onder H_o asimptoties, vir $m \to \infty$, 'n χ^2-verdeling met $(k-1)$ grade van vryheid.

Bewys: T_k is 'n spesiale geval van T wat onder voorwaardes (4.41) en (4.44) asimptoties χ^2 verdeel is as $m \to \infty$ (stelling 4.2).

Nou is $\lim_{m \to \infty} m^{-1} k_{.j} = \lim_{m \to \infty} m^{-1} mb = b > 0$ sodat (4.41) bevredig word. Aan voorwaarde (4.44) word ook voldoen want $|\Lambda| = K^{k-1} k^{k-2} > 0$.

Die stelling volg dus direk.

Opmerkings.

1). FRIEDMAN(1937) se bewys deur karakteristieke funksies kan uitgebrei word na hierdie geval.

2). $$\begin{aligned} E(T_k) &= \frac{(k-1)}{k\sigma^2} E(\sum_j u_j^2) \\ &= \frac{(k-1)}{k\sigma^2} \sum_j E(u_j^2) \\ &= \frac{(k-1)}{k\sigma^2} \cdot k\sigma^2 \\ &= (k-1) \\ &= \text{aantal g.v.v. van } T_k. \end{aligned}$$

4.4.3. Spesiale gevalle van T_k.

4.4.3.1. BENARD en VAN ELTEREN(1953).

Stel $\Psi_i(\delta_{ij\eta}) = r_{ij\eta}$, dan herlei

$K_i = [kb(kb-1)]^{-1} \sum_p (a_{ip} - \bar{\Psi}_i)^2$ – sien (4.22) – na

(4.64) $K_i = [k^3b^3 - \sum_{\alpha_i} g_{\alpha_i}^3]/[12bk(kb-1)]$ uit (2.61), sodat

(4.65) $$\begin{aligned} \sigma^2 &= (k-1)K \text{ uit (4.54) en (4.56)} \\ &= m^{-1}(k-1)b^2 \sum_i K_i \\ &= b(k-1)[12mk(kb-1)]^{-1} \sum_i (k^3b^3 - \sum_{\alpha_i} g_{\alpha_i}^3). \end{aligned}$$

Verder is

(4.66) $u_j = m^{-\frac{1}{2}} \sum_i \sum_\eta [r_{ij\eta} - \frac{1}{2}(kb+1)]$ uit (4.53)

sodat T_k herlei na

(4.67) $$T_B = \frac{12(kb-1)}{b\sum_i (k^3b^3 - \sum_{\alpha_i} g_{\alpha_i}^3)} \sum_j [\sum_i \sum_\eta \left\{r_{ij\eta} - \tfrac{1}{2}(kb+1)\right\}]^2.$$

T_B is die toetsingsgrootheid wat deur BENARD en VAN ELTEREN(1953) in §3.3 bespreek word waarby aangeneem word dat alle kovariansies $\sigma_{jj'}$ (sien (4.30) waarby $\Psi_i(\delta_{ij\eta}) = r_{ij\eta}$) gelyk is. Nodig en voldoende hiervoor is dat alle k_{ij} gelyk is, sê $k_{ij} = b > 0$ vir alle i en j, en dit is presies die geval wat hierbo bespreek is.

Opmerking.

Indien $\Psi_i(\delta_{ij\eta}) = \delta_{ij\eta}$ geneem word, herlei T_k ook na T_B.

4.4.3.2. FRIEDMAN(1937).

Stel $k_{ij} = 1$ vir alle i en j, d.w.s. b = 1 in §4.4.3.1. Dan herlei T_k na

(4.68) $T_C = \frac{(k-1)}{k\sigma^2} \sum_j [m^{-\frac{1}{2}} \sum_i (a_{ij} - \bar{\Psi}_i)]^2$ uit (4.53) en (4.63)

waarby a_{ij} gedefinieer is in (4.12) en

(4.69) $\sigma^2 = (mk)^{-1} \sum_i \sum_j (a_{ij} - \bar{\Psi}_i)^2$ uit (4.54) en (4.55) met b = 1

$= m^{-1} \sum_i \sigma^2_{\Psi_i}$ waar

(4.70) $\sigma^2_{\Psi_i} = k^{-1} \sum_j (a_{ij} - \bar{\Psi}_i)^2$, sodat dan volg

(4.71) $T_C = \frac{(k-1)}{k \sum_i \sigma^2_{\Psi_i}} \sum_j [\sum_i (a_{ij} - \bar{\Psi}_i)]^2.$

Definieer

(4.72) $S^* = \sum_j [\sum_i \{r_{ij} - \frac{1}{2}(k+1)\}]^2$ (vergelyk (4.2)) en

(4.73) $T^* = (k-1)^{-1} \sum_i \sum_{\alpha_i} (g^3_{\alpha_i} - g_{\alpha_i})$ (vergelyk BENARD en VAN ELTEREN(1953) vergelyking (3.4.2)).

Stel weer $\Psi_i(\delta_{ij}) = r_{ij}$, dan is

$\sigma^2 = [12mk]^{-1} \sum_i (k^3 - \sum_{\alpha_i} g^3_{\alpha_i})$ uit (4.65) met b=1 sodat

(4.74) $\sigma^2_{\Psi_i} = [12k]^{-1}(k^3 - \sum_{\alpha_i} g^3_{\alpha_i})$, en herlei T_C na:

(4.75) $T_D = 12(k-1)S^* / \sum_i (k^3 - \sum_{\alpha_i} g^3_{\alpha_i})$

$= 12(k-1)S^*/[mk(k^2-1) - \sum_i \{\sum_{\alpha_i} g^3_{\alpha_i} - \sum_{\alpha_i} g_{\alpha_i}\}]$

$= 12S^*/[mk(k+1) - (k-1)^{-1} \sum_i \sum_{\alpha_i} (g^3_{\alpha_i} - g_{\alpha_i})]$

$= 12S^*/[mk(k+1) - T^*]$

$= \chi^2$ van FRIEDMAN(1937) soos aangegee deur B+v.E(1953) vergelykings (3.4.1) en (3.4.2).

Indien daar geen knope voorkom nie, d.i. as alle $g_{\alpha_i} = 1$, is $T^{*} = 0$ en herlei T_D na:

(4.76) $T_F = 12S^{*}/mk(k+1)$

$$= 12[mk(k+1)]^{-1} \sum_j (\sum_i \tilde{r}_{ij})^2 \quad \text{waar}$$

(4.77) $\tilde{r}_{ij} = r_{ij} - \frac{1}{2}(k+1)$.

Vergelyk FRIEDMAN(1937) vergelyking (18) en FRASER(1957) p. 262 probleem 22.

FRIEDMAN(1937) het bewys dat T_F asimptoties 'n χ^2-verdeling met (k-1) g.v.v. besit as $m\to\infty$.

Uit stelling 4.3 volg dat die meer algemene toetsingsgrootheid T_C, wat na $T_D = \chi^2$ herlei indien $\Psi_i(\delta_{ij})=r_{ij}$ en voorts na T_F indien daar geen knope voorkom nie, asimptoties χ^2 verdeel is onder H_0 met (k-1) g.v.v. as $m\to\infty$.

Opmerkings.

1). FRIEDMAN(1937) se toetsingsgrootheid is 'n spesiale geval van BENARD en VAN ELTEREN(1953) se toetsingsgrootheid T_B.

2). KENDALL(1948) het die grootheid T_F geskryf in die vorm

(4.78) $T_F = m(k-1)W$ waar

(4.79) $W = 12S^{*}/[m^2k(k^2-1)]$.

Kendall noem W die „coefficient of concordance".

3). Indien daar geen gelyke waarnemings voorkom nie, alle $\Psi_i(\delta) = \Psi(\delta)$ en alle $k_{ij} = 1$, dan herlei $\sigma^2_{\Psi_i}$ uit (4.70) na

(4.80) $$\sigma^2_{\Psi_i} = k^{-1}\sum_j [\Psi(\delta_{ij}) - \bar{\Psi}]^2 \quad \text{waar } \bar{\Psi} = k^{-1}\sum_j \Psi[j/(k+1)]$$

$$= \sigma^2_{\Psi} \text{ (sê).}$$

Dus $$\sigma^2_j = m^{-1}\sum_i \sigma^2_{ij} \quad \text{uit (4.24)}$$

$$= m^{-1}\sum_i \sigma^2_{\Psi}$$

$$= \sigma^2_{\Psi} .$$

Laat S_m die maksimum-waarde wees wat $S = \sum_j u_j^2$ onder bogenoemde voorwaardes kan aanneem. Dit word verkry as alle rangskikkings identies is. Die kolomtotale is dan (nie noodwendig in dié volgorde nie):

$\sqrt{m}\ \Psi(\frac{1}{k+1}),\ \sqrt{m}\ \Psi(\frac{2}{k+1}),\ \ldots\ ,\ \sqrt{m}\ \Psi(\frac{k}{k+1})$ sodat

$$S_m = \sum_j [\sqrt{m}\ \Psi(\frac{j}{k+1}) - \sqrt{m}\ \bar{\Psi}]^2$$

$$= m \sum_j [\Psi(\frac{j}{k+1}) - \bar{\Psi}]^2$$

$$= mk\sigma_\Psi^2 \text{ uit (4.80).}$$

Gevolglik herlei T_C uit (4.68) na

$$T_C' = \frac{(k-1)}{k\sigma_\Psi^2} \sum_j u_j^2 \quad \text{m.b.v. (4.69) en (4.80)}$$

$$= m(k-1)S/S_m \ .$$

4). TERPSTRA(1962) wys daarop dat KENDALL(1948) se bewys dat $m(k-1)W$ asimptoties χ^2 verdeel is as $m \to \infty$, foutief is (stelling VI).

5). In EHRENBERG, A.S.C.(1952):Biometrika 39, pp. 82-87, is 'n toetsingsgrootheid voorgestel vir die probleem van m-rangskikkings. TERPSTRA(1955,1956) het die asimptotiese verdeling daarvan bepaal vir m eindig en $k \to \infty$, terwyl VAN ELTEREN(1957) die asimptotiese verdeling bepaal het vir k eindig en $m \to \infty$ en aangetoon het dat die toetsingsgrootheid in 'n triviale geval ekwivalent is aan die toetsingsgrootheid van FRIEDMAN(1937).

4.4.3.3. $\Psi_i(\delta_{ij\eta}) \equiv \Xi_{r_{ij\eta}} \equiv E(\xi_{r_{ij\eta}})$

waar ξ_{ip_i} $(\xi_{i1} < \xi_{i2} < \ldots < \xi_{ik_i})$ die waardes in rang van 'n ewekansige steekproef van grootte $k_i = kb$ uit 'n normaal (0,1)-populasie is (vergelyk TERRY(1952) en §2.5.2.2).

In hierdie geval is: [9)]

$$\bar{\Xi}_i = \bar{\Psi}_i$$
$$= (kb)^{-1} \sum_p \Xi_{ip}$$
$$= 0 \text{ weens simmetrie.}$$

Verder is

(4.81) $u_j = m^{-\frac{1}{2}} \sum_i \sum_\eta a_{ij\eta}$ en

(4.82) $\sigma^2 = (k-1)K$ uit (4.54)

$$= (k-1)b[mk(kb-1)]^{-1} \sum_i \sum_p a_{ip}^2 \quad \text{waar}$$

(4.83) $a_{ip} = g_{\alpha_i}^{-1} \sum_{\mu_i} \Psi_i[(G_{\alpha_i-1} + \mu_i)/(k_i+1)]$

wanneer $G_{\alpha_i-1} + 1 \leq p \leq G_{\alpha_i}$

met $\Psi_i(\delta_{ip}) = \Xi_{ip}$ in hierdie geval.

Uit stelling 4.3 volg dat

$$T_k = \frac{(k-1)}{k\sigma^2} \sum_j u_j^2$$ onder H_0 asimptoties, vir $m \to \infty$,

'n χ^2-verdeling met (k-1) g.v.v. besit.

Opmerking.

Hierdie is 'n veralgemening van 'n probleem voorgestel deur FRASER(1957) (p. 263 probleem 23). Stel ons b = 1, dan herlei bostaande resultate na genoemde probleem van FRASER(1957) en kry ons:

$$u_j = m^{-\frac{1}{2}} \sum_i a_{ij} \quad \text{en}$$

$$\sigma^2 = (mk)^{-1} \sum_i \sum_j a_{ij}^2 \quad \text{sodat}$$

(4.84) $$T_k = \frac{m(k-1)}{\sum_i \sum_j a_{ij}^2} \sum_j u_j^2 .$$

---

9). Ons kan in hierdie geval ook met 'n toetsingsgrootheid van die vorm van (4.46) werk indien die k_{ij} 's nie almal gelyk is nie.

Indien geen knope voorkom nie, is $a_{ij} = \Xi_{ij}$, sodat

(4.85) $$T_k = \frac{(k-1)}{\sum_i \sum_j \Xi_{ij}^2} \sum_j [\sum_i \Xi_{ij}]^2 .$$

T_k is onder H_o asimptoties χ^2 verdeel met $(k-1)$ g.v.v. as $m \to \infty$.

4.4.3.4. $\Psi_i(\delta_{ij\eta}) \equiv Q(\delta_{ij\eta})$

waar $Q(q)$ die q^{de} kwantiel van 'n normaal (0,1)-verdeling is (sien VAN DER WAERDEN(1957) en §2.5.2.3).

In hierdie geval is [10)]

$$\begin{aligned} \bar{Q}_i &= \bar{\Psi}_i \\ &= (kb)^{-1} \sum_p Q(\delta_{ip}) \\ &= 0 \text{ weens simmetrie.} \end{aligned}$$

Verder is

(4.86) $$u_j = m^{-\frac{1}{2}} \sum_i \sum_\eta a_{ij\eta} \quad \text{en}$$

(4.87) $$\sigma^2 = (k-1)b[mk(kb-1)]^{-1} \sum_i \sum_p a_{ip}^2 \quad \text{uit (4.82)}$$

met a_{ip} soos in (4.83) waarby $\Psi_i(\delta_{ip}) = Q(\delta_{ip})$.

Soos in die geval van §4.4.3.3 volg dat

$$T_k = \frac{(k-1)}{k\sigma^2} \sum_j u_j^2$$ onder H_o asimptoties, vir $m \to \infty$, 'n χ^2-verdeling met $(k-1)$ g.v.v. besit.

Opmerking.

Stel $b = 1$, dan is

$$u_j = m^{-\frac{1}{2}} \sum_i a_{ij} \quad \text{en}$$

$$\sigma^2 = (mk)^{-1} \sum_i \sum_j a_{ij}^2 \quad \text{sodat}$$

(4.88) $$T_k = \frac{m(k-1)}{\sum_i \sum_j a_{ij}^2} \sum_j u_j^2 .$$

Indien daar geen knope voorkom nie, is $a_{ip} = Q(\delta_{ip})$ sodat

(4.89) $$T_k = \frac{(k-1)}{\sum_i \sum_j Q^2(\delta_{ij})} \sum_j [\sum_i Q(\delta_{ij})]^2 .$$

---

10). Sien voetnoot 9.

T_k is onder H_o asimptoties χ^2 verdeel met $(k-1)$ g.v.v. as $m\to\infty$.

4.4.3.5. Neem vervolgens die meer algemene funksies $\Psi_i(\delta_{ijh}, x_{ijh})$, $i=1,2,\ldots,m$, wat strengstygend (of strengdalend) is in beide argumente en eindig is vir alle eindige x_{ijh} en vir $0 < \delta_{ijh} < 1$, met

$$\left|\frac{\partial\Psi_i}{\partial x}\right| \geq \varepsilon > 0 \text{ vir alle } i \text{ en } x.$$

Laat (sien byvoorbeeld vergelyking (4.12)):

$$(4.90)\quad a_{ip_i} = g_{\alpha_i}^{-1} \sum_{\mu_i} \Psi_i [\, z_{i,G_{\alpha_i-1}+\mu_i}, (G_{\alpha_i-1}+\mu_i)/(k_i+1)]$$

vir $G_{\alpha_i-1}+1 \leq p_i \leq G_{\alpha_i}$.

Definieer

$$(4.91)\quad a_{ijh} = a_{ip_i} \quad \text{indien } x_{ijh} = z_{ip_i} ,$$

$$(4.92)\quad \bar{\Psi}_i = k_i^{-1} \sum_{p_i} \Psi_i(x_{ip_i}, \delta_{ip_i})$$

$$= k_i^{-1} \sum_{p_i} a_{ip_i} .$$

Laat weer

$$(4.93)\quad u_{ij} = \sum_h (a_{ijh} - \bar{\Psi}_i) \quad \text{indien } k_{ij} > 0,$$

$$(4.94)\quad u_j = m^{-\frac{1}{2}} \sum_i u_{ij}.$$

Die resultate van §^{e} 4.2 tot 4.4.2 kan vir hierdie geval aangepas word. Ons gee slegs 'n kort aanduiding.

Dui die x_{ijh}-waardes aan deur $z_1, z_2, \ldots, z_N$ en laat $Z = (z_1, z_2, \ldots, z_N)$.

Soos in lemmas 4.1 tot 4.3 kan bewys word dat

$$(4.95)\quad E(u_j \mid H_o; Z) = 0,$$

$$(4.96)\quad \sigma_j^2 = \text{var}(u_j \mid H_o; Z)$$
$$= m^{-1} \sum_i k_{ij}(k_i - k_{ij}) k_i^{-1} (k_i - 1)^{-1} \sum_{p_i} (a_{ip_i} - \bar{\Psi}_i)^2 \quad \text{en}$$

$$(4.97)\quad \sigma_{jj'} = \text{kov}(u_j, u_{j'} \mid H_o; Z)$$
$$= -\, m^{-1} \sum_i k_{ij} k_{ij'} k_i^{-1} (k_i - 1)^{-1} \sum_{p_i} (a_{ip_i} - \bar{\Psi}_i)^2 .$$

Ons kies 'n getal $\omega > 0$ en vereis dat in elke ry tenminste twee x'e in verskillende „selle" met minstens ω verskil. Alle rye wat nie aan hierdie voorwaarde voldoen nie, word weggelaat.

Omdat alle x_{ijh} eindig is, volg soos in die opmerking na lemma 4.2 dat $\sigma_j^2 > 0$ vir m eindig, en $\lim_{m\to\infty} \sigma_j^2 \geq \omega' > 0$.

Neem ons weer $v_i = \sum_j c_j u_{ij}$ waar die c_j 's eindige reële getalle is wat nie almal gelyk is nie, en $\sigma_{i.}^2 = \text{var}(v_i)$, dan kan ons op analoë wyse as tevore die volgende stellings bewys:

STELLING 4.4.

Indien

(4.98) $\sigma_{i.}^2 > 0$ vir $i=1,2,\ldots,m$,

besit $\sum_j c_j u_j$ asimptoties, vir $m\to\infty$, onder H_0 'n normaalverdeling.

Bewys: Soos stelling 4.1.

Soortgelyke resultate word ook gevind as in lemmas 4.4 en 4.5, sodat ons kry:

STELLING 4.5.

Indien

(4.41) $\lim_{m\to\infty} m^{-1}k_{.j} > 0$ vir $j=1,2,\ldots,k$ en

(4.99) $\Sigma = \{\sigma_{jj'}\}$ van rang $(k-1)$ is vir elke gegewe Z

dan besit

$$T = \sum_{\zeta}\sum_{\zeta} \frac{|\Lambda_{\zeta\zeta'}|}{|\Lambda|} u_\zeta u_{\zeta'}$$

asimptoties, vir $m\to\infty$, onder H_0 'n χ^2-verdeling met $(k-1)$ g.v.v.

Bewys: Soos stelling 4.2.

Stel $k_{ij} = b > 0$ vir alle i en j, dan is (sien §4.4.2)

(4.100) $\sigma_j^2 = (k-1)b[mk(kb-1)]^{-1} \sum_i \sum_p (a_{ip} - \bar{\Psi}_i)^2$ vir alle j

$= \sigma^2$ (sê), en

$$u_j = m^{-\frac{1}{2}} \sum_i \sum_{\eta=1}^{b} (a_{ij\eta} - \bar{\Psi}_i).$$

Definieer weer

$$T_k = \frac{(k-1)}{k\sigma^2} \sum_j u_j^2$$

dan besit T_k asimptoties, vir $m \to \infty$, onder H_0 'n χ^2-verdeling met $(k-1)$ g.v.v. (sien stelling 4.3).

Stel nou $\Psi_i(x_{ij\eta}, \delta_{ij\eta}) \equiv x_{ij\eta}$ (vergelyk PITMAN(1937) en §2.6.2).

Laat $\bar{x}_i = \bar{\Psi}_i$

$$= (kb)^{-1} \sum_p x_{ip}.$$

Dan is

(4.101) $u_j = m^{-\frac{1}{2}} \sum_i \sum_\eta (x_{ij\eta} - \bar{x}_i)$ en

(4.102) $\sigma^2 = (k-1)b[mk(kb-1)]^{-1} \sum_i \sum_p (x_{ip} - \bar{x}_i)^2.$

$T_k = \frac{(k-1)}{k\sigma^2} \sum_j u_j^2$ besit asimptoties 'n χ^2-verdeling onder H_0 met $(k-1)$ g.v.v. as $m \to \infty$.

Opmerking.

Stel $b = 1$, dan reduseer bostaande geval na dié voorgestel deur FRASER(1957) (p.262 probleem 21). In dié geval is:

$$\bar{x}_i = k^{-1} \sum_j x_{ij} \;,\quad \bar{x}_{.j} = m^{-1} \sum_i x_{ij} \;,\quad \bar{x} = (mk)^{-1} \sum_i \sum_j x_{ij} \;,$$

$$u_j = m^{-\frac{1}{2}} \sum_i (x_{ij} - \bar{x}_i),$$

$$\sigma^2 = (mk)^{-1} \sum_i \sum_j (x_{ij} - \bar{x}_i)^2$$

$$= m^{-1} \sum_i S_i^2 \quad \text{waar}$$

(4.103) $S_i^2 = k^{-1} \sum_j (x_{ij} - \bar{x}_i)^2$ sodat

(4.104) $T_k = \frac{(k-1)}{k \sum_i S_i^2} \sum_j [\sum_i (x_{ij} - \bar{x}_i)]^2$

$$= \frac{m^2(k-1)}{k \sum_i S_i^2} \sum_j (\bar{x}_{.j} - \bar{x})^2.$$

T_k besit asimptoties, vir $m \to \infty$, onder H_o 'n χ^2-verdeling met (k-1) grade van vryheid.

4.5. ANDER UITBREIDINGS.

4.5.1. Inleiding.

In die probleem van m-rangskikkings ($§^e$ 4.1 tot 4.4) is die asimptotiese verdelings van voorgestelde toetsingsgroothede ondersoek indien die aantal rangskikkings $m \to \infty$. Die vraag ontstaan wat sal gebeur indien alle $k_{ij} \to \infty$, of indien $k \to \infty$.

Alleen eersgenoemde geval lewer 'n sinvolle bydrae tot die studie van ons probleem. (Sien ook §4.4.3.2 opmerking 5 waar 'n geval aangehaal word waar laasgenoemde geval $(k \to \infty)$ bestudeer word).

4.5.2. Geval $k_{ij} \to \infty$ vir $j=1,2,\ldots,k$; $i=1,2,\ldots,m$, waarby k en m eindig is.

Hierdie is 'n twee-faktor variansie-analise waar rangnommers afsonderlik binne elke ry toegeken word. Elke ry kan beskou word as bestaande uit k willekeurige groot steekproewe (die m rye dui m replikate aan). Die teorie van hoofstuk II is dus toepasbaar op elke ry. In notasie sluit ons aan by dié van hoofstuk II.

Die waarnemings in die sel (i,j) kan beskou word as 'n steekproef afkomstig uit 'n populasie met verdelingsfunksie F_{ij}. Onder H_o' nl. dat alle verdelingsfunksies F_{ij} identies is vir gegewe i, sê $= F_i$, is alle permutasies van waarnemings in elke ry ewekansig (H_o).

Neem

$$(4.105)\quad y_{ij} = \begin{cases} k_{ij}^{-\frac{1}{2}} \sum\limits_h (a_{ijh} - \bar{\Psi}_i) & \text{indien } k_{ij} > 0 \\ 0 & \text{indien } k_{ij} = 0 \end{cases}$$

en $y_j = \sum\limits_i y_{ij}$.

Hier word veronderstel dat die Ψ_i funksies is van rangnommers alleen, d.w.s. van die vorm $\Psi_i(\delta_{ijh})$.

Lemma 4.7. Indien

(4.106) $\max_{p_i}[\Psi_i(\delta_{ip_i})-\bar{\Psi}_i]^2 = o(k_i)$ vir $k_i \to \infty$ en $i=1,2,..,m$

(4.107) (Iedere diskontinuïteit van $F_i(x)) \leq \theta < 1$ vir alle i

(4.108) $\lim_{k_i \to \infty} k_{ij}/k_i \geq \varepsilon > 0$ vir alle i en j,

dan is

(4.109) $w_i = \sum_j c_{ij} y_{ij}$

waar die c_{ij}'s willekeurige eindige reële getalle is, asimptoties normaal verdeel as $k_i \to \infty$, $i=1,2,\dots,m$.

Bewys: Die bewys is, vir vaste i, analoog aan dié van stelling 2.3.

Opmerking.

Voorwaarde (4.108) is nie nodig nie solank $k_{ij} \to \infty$ as $k_i \to \infty$ (vergelyk stelling 2.3 opmerking 2).

STELLING 4.6.

Onder voorwaardes (4.106), (4.107) en (4.108) en onder H_0' is $\sum_j e_j \hat{y}_j$, waarby die e_j's willekeurige eindige reële getalle is en

(4.110) $\hat{y}_j = (N-k_{.j})^{\frac{1}{2}} N^{-\frac{1}{2}} y_j \sigma_j^{-1}$

met $\sigma_j^2 = \text{var}(y_j)$ en $k_{.j} = \sum_i k_{ij}$,

asimptoties normaal verdeel as $N \to \infty$.

Bewys: w_i en $w_{i'}$ is onderling onafhanklik indien $i \neq i'$.

Met behulp van lemma 4.7 volg dus dat

(4.111) $W = \sum_i^{l} w_i$ asimptoties normaal verdeel is as $N \to \infty$

(as $N \to \infty$, geld dat alle $k_{ij} \to \infty$ omdat k en m eindig is en omdat ons aanneem dat alle k_{ij} van dieselfde orde grootte is - voorwaarde (4.108)).

Omdat die c_{ij} 's willekeurig is (behalwe in soverre dat almal nie gelyk mag wees nie) kan ons sonder verlies aan algemeenheid stel

(4.112) $c_{ij} = (N-k_{.j})^{\frac{1}{2}} N^{-\frac{1}{2}} e_j \sigma_j^{-1}$ vir alle i.

Dan is

$$(4.113) \quad W = \sum_i \sum_j c_{ij} y_{ij}$$
$$= \sum_i \sum_j (N-k_{.j})^{\frac{1}{2}} N^{-\frac{1}{2}} e_j \sigma_j^{-1} y_{ij}$$
$$= \sum_j (N-k_{.j})^{\frac{1}{2}} N^{-\frac{1}{2}} e_j \sigma_j^{-1} \sum_i y_{ij}$$
$$= \sum_j (N-k_{.j})^{\frac{1}{2}} N^{-\frac{1}{2}} e_j \sigma_j^{-1} y_j$$
$$= \sum_j e_j \hat{y}_j \quad \text{m.b.v. } (4.110).$$

$\sum_j e_j \hat{y}_j$ is dus asimptoties normaal verdeel as $N \to \infty$.

Lemma 4.8. Onder die voorwaardes (4.106), (4.107) en (4.108) en onder H_0 geld die voorwaardes

$$(4.114) \quad \lim_{N\to\infty} \text{var}(\hat{y}_j) > 0 \quad \text{en}$$

$$(4.115) \quad \lim_{N\to\infty} E|\hat{y}_j|^{2+\delta} < \infty \text{ vir } j=1,2,\ldots,k \text{ en } \delta > 0.$$

Bewys: Soos in die opmerking na stelling 2.2 volg m.b.v. STOKER(1955) stelling 2.4 dat

$$\lim_{N\to\infty} \inf. \text{ var}(y_{ij}) \geq \varepsilon > 0.$$

Maar $\text{var}(y_j) = \sum_i \text{var}(y_{ij})$ – sien byvoorbeeld (4.24).

Dus $\lim_{N\to\infty} \inf. \text{ var}(y_j) \geq \lim_{N\to\infty} \inf. \text{ var}(y_{ij}) > 0$ vir enige i.

$$\text{var}(\hat{y}_j) = (N-k_{.j})N^{-1}\text{var}(y_j)\ \sigma_j^{-2}$$
$$= 1 - k_{.j}N^{-1} \quad \text{sodat}$$
$$\lim_{N\to\infty} \text{var}(\hat{y}_j) = 1 - \lim_{N\to\infty} k_{.j}N^{-1} > 0 \quad \text{waarmee } (4.114)$$

bevredig word.

Indien, wat verder deurgaans veronderstel word,

$$k_i^{-1} \sum_{p_i} (a_{ip_i} - \bar{\Psi}_{i.})^4 = O(k_i) \text{ vir } i = 1,2,\ldots,m$$

dan volg op analoë wyse as in bylaag A dat

$$E(y_{ij})^4 < \infty \text{ vir alle i en j.}$$

Aangesien y_{ij} en $y_{i'j}$ onafhanklik is indien $i \neq i'$, volg maklik dat voorwaarde (4.115) bevredig word.

Lemma 4.9. Die variante $\hat{y}_j$ is onder voorwaardes (4.106), (4.107) en (4.108) gesamentlik asimptoties normaal verdeel as $N \to \infty$.

Bewys: Die bewys volg direk uit stelling 4.6 en lemma 4.8 met behulp van lemma 2.5.

Opmerking.

Die karakteristieke funksie van die $\hat{y}_j$'s is dus van die vorm

$$e^{-\frac{1}{2}\vec{\theta}'\Sigma\vec{\theta}}$$ waar Σ die momentematriks van die $\hat{y}_j$'s voorstel, $\vec{\theta}' = \{\theta_1,\theta_2,\ldots,\theta_k\}$ en $\vec{\theta}$ die ooreenkomstige kolomvektor.

Lemma 4.10. Die momentematriks Σ van die variante $\hat{y}_j$, $j=1,2,\ldots,k$ besit onder H_0 'n rang van (k-1).

Bewys: Met behulp van (4.23) en (4.30) volg:

$$(4.116)\quad \mathrm{var}(y_{ij}) = (k_i-k_{ij})k_i^{-1}(k_i-1)^{-1}\sum_{p_i}(a_{ip_i}-\bar{\Psi}_i)^2$$
$$= (k_i-k_{ij})K_i \quad \text{en}$$

$$(4.117)\quad \mathrm{kov}(y_{ij},y_{ij'}) = -(k_{ij}k_{ij'})^{\frac{1}{2}}k_i^{-1}(k_i-1)^{-1}\sum_{p_i}(a_{ip_i}-\bar{\Psi}_i)^2$$
$$= -(k_{ij}k_{ij'})^{\frac{1}{2}}K_i \quad \text{vir } j \neq j' \quad \text{waar}$$

$$(4.22)\qquad K_i = k_i^{-1}(k_i-1)^{-1}\sum_{p_i}(a_{ip_i}-\bar{\Psi}_i)^2.$$

Die momentematriks van die y_{ij} 's (vir vaste i) is van rang (k-1), want tussen die elemente van elke ry bestaan 'n lineêre verband, naamlik

$$(4.118)\quad \sum_j k_{ij}\,\mathrm{var}(y_{ij}) + \sum_{j\neq j'}\sum (k_{ij}k_{ij'})^{\frac{1}{2}}\,\mathrm{kov}(y_{ij},y_{ij'})$$
$$= \sum_j k_{ij}(k_i-k_{ij})K_i - \sum_{j\neq j'}\sum k_{ij}k_{ij'}K_i$$
$$= \sum_j k_{ij}K_i[k_i-k_{ij}-\sum_{j'(\neq j)}k_{ij'}]$$
$$= 0$$

en geen ander lineêre onafhanklike verband bestaan nie.

Die transformasie van y_{ij} na $\hat{y}_j$ is lineêr en nie-singulier. Gevolglik is die rang van Σ, die momentematriks van die $\hat{y}_j$'s, ook (k-1).

Opmerking.

Dui, soos in die opmerking na lemma 4.5, die momentematriks van enige (k-1) $\hat{y}_j$'s aan deur $\Lambda = \{\lambda_{\zeta\varsigma}\}$ met kofaktore $|\Lambda_{\zeta\varsigma}|$.

STELLING 4.7.

Indien

(4.106) $\underset{p_i}{\text{maks}}[\Psi_i(\delta_{ip_i}) - \bar{\Psi}_i]^2 = o(k_i)$ vir $k_i \to \infty$ en $i=1,2,..,n$

(4.107) (Iedere diskontinuïteit van $F_i(x)) \leq \theta < 1$ vir alle i

(4.108) $\lim\limits_{k_i \to \infty} k_{ij}/k_i \geq \varepsilon > 0$ vir alle i en j,

dan is

$$T = \sum_{\zeta=1}^{k-1} \sum_{\varsigma=1}^{k-1} \frac{|\Lambda_{\zeta\varsigma}|}{|\Lambda|} \hat{y}_\zeta \hat{y}_\varsigma \quad \text{onder } H_0' \text{ asimptoties}$$

χ^2 verdeel met (k-1) grade van vryheid as $N \to \infty$.

Bewys: Uit lemma 4.9 volg dat die momentematriks van die $\hat{y}_\zeta$'s van die vorm

$$e^{-\frac{1}{2}\vec{t}'\Lambda\vec{t}} \quad \text{is,}$$

en m.b.v. lemma 4.10 dat Λ van rang (k-1) is.

Die bewys volg nou m.b.v. §1.4.

Opmerking.

T kan tot meer hanteerbare vorms vereenvoudig word deur sekere beperkings te stel, byvoorbeeld alle $k_{ij} = b$ (sien §4.4). Aangesien die prosedure dieselfde is as in hoofstuk II, gaan ons nie verder daarop in nie.

4.6. DIE ALTERNATIEWE HIPOTESE.

4.6.1. Inleiding.

Tot dusver in hierdie hoofstuk het ons die geval onder H_0 beskou waarby veronderstel word dat alle moontlike rangskikkings van rangnommers binne elke ry gelykkansig is. Hier word egter die meer algemene hipotese H beskou waarby aangeneem word dat al die km items uit dieselfde of verskillende populasies afkomstig is, en wel dat die waarnemings x_{ijh}, $h=1,2,\ldots,k_{ij}$ van die item x_{ij}, afkomstig is uit 'n populasie met kontinue verdelingsfunksie F_{ij}, $i=1,2,\ldots,m$; $j=1,2,\ldots,k$.

Die limietverdeling van die voorgestelde toetsingsgrootheid word onder H bepaal indien $m \to \infty$.

4.6.2. Definisies.

Neem

(4.119) $$t_{ij} = k_{ij}^{-1} \sum_{h=1}^{k_{ij}} \Psi_i(\delta_{ijh})$$

waar $\Psi_i(\delta)$ eindig is vir $0 < \delta < 1$ en $\delta_{ijh} = r_{ijh}/(k_i+1)$ soos in §4.1. Stel

(4.120) $$t'_{ij} = t_{ij} - E(t_{ij}|H) \ ,$$

(4.121) $$t_j = m^{-\frac{1}{2}} \sum_{i=1}^{m} t'_{ij} \ ,$$

(4.122) $$\sigma_{jj'}^{(i)} = \text{kov}(t'_{ij}, t'_{ij'}|H) = E(t'_{ij}, t'_{ij'}|H) \text{ vir } j \neq j' \text{ want}$$

(4.123) $$E(t'_{ij}|H) = 0;$$

(4.124) $$\sigma_{jj}^{(i)} = \text{var}(t'_{ij}|H),$$

(4.125) $$\sigma'_{jj'} = \text{kov}(t_j, t_{j'}|H) = E(t_j . t_{j'}|H) \text{ vir } j \neq j' \quad \text{want}$$

(4.126) $$E(t_j|H) = 0 \quad \text{en}$$

(4.127) $$\sigma'_{jj} = \sigma_j'^2 = \text{var}(t_j|H).$$

Hieruit volg maklik dat

(4.128) $$\sigma'_{jj'} = m^{-1} \sum_i \sigma_{jj'}^{(i)} \quad \text{met}$$

(4.129) $$\sigma_j'^2 = m^{-1} \sum_i \sigma_{jj}^{(i)}.$$

4.6.3. Die Toetsingsgrootheid T_A. Limietverdeling onder H.

Stel

(4.130) $$V_i = \sum_j c_j t'_{ij}$$ waar die c_j 's willekeurige eindige reële getalle is wat nie almal gelyk is nie.

STELLING 4.8.

Indien

(4.131) $\mathrm{var}(V_i|H) > 0$ vir $i=1,2,\ldots,m$,

dan besit $\sum_j c_j t_j$ onder H asimptoties, vir $m\to\infty$, 'n normaalverdeling.

Bewys: Uit die definisie van V_i volg dat t_{ij}, en dus ook t'_{ij}, eindig is vir alle i en j ($k_i < \infty$ vir alle i).

Dus

(4.132) $E|t'_{ij}|^{2+\delta} < \infty$ vir alle i en j en vir $\delta > 0$.

Soos in stelling 4.1 volg eerstens dat

(4.133) $E|V_i|^{2+\delta} < \infty$ vir alle $i=1,2,\ldots,m$

en vervolgens dat $\sum_j c_j t_j = m^{-\frac{1}{2}} \sum_i V_i$ asimptoties, vir $m\to\infty$, 'n normaalverdeling besit.

Opmerking.

$\mathrm{var}(V_i|H)$ kan ook geskryf word:

$$
\begin{aligned}
(4.134)\quad \mathrm{var}(V_i) &= E[V_i - E(V_i)]^2 \\
&= E[\sum_j c_j t'_{ij} - E(\sum_j c_j t'_{ij})]^2 \\
&= E[\sum_j c_j t'_{ij}]^2 \qquad \text{want } E(t'_{ij}) = 0 \\
&= \sum_j \sum_{j'} c_j c_{j'} E(t'_{ij}.t'_{ij'}) \\
&= \sum_j \sum_{j'} c_j c_{j'} \sigma^{(i)}_{jj'} .
\end{aligned}
$$

Lemma 4.11. Indien

(4.131) $\mathrm{var}(V_i|H) > 0$ vir $i=1,2,\ldots,m$ en

(4.135) $\sigma^{(i)}_{jj} > 0$ vir alle i en j,

dan voldoen die t_j 's aan die voorwaardes

(4.136) $\lim\limits_{m\to\infty} \mathrm{var}(t_j|H) > 0$ vir $j=1,2,\ldots,k$ en

(4.137) $\lim\limits_{m\to\infty} E|t_j|^{2+\delta} < \infty$ vir $j=1,2,\ldots,k$ en $\delta > 0$.

Bewys: Uit (4.129) volg

$\mathrm{var}(t_j|H) = m^{-1}\sum_i \sigma_{jj}^{(i)}$ sodat m.b.v. (4.135) volg

(4.136) $\lim_{m\to\infty} \mathrm{var}(t_j|H) > 0.$

In stelling 4.8 is bewys dat

(4.132) $E|t'_{ij}|^{2+\delta} < \infty$ vir alle i en j en $\delta > 0$.

Soos in lemma 4.4 volg nou dat

(4.137) $\lim_{m\to\infty} E|t_j|^{2+\delta} < \infty$ vir $j=1,2,\ldots,k$ en $\delta > 0$.

Lemma 4.12. Indien voorwaardes (4.131) en (4.135) geld, is die gesamentlike verdeling van die variante t_j asimptoties normaal as $m\to\infty$.

Bewys: Omdat

(4.136) $\lim_{m\to\infty} \mathrm{var}(t_j|H) > 0$ vir $j=1,2,\ldots,k$,

(4.137) $\lim_{m\to\infty} E|t_j|^{2+\delta} < \infty$ vir $j=1,2,\ldots,k$ en $\delta > 0$ (lemma 4.11) en omdat $\sum_j c_j t_j$ asimptoties normaal verdeel is as $m\to\infty$ (stelling 4.8), volg die resultaat m.b.v. lemma 2.5.

Opmerkings.

1). Die karakteristieke funksie van die gesamentlike verdeling van die t_j 's is dus

$e^{-\frac{1}{2}\vec{\theta}'\Sigma\vec{\theta}}$ waar $\Sigma = \{\sigma'_{jj}\}$ die momentematriks van die t_j 's voorstel.

2). Neem as voorwaarde

(4.138) Σ is van rang (k-1).

Dui, soos in opmerking 2 na lemma 4.5, die momentematriks van enige (k-1) t_j 's aan deur $\Lambda' = \{\lambda'_{\zeta\varsigma}\}$ met kofaktore $|\Lambda'_{\zeta\varsigma}|$ waar $\zeta,\varsigma = 1,2,\ldots,(k-1)$.

STELLING 4.9.

Indien

(4.131) $var(V_i|H) > 0$ vir $i=1,2,...,m$,

(4.135) $\sigma_{jj}^{(i)} > 0$ vir alle i en j en

(4.138) Σ van rang $(k-1)$ is,

dan besit die toetsingsgrootheid

(4.139) $$T_A = \sum_{\zeta}\sum_{\zeta'} \frac{|\Lambda'_{\zeta\zeta'}|}{|\Lambda'|} t_\zeta t_{\zeta'}$$

onder H asimptoties, vir $m\to\infty$, 'n χ^2-verdeling met $(k-1)$ grade van vryheid.

Bewys: Die stelling volg direk uit lemma 4.12 en §1.4.

Opmerkings.

1). Die berekening van $\sigma_{jj'}^{(i)}$ skyn baie ingewikkeld te wees, behalwe onder H_o in welke geval ons dit in lemmas 4.2 en 4.3 aangegee het.

2). Deur allerlei vereenvoudigings kan spesiale gevalle uit T_A afgelei word.

3). Die stelling kan ook bewys word deur van karakteristieke funksies gebruik te maak soos FRIEDMAN(1937), maar die voorwaardes is baie meer beperkend as in ons geval (byvoorbeeld: Vir elke vaste j en j' moet $\sigma_{jj'}^{(i)}$ gelyk wees vir alle waardes van $i=1,2,...,m$).

4.7. ASIMPTOTIESE RELATIEWE DOELTREFFENDHEID.

Beskou die alternatiewe hipotese

(4.140) H_a: $F_{ij}(x) = F_i(x+d_j m^{-\frac{1}{2}})$

waar die d_j 's willekeurige eindige reële getalle is waarvoor sonder verlies aan algemeenheid veronderstel kan word dat

(4.141) $\sum_j d_j = 0$.

Ons neem nou deurgaans as voorwaardes

(4.142) $\lim_{m\to\infty} E(u_j|H_a) = B_j < \infty$ vir alle j en

(4.143) $\lim\limits_{m\to\infty} \dfrac{\text{kov}(u_j, u_{j'}|H_a)}{\text{kov}(u_j, u_{j'}|H_o)} = 1$ vir alle j en j'.

Aan laasgenoemde voorwaarde word waarskynlik in die algemeen voldoen, d.w.s. onder algemene reëlmatigheidsvoorwaardes en kontinuïteit van die funksies Ψ_i en f_i, waar $f_i \equiv F_i'$.

In die res van hierdie hoofstuk word veronderstel dat alle Ψ_i en f_i kontinu is, dat daar geen gelyke waarnemings voorkom nie (alle F_i is ook kontinu) en dat alle $k_{ij} = b > 0$.

Nou volg uit (4.17), (4.119), (4.120) en (4.121) met H_a in plaas van H:

(4.144) $t_j = b^{-1} u_j - b^{-1} E(u_j|H_a)$

met $b^{-1} E(u_j|H_a) = m^{-\frac{1}{2}} \sum\limits_i [b^{-1} \sum\limits_\eta E\{\Psi_i(\delta_{ij\eta})|H_a\} - \bar{\Psi}_i]$.

Uit stelling 4.2 volg dat

(4.46) $T = \sum\limits_\zeta \sum\limits_{\zeta'} \dfrac{|\Lambda_{\zeta\zeta'}|}{|\Lambda|} u_\zeta u_{\zeta'}$

onder H_o asimptoties 'n χ^2-verdeling met (k-1) g.v.v. besit as $m\to\infty$. (Aan die voorwaardes van stelling 4.2 word voldoen - sien stelling 4.3 - want vir alle $k_{ij} = b$ herlei T na T_k van stelling 4.3).

Omdat $\lim\limits_{m\to\infty} \text{var}(t_j|H_a)$ eindig en positief is (lemma 4.11), volg dat

(4.145) $0 < \lim\limits_{m\to\infty} \text{var}(u_j|H_a) < \infty$ vir alle j.

Uit stelling 4.8 volg dat $\sum\limits_j c_j t_j$ asimptoties normaal verdeel is onder H_a en voorwaarde (4.131) as $m\to\infty$, waar die c_j 's willekeurige eindige reële getalle is.

Maar

(4.146) $\sum\limits_j c_j t_j = \sum\limits_j c_j [b^{-1}u_j - b^{-1}E(u_j|H_a)]$ uit (4.144)

$= b^{-1} \sum\limits_j c_j u_j - b^{-1} \sum\limits_j c_j E(u_j|H_a)$, sodat ook

$\sum\limits_j c_j u_j$ asimptoties normaal verdeel is onder bogenoemde voorwaardes.

Gevolglik besit die u_j 's onder H_a asimptoties, vir $m\to\infty$, gesamentlik 'n normaalverdeling indien voorwaarde (4.142) ook nog bevredig word.

Uit voorwaarde (4.143) volg dat die u_j 's onder H_a asimptoties, vir $m\to\infty$, dieselfde momentematriks besit as onder H_o, naamlik Λ. Nou volg uit §1.4 dat T onder H_a asimptoties, vir $m\to\infty$, 'n nie-sentrale χ^2-verdeling besit met (k-1) g.v.v. en nie-sentraliteitsparameter

$$(4.147)\quad \lambda^2 = \sum_{\zeta}\sum_{\zeta'} \lim_{m\to\infty} \frac{|\Lambda_{\zeta\zeta'}|}{|\Lambda|} E(u_\zeta|H_a).E(u_{\zeta'}|H_a) .$$

Uit (4.142) en (4.145) volg dat λ^2 eindig is.

Aangesien alle $k_{ij} = b > 0$ en daar geen knope voorkom nie, herlei T na:

$$(4.148)\quad T_k = \frac{(k-1)}{k\sigma^2}\sum_j u_j^2 \quad \text{uit (4.63)}$$

$$= \frac{(k-1)}{k\sigma^2}\sum_j[m^{-\frac{1}{2}}\sum_i\sum_{\eta=1}^{b}\{\Psi_i(\delta_{ij\eta}) - \bar{\Psi}_i\}]^2 \text{ uit (4.17)}$$

met σ^2 gegee deur (4.54), (4.55) en (4.56).

Soortgelyk aan T herlei λ^2 m.b.v. (4.59) en (4.60) na:

$$(4.149)\quad \lambda_k^2 = \lim_{m\to\infty}\left\{\frac{2}{Kk}\sum_\zeta[E(u_\zeta|H_a)]^2+\frac{1}{Kk}\sum_{\zeta\neq\zeta'}\sum E(u_\zeta|H_a)E(u_{\zeta'}|H_a)\right\}$$

$$= \lim_{m\to\infty}\frac{(k-1)}{k\sigma^2}\left\{2\sum_\zeta[E(u_\zeta|H_a)]^2+\sum_{\zeta\neq\zeta'}\sum E(u_\zeta|H_a)E(u_{\zeta'}|H_a)\right\} .$$

Hierbo is

$$(4.150)\quad E(u_j|H_a) = E[\{m^{-\frac{1}{2}}\sum_i\sum_\eta[\Psi_i(\delta_{ij\eta}) - \bar{\Psi}_i]\}|H_a]$$

$$= m^{-\frac{1}{2}}\sum_i\sum_\eta E[\Psi_i(\delta_{ij\eta})|H_a] - bm^{-\frac{1}{2}}\sum_i\bar{\Psi}_i .$$

Nou is, vir vaste i,

$$(4.151)\quad E[\Psi_i(\delta_{ij\eta})|H_a) = E[\Psi_i(\frac{r_{ij\eta}}{kb+1})|H_a]$$

$$= \sum_{p=1}^{kb}\Psi_i(\frac{p}{kb+1})P[r_{ij\eta}=p|H_a] .$$

Dui vervolgens die verdelingsfunksie van die verdeling van $x_{ij\eta}$ aan deur $F''_{ij\eta}$ waarby $i=1,2,\ldots,m$; $j=1,2,\ldots,k$ en $\eta=1,2,\ldots,b$. Dan geld volgens §4.6.1 dat

(4.152) $F''_{ij1} \equiv F''_{ij2} \equiv \ldots \equiv F''_{ijb} \equiv F_{ij}$.

Vir vaste i is daar kb verdelingsfunksies $F''_{ij\eta}$; $j=1,2,\ldots,k$; $\eta=1,2,\ldots,b$. Dui nou hierdie verdelingsfunksies aan deur F'_{ig} waar $g=1,2,\ldots,kb$ sodanig dat $F'_{i1} \equiv F''_{i11}$, $F'_{i2} \equiv F''_{i12}$, ... , $F'_{ib} \equiv F''_{i1b}$, $F'_{i,b+1} \equiv F''_{i21}$, $F'_{i,b+2} \equiv F''_{i22}$, ens.

Uit (4.151) volg dus nou verder

$$(4.153)\quad E[\Psi_i(\delta_{ij\eta})|H_a] = \sum_p \Psi_i\left(\frac{p}{kb+1}\right) P[r_{ij\eta}=p|H_a]$$

$$= \Psi_i\left(\frac{1}{kb+1}\right) \int_{-\infty}^{\infty} \prod_{\substack{p'=1\\(\neq g)}}^{kb} [1-F'_{ip'}]\, dF'_{ig} +$$

$$+ \Psi_i\left(\frac{2}{kb+1}\right) \sum_{\substack{g'=1\\(\neq g)}}^{kb} \int_{-\infty}^{\infty} \prod_{\substack{p'=1\\(\neq g,g')}}^{kb} [1-F'_{ip'}]F'_{ig'}\, dF'_{ig} +$$

$$+ \Psi_i\left(\frac{3}{kb+1}\right) \sum_{\substack{g'<g''\\(\neq g)}}^{kb} \sum \int_{-\infty}^{\infty} \prod_{\substack{p'=1\\(\neq g,g',g'')}}^{kb} [1-F'_{ip'}]F'_{ig'}\, F'_{ig''}\, dF'_{ig} + \ldots\ldots +$$

$$+ \Psi_i\left(\frac{kb}{kb+1}\right) \int_{-\infty}^{\infty} \prod_{\substack{p'=1\\(\neq g)}}^{kb} F'_{ip'}\, dF'_{ig}$$

waar $F'_{ig} \equiv F''_{ij\eta}$, d.w.s $g = (j-1)b + \eta$ 11)

$$= \sum_{p=1}^{kb} \Psi_i\left(\frac{p}{kb+1}\right) \cdot$$

$$\cdot \sum_{\substack{g'<g''<\ldots<g^{(p-1)}\\(\text{almal } \neq g)}}^{kb} \sum \ldots \sum \int_{-\infty}^{\infty} \prod_{\substack{p'=1\\(\neq \{g,g',\ldots,g^{(p-1)}\})}}^{kb} [1-F'_{ip'}]F'_{ig'}F'_{ig''}\ldots F'_{ig^{(p-1)}}\, dF'_{ig} \,.$$

λ_k^2 word nou gevind deur terugsubstitusie van (4.153) en (4.150) in (4.149). Bostaande uitdrukking is in die algemeen onhanteerbaar. Ons beskou slegs die volgende spesiale geval.

11). $\prod_{p'=1}^{kb} F'_{ip'}$ dui die produk $F'_{i1}F'_{i2}\cdot \ldots \cdot F'_{i,kb}$ aan.

Spesiale geval: $\Psi_i(\delta_{ij\eta}) \equiv r_{ij\eta}$.

Nou is $\bar{\Psi}_i = (kb)^{-1} \sum_j \sum_\eta r_{ij\eta}$ uit (4.14)

$$= \tfrac{1}{2}(kb+1),$$

(4.154) $\quad u_j = m^{-\frac{1}{2}} \sum_i \sum_\eta [r_{ij\eta} - \tfrac{1}{2}(kb+1)]$ en

(4.155) $\quad \sigma^2 = b^2(k-1)(kb+1)/12$ uit (4.65).

T_k herlei dus tot

(4.156) $\quad T_k' = 12[mb^2k(kb+1)]^{-1} \sum_j [\sum_i \sum_\eta \{r_{ij\eta} - \tfrac{1}{2}(kb+1)\}]^2$.

Uit (4.153) volg:

$$E(r_{ij\eta}|H_a) = \int_{-\infty}^{\infty} \prod_{\substack{p'=1\\(\neq g)}}^{kb} [1 - F_{ip'}'] \, dF_{ig}' +$$

$$+ 2 \sum_{\substack{g'=1\\(\neq g)}}^{kb} \int_{-\infty}^{\infty} \prod_{\substack{p'=1\\(\neq g,g')}}^{kb} [1 - F_{ip'}'] F_{ig'}' \, dF_{ig}' +$$

$$+ 3 \sum_{\substack{g'<g''\\(\neq g)}}^{kb} \sum \int_{-\infty}^{\infty} \prod_{\substack{p'=1\\(\neq g,g',g'')}}^{kb} [1-F_{ip'}'] F_{ig'}' F_{ig''}' dF_{ig}' + \ldots . + $$

$$+ kb \int_{-\infty}^{\infty} \prod_{\substack{p'=1\\(\neq g)}}^{kb} F_{ip'}' \, dF_{ig}' .$$

Deur die terme uit te skryf en te hergroepeer, volg dan:

(4.157) $\quad E(r_{ij\eta}|H_a) = 1 + \sum_{g'(\neq g)}^{kb} \int_{-\infty}^{\infty} F_{ig'}' \, dF_{ig}'$ waarby

$F_{ig}' \equiv F_{ij\eta}'' \equiv F_{ij}$, d.w.s. $g = (j-1)b + \eta$.

Hierby is

(4.158) $$\sum_{g'(\neq g)}^{kb} \int_{-\infty}^{\infty} F_{ig'}' \, dF_{ig}' = \sum_{j'=1}^{k} \sum_{\eta'=1}^{b} \int_{-\infty}^{\infty} F_{ij'\eta'}'' \, dF_{ij\eta}'' - $$

$$- \int_{-\infty}^{\infty} F_{ig}' \, dF_{ig}'$$

$$= b \sum_{j'} \int_{-\infty}^{\infty} F_{ij'} \, dF_{ij} - \tfrac{1}{2}$$

$$= b \sum_{j'} \int_{-\infty}^{\infty} F_i(x+d_{j'}m^{-\frac{1}{2}}) \, dF_i(x+d_j m^{-\frac{1}{2}}) - \tfrac{1}{2}$$ met behulp van (4.140),

$$= b \sum_{j'} \int_{-\infty}^{\infty} F_i[x + (d_{j'} - d_j)m^{-\frac{1}{2}}]\, dF_i(x) - \tfrac{1}{2}$$

$$= b \sum_{j'} \int_{-\infty}^{\infty} [F_i(x) + (d_{j'} - d_j)m^{-\frac{1}{2}} f_i\{x+\Theta(d_{j'} - d_j)m^{-\frac{1}{2}}\}]dF_i(x) - \tfrac{1}{2}$$

m.b.v. Taylor se middelwaardestelling, waar $0<\Theta<1$, aangesien f_i orals kontinu veronderstel word

$$= b \sum_{j'}[\tfrac{1}{2} + (d_{j'} - d_j)m^{-\frac{1}{2}} \int_{-\infty}^{\infty} f_i\{x+\Theta(d_{j'} - d_j)m^{-\frac{1}{2}}\}dF_i(x)] - \tfrac{1}{2}$$

$$= \tfrac{1}{2}(kb-1) + bm^{-\frac{1}{2}} \sum_{j'} (d_{j'} - d_j) \int_{-\infty}^{\infty} f_i(x')dF_i(x) \text{ waar}$$

$$x' = x+\Theta(d_{j'} - d_j)m^{-\frac{1}{2}}.$$

Nou is vir die geval $\Psi_i(\delta_{ij\eta}) \equiv r_{ij\eta}$,

$$(4.159) \quad E(u_j|H_a) = m^{-\frac{1}{2}} \sum_i \sum_\eta E[r_{ij\eta}|H_a] - \tfrac{1}{2}m^{+\frac{1}{2}}b(kb+1) \text{ met behulp van } (4.154)$$

$$= m^{-\frac{1}{2}} \sum_i \sum_\eta [1+\tfrac{1}{2}(kb-1) + bm^{-\frac{1}{2}}\sum_{j'}(d_{j'} - d_j) \int_{-\infty}^{\infty} f_i(x')dF_i(x)] -$$

$$- \tfrac{1}{2}m^{+\frac{1}{2}}b(kb+1) \text{ uit } (4.157) \text{ en } (4.158)$$

$$= b^2m^{-1} \sum_{j'} (d_{j'} - d_j)\sum_i \int_{-\infty}^{\infty} f_i(x')dF_i(x).$$

Indien vir m voldoende groot geld

$$(4.160) \quad \int_{-\infty}^{\infty} f_i(x+cm^{-\frac{1}{2}})dF_i(x) < \infty \text{ waar c 'n willekeurige}$$

konstante is, dan geld, omdat f_i kontinu veronderstel word,

$$\int_{-\infty}^{\infty} f_i[x+\Theta(d_{j'} - d_j)m^{-\frac{1}{2}}]dF_i(x) = \int_{-\infty}^{\infty} f_i(x)dF_i(x) + o(1),$$

waarby $0<\Theta<1$.

Aan voorwaarde (4.160) word byvoorbeeld voldoen indien f_i gelykmatig begrens is.

Nou is

$$(4.161) \quad E(u_j|H_a) = b^2m^{-1}\sum_{j'}(d_{j'} - d_j)\sum_i \int_{-\infty}^{\infty} f_i(x)dF_i(x) + o(1)$$

$$= -b^2km^{-1}d_j \sum_i \int_{-\infty}^{\infty} f_i(x)dF_i(x) + o(1) \text{ met behulp van } (4.141).$$

Daar word dus in hierdie geval voldoen aan voorwaarde (4.142).

Uit (4.149), (4.155) en (4.161) volg:

$$(4.162)\quad \lambda_k^2 = \frac{12}{b^2k(kb+1)} \lim_{m\to\infty} [2\sum_\zeta \{E(u_\zeta|H_a)\}^2 + \sum_{\zeta\neq\zeta'}\sum E(u_\zeta|H_a)E(u_{\zeta'}|H_a)]$$

$$= \frac{12}{b^2k(kb+1)} \lim_{m\to\infty} [2\sum_\zeta\left\{\frac{b^2k}{m}d_\zeta \sum_i \int_{-\infty}^{\infty} f_i(x)dF_i(x)\right\}^2 + \sum_{\zeta\neq\zeta'}\sum\left\{\frac{b^2k}{m}d_\zeta\sum_i\int_{-\infty}^{\infty} f_i(x)dF_i(x)\right\}\left\{\frac{b^2k}{m}d_{\zeta'}\sum_i\int_{-\infty}^{\infty}f_i(x)dF_i(x)\right\}]$$

$$= \frac{12b^2k}{(kb+1)} \lim_{m\to\infty} m^{-2}[\sum_i \int_{-\infty}^{\infty} f_i(x)dF_i(x)]^2[2\sum_\zeta d_\zeta^2 + \sum_{\zeta\neq\zeta'}\sum d_\zeta d_{\zeta'}]$$

$$= \frac{12b^2k}{(kb+1)} \lim_{m\to\infty} m^{-2}[\sum_i \int_{-\infty}^{\infty} f_i(x)dF_i(x)]^2[\sum_\zeta d_\zeta^2 + (\sum_\zeta d_\zeta)^2]$$

$$= \frac{12b^2k}{(kb+1)} \lim_{m\to\infty} m^{-2}[\sum_i \int_{-\infty}^{\infty} f_i(x)dF_i(x)]^2 \cdot \sum_{j=1}^{k} d_j^2$$ met behulp van (4.141) aangesien $\zeta=1,2,\dots,(k-1)$.

Vir $b = 1$ herlei hierdie uitdrukking tot:

$$(4.163)\quad \lambda_k^2 = \frac{12k}{(k+1)} \lim_{m\to\infty} m^{-2}[\sum_i \int_{-\infty}^{\infty} f_i^2(x)\,dx]^2 \sum_j d_j^2$$

wat dieselfde is as dié in VAN ELTEREN en NOETHER(1959) vergelyking (9).

Beskou ons vervolgens die meer eenvoudige hipotese

$$(4.164)\quad H_a':\ F_{ij}(x) = F[x+(\delta_i+d_j)m^{-\frac{1}{2}}]$$

waar die δ_i 's en d_j 's willekeurige eindige reële getalle is en die d_j 's nog voldoen aan die voorwaarde

$$(4.141)\quad \sum_j d_j = 0,$$

dan herlei λ_k^2 uit (4.162) na:

$$(4.165)\quad \lambda_k^2 = \frac{12b^2k}{(kb+1)} \lim_{m\to\infty} m^{-2}[m\int_{-\infty}^{\infty} f^2(x)\,dx]^2 \sum_j d_j^2$$

$$= \frac{12b^2k}{(kb+1)}\left[\int_{-\infty}^{\infty} f^2(x)\,dx\right]^2 \sum_j d_j^2 .$$

Die nie-sentraliteitsparameter van die F-toets vir hierdie geval kan gevind word deur die voorbeeld van TANG(1938) op bladsye 137-138 na hierdie geval te veralgemeen, waardeur ons vind

(4.166) $\lambda_F^2 = b\sigma_F^{-2} \sum_j d_j^2$. (Vergelyk ook VAN ELTEREN en NOETHER(1959) vir die geval b = 1).

Die asimptotiese relatiewe doeltreffendheid van T_k' met betrekking tot die F-toets is dus

(4.167) $$E_{T_k',F} = \lambda_k^2/\lambda_F^2$$

$$= \frac{12kb}{(kb+1)} \sigma_F^2 \left[\int_{-\infty}^{\infty} f^2(x)\, dx \right]^2 .$$

Opmerkings.

1). Stel b = 1, dan herlei T_k' in (4.156) na

(4.168) $$T_k'' = \frac{12}{mk(k+1)} \sum_j [\sum_i \{r_{ij} - \tfrac{1}{2}(k+1)\}]^2$$

$$= \frac{12}{mk(k+1)} \sum_j [\sum_i r_{ij} - \tfrac{1}{2}m(k+1)]^2 .$$

2). Vir b = 1 is

(4.169) $$E_{T_k'',F} = \frac{12k}{(k+1)} [\sigma_F \int_{-\infty}^{\infty} f^2(x)\, dx]^2$$

wat dieselfde is as die resultaat gevind deur VAN ELTEREN en NOETHER(1959) vergelyking (7).

3). Vir die normaalverdeling is uit (4.167)

(4.170) $$E_{T_k'',F} = \frac{kb}{(kb+1)} . \frac{3}{\pi} .$$

Namate kb toeneem, benader $E_{T_k'',F}$ die waarde $3/\pi = 0.955$ (sien VAN ELTEREN en NOETHER(1959) vergelyking (7N) en die daaropvolgende tabel).

4). Vir die homogene verdeling is uit (4.167)

(4.171) $$E_{T_k'',F} = \frac{kb}{(kb+1)}$$

wat die waarde 1 benader as kb toeneem.

Opmerking.

Dit blyk dat T_k' met $\Psi_i(\delta_{ij\eta}) \equiv r_{ij\eta}$ asimptoties onafhanklik is van verskuiwingseffekte tussen die verskillende rye - sien (4.162). Die toets gebaseer op T_k' kan dus gebruik word om vir verskuiwingseffekte tussen die kolomme te toets onafhanklik van eventuele verskuiwingseffekte tussen rye. Soortgelyk kan rangnommers opnuut binne die kolomme toegeken word en die toets aangepas word om te toets vir verskuiwings tussen die rye (onafhanklik van die voorkoms, al dan nie, van kolom-verskuiwings). Die toetse vir ry- en kolom-effekte moet egter op afsonderlike onafhanklike stelle waarnemings toegepas word.

Slotopmerking.

In bostaande spesiale geval is dit redelik maklik om in te sien dat onder sekere reëlmatigheidsvoorwaardes, waaronder die kontinuïteit van alle F_i, aan (4.143) voldoen word.

HOOFSTUK V.

DIE ANALISE VAN VARIANSIE.

5.1. INLEIDING.

Word in die geval van §4.5.2 rangnommers aan al die N waarnemings gesamentlik toegeken, kry ons 'n twee-faktor variansie-analise. Ons gaan dié geval in hierdie hoofstuk behandel.

5.2. KORT OORSIG OOR REEDS BEKENDE WERK.

MOOD(1950) het 'n verdelingsvrye prosedure behandel waarvolgens vir ry-effekte, kolomeffekte, interaksie en kombinasies daarvan getoets kan word in die tweefaktor rangskikking met b waarnemings per sel. Die toetsingsgroothede is gebaseer op die afwykings van die waarnemings vanaf hulle mediane. Die verskillende toetsingsgroothede word kortliks bespreek (sien MOOD(1950) en TATE en CLELLAND(1957)).

Gestel daar is m rye, k kolomme en b waarnemings per sel. Laat $A = \frac{1}{2}mb$ of $\frac{1}{2}(mb-1)$, naamlik dié een waarvoor A 'n heelgetal is. Laat $\hat{x}_{.j}$ die mediaan van die mb waarnemings in die j^{de} kolom wees en $\hat{x}_{i.}$ die mediaan van die kb waarnemings in die i^{de} ry.

5.2.1. Mediaantoets vir Interaksie.

MOOD(1950) het as toetsingsgrootheid vir interaksie voorgestel

$$\chi_I^2 = P^2 b \sum_i \sum_j \frac{(P_{ij} - P_{i.}P_{.j}/P)^2}{P_{i.}P_{.j}(P_{i.}P_{.j} - b)} \qquad 1)$$

waar die simbole die volgende betekenis het: Word ry- en/of kolommediane van die waarnemings in die tabel afgetrek, kry ons negatiewe waardes (aangedui deur

---

1). In hierdie hoofstuk deurloop i,i' die waardes 1,2,...,m en j,j',j" die waardes 1,2,...,k, tensy anders gespesifiseer.

minustekens), nulle, en positiewe waardes (aangedui deur plustekens). Nou is

P = aantal plustekens plus helfte van die nulle in die tabel,

P_{ij} = aantal plustekens plus helfte van die nulle in die sel (i,j),

$P_{i.}, P_{.j}$ = aantal plustekens plus helfte van die nulle in die i^{de} ry respektiewelik j^{de} kolom.

Hierdie toets vir interaksie word toegepas nadat ry- en kolomeffekte (indien enige) verwyder is deur herhaalde aftrekking. Trek naamlik die kolommediane af van die waarnemings in die onderskeie kolomme. (Daar kan netsowel met rye begin word). Vir die nuwe stelsel word ry-mediane bepaal en van die waarnemings in die onderskeie rye afgetrek. Hierdie prosedure word herhaal totdat die ry- en kolommediane almal prakties nul is.

Volgens MOOD(1950) is χ_I^2 asimptoties χ^2 verdeel met $(m-1)(k-1)$ g.v.v. as $b \to \infty$. (Die benadering is bevredigend as $P_{i.} P_{.j}/P \geq 2$ en $(m-1)(k-1) \geq 2$).

5.2.2. Mediaantoets vir ry- en interaksie-effekte gesamentlik.

Reduseer die waarnemings binne elke kolom om hul kolommediaan sodat die toets onafhanklik is van moontlike kolomeffekte.

Die toetsingsgrootheid

$$\chi_{mI}^2 = \frac{m(mb-1)}{A(mb-A)} \sum_i \sum_j (P_{ij} - \frac{A}{m})^2$$

besit asimptoties 'n χ^2-verdeling met $k(m-1)$ g.v.v. as $b \to \infty$. (Die benadering is goed vir $b \geq 5$ of $mkb \geq 20$).

Soortgelyk vir kolom- en interaksie-effekte gesamentlik.

5.2.3. Mediaantoets vir ry-effekte as interaksie nie betekenisvol is nie.

Reduseer die waarnemings binne elke kolom om hul kolommediaan (om moontlike kolomeffekte uit te skakel). Die toetsingsgrootheid

$$\chi_m^2 = \frac{m(mb-1)}{kA(mb-A)} \sum_i (P_{i.} - \frac{kA}{m})^2$$

is asimptoties χ^2 verdeel met $(m-1)$ g.v.v. as $b \to \infty$.

Soortgelyk vir kolomeffekte.

5.3. 'n NUWE UITBREIDING.

5.3.1. Inleiding.

In hierdie hoofstuk word twee tipes toetse bespreek, naamlik toetse vir homogeniteit van gemiddeldes en homogeniteit van variansies.

In die behandeling van MOOD(1950) hierbo word gewerk met die afwykings van die gegewe waarnemings vanaf hulle mediane.

By toetse vir die homogeniteit van variansies kan dieselfde gedagte verder uitgebrei word deur die waarnemings te reduseer om 'n bepaalde lokaliteitsparameter (rekenkundige gemiddeld, mediaan, modus, ens.), aan die gereduseerde waardes rangnommers toe te ken en dan 'n toetsingsgrootheid te konstrueer as 'n funksie van die rangnommers. Die resultate van hoofstuk II kan vermoedelik na hierdie geval uitgebrei word onderhewig aan bepaalde voorwaardes, bv. simmetrie van die verdeling.

SUKHATME(1958) het in die twee-steekproef geval vir verspreidingsalternatiewe soos volg te werk gegaan: Hy beskou onafhanklike steekproewe $X_1, X_2, \ldots, X_m$ en $Y_1, Y_2, \ldots, Y_n$ uit populasies met verdelingsfunksies $F(x-\xi)$ en $G(x-\eta)$ respektiewelik waar ξ en η onbekende populasieparameters is. SUKHATME(1958) maak skattings

$\hat{\xi}(X_1,X_2,\ldots,X_m)$ en $\hat{\eta}(Y_1,Y_2,\ldots,Y_n)$ van ξ en η respektiewelik. Dan werk hy met 'n toetsingsgrootheid gebaseer op die U-statistiese groothede van HOEFFDING(1948), toegepas op die waarnemings $X_i - \hat{\xi}$ en $Y_j - \hat{\eta}$. SUKHATME toon onder andere aan dat sy toetsingsgrootheid asimptoties normaal verdeel en asimptoties verdelingsvry is.

Dié geval is deur CROUSE(1960) uitgebrei na m steekproewe.

In hierdie hoofstuk gebruik ons egter 'n geheel ander benadering wat, soos aangetoon sal word, 'n groot ooreenkoms toon met die gewone analise van variansie.

5.3.2. Definisies.

Gestel ons het 'n stel waarnemings x_{ijh} verdeel in m rye en k kolomme met b waarnemings per sel waar m en k eindig is. Laat N=mkb. Ons werk onder die nulhipotese H_o dat alle waarnemings x_{ijh} afkomstig is uit dieselfde populasie met verdelingsfunksie F.

Rangskik die waarnemings x_{ijh} gesamentlik volgens grootte in stygende volgorde. Ken aan al die N waardes x_{ijh} rangnommers toe waarby r_{ijh} die rangnommer is van x_{ijh}. Dui die gerangskikte x_{ijh}-waardes aan deur

$$z_1 \leq z_2 \leq \ldots \leq z_N.$$

Ons beskou nou 'n funksie van die rangnommers, naamlik $\Psi_N[r_{ijh}/(N+1)]$.

Indien daar tussen die waardes x_{ijh} knope van lengtes $g_1,g_2,\ldots,g_\lambda$ voorkom, definieer ons (sien hoofstuk II):

$$(5.1)\quad G_\alpha = g_1 + g_2 + \ldots + g_\alpha \quad \text{met } \alpha \in \{1,2,\ldots,\lambda\} \quad \text{en}$$
$$G_o = 0,$$

waarby aangeneem word dat daar λ knope voorkom. Ongelyke waardes word beskou as knope van lengte een.

Laat

$$(5.2)\quad a_p = g_\alpha^{-1} \sum_{\mu=1}^{g_\alpha} \Psi_N[(G_{\alpha-1}+\mu)/(N+1)] \text{ wanneer}$$

$G_{\alpha-1}+1 \leq p \leq G_\alpha$.

Definieer

$$(5.3)\quad a_{ijh} = a_p \text{ indien } x_{ijh} = z_p.$$

Indien alle knope van lengte een is, is

$$a_{ijh} = \Psi_N[r_{ijh}/(N+1)] \text{ vir alle } i,\ j \text{ en } h.$$

Laat

$$(5.4)\quad a_{ij.} = b^{-1}\sum_h a_{ijh} \qquad [2)]$$

$$(5.5)\quad a_{i..} = (kb)^{-1}\sum_j\sum_h a_{ijh}$$

$$(5.6)\quad a_{.j.} = (mb)^{-1}\sum_i\sum_h a_{ijh}$$

$$(5.7)\quad a_{...} = (mkb)^{-1}\sum_i\sum_j\sum_h a_{ijh} = N^{-1}\sum_{p=1}^{N} a_p$$

$$(5.8)\quad \sigma_a^2 = N^{-1}\sum_i\sum_j\sum_h (a_{ijh}-a_{...})^2 = N^{-1}\sum_p (a_p - a_{...})^2 .$$

5.3.3. Die Toetsingsgroothede T_s, T_m, T_k en T_I .

Ons definieer die volgende toetsingsgroothede:

$$(5.9)\quad T_s = (N-1)bN^{-1}\sigma_a^{-2}\sum_i\sum_j (a_{ij.}-a_{...})^2,$$

$$(5.10)\quad T_m = (N-1)kbN^{-1}\sigma_a^{-2}\sum_i (a_{i..}-a_{...})^2,$$

$$(5.11)\quad T_k = (N-1)mbN^{-1}\sigma_a^{-2}\sum_j (a_{.j.}-a_{...})^2 \quad \text{en}$$

$$(5.12)\quad T_I = (N-1)bN^{-1}\sigma_a^{-2}\sum_i\sum_j (a_{ij.}-a_{i..}-a_{.j.}+a_{...})^2 .$$

Lemma 5.1. $T_s = T_m + T_k + T_I$.

Bewys: $\sum_i\sum_j (a_{ij.}-a_{...})^2 =$

$$= \sum_i\sum_j [a_{ij.}-a_{i..}-a_{.j.}+a_{...}+(a_{i..}-a_{...})+(a_{.j.}-a_{...})]^2$$

$$= \sum_i\sum_j (a_{ij.}-a_{i..}-a_{.j.}+a_{...})^2 + k\sum_i (a_{i..}-a_{...})^2 + m\sum_j (a_{.j.}-a_{...})^2 .$$

2). In hierdie hoofstuk deurloop h,h' die waardes 1,2,..,b en p die waardes 1,2,...,N, tensy anders gespesifiseer.

Nou is

$$T_s = (N-1)bN^{-1}\sigma_a^{-2}\sum_i\sum_j(a_{ij.}-a_{...})^2$$

$$= \frac{(N-1)b}{N\sigma_a^2}\sum_i\sum_j(a_{ij.}-a_{i..}-a_{.j.}+a_{...})^2 + \frac{(N-1)bk}{N\sigma_a^2}\sum_i(a_{i..}-a_{...})^2 + $$

$$+ \frac{(N-1)mb}{N\sigma_a^2}\sum_j(a_{.j.}-a_{...})^2$$

$$= T_I + T_m + T_k .$$

5.3.4. Limietverdelings onder H_o van T_s, T_m, T_k en T_I as $b\to\infty$.

STELLING 5.1.

Indien

(5.13) $\max_p (a_p-a_{...})^2 = o(N)$ en

(5.14) (Iedere diskontinuïteit van $F(x)$) $\leq \theta < 1$,

dan is T_s onder H_o asimptoties χ^2 verdeel met (mk-1) grade van vryheid as $N\to\infty$ [3].

Bewys: Definieer

(5.15) $t'_{ij} = \sum_h(a_{ijh} - a_{...})$

$$= b(a_{ij.} - a_{...}).$$

Omdat t'_{ij} analoog is aan t_i van hoofstuk II, volg netsoos in lemmas 2.1, 2.2 en 2.3 dat

(5.16) $E(t'_{ij}) = bN^{-1}\sum_i\sum_j\sum_h(a_{ijh} - a_{...}) = 0,$

(5.17) $\mathrm{var}(t'_{ij}) = b(N-b)(N-1)^{-1}\sigma_a^2$ en

(5.18) $\mathrm{kov}(t'_{ij},t'_{i'j'}) = -\,b^2(N-1)^{-1}\sigma_a^2$ mits $i'\neq i$ en/of $j'\neq j$.

Soos in stelling 2.4 volg dat die variansie-kovariansiematriks van die variante

$$\hat{t}'_{ij} = (N-b)^{\frac{1}{2}}\, t'_{ij}[N\,\mathrm{var}(t'_{ij})]^{-\frac{1}{2}}$$

van rang (mk-1) is.

---

3). Die bewerings $N\to\infty$ en $b\to\infty$ kan as ekwivalent beskou word omdat m en k vas gehou word as $b\to\infty$.

Kies nou willekeurige eindige konstantes c'_{ij} wat nie almal gelyk is nie.

Onder voorwaardes (5.13) en (5.14) geld, soos in stelling 2.3, dat $\sum_i \sum_j c'_{ij} t'_{ij}$ asimptoties normaal verdeel is as $N \to \infty$. Ten slotte volg, soos in stelling 2.5, dat $\sum_i \sum_j (N-b)N^{-1}(t'_{ij})^2[\mathrm{var}(t'_{ij})]^{-1}$ asimptoties, vir $N \to \infty$, 'n χ^2-verdeling met (mk-1) g.v.v. besit.

Hierby is:

$$\sum_i \sum_j (N-b)N^{-1}(t'_{ij})^2[\mathrm{var}(t'_{ij})]^{-1}$$

$$= (N-b)N^{-1} \sum_i \sum_j [b(a_{ij.} - a_{...})]^2 [b(N-b)(N-1)^{-1}\sigma_a^2]^{-1}$$

$$= (N-1)bN^{-1} \sigma_a^{-2} \sum_i \sum_j (a_{ij.} - a_{...})^2$$

$$= T_s \ .$$

Hiermee is die stelling bewys.

Opmerking.

$$(5.19) \quad E(T_s) = E[(N-1)bN^{-1}\sigma_a^{-2} \sum_i \sum_j (a_{ij.} - a_{...})^2]$$

$$= (N-1)bN^{-1}\sigma_a^{-2} \sum_i \sum_j E(a_{ij.} - a_{...})^2$$

$$= (N-1)bN^{-1}\sigma_a^{-2} \sum_i \sum_j E[b^{-1}\sum_h (a_{ijh} - a_{...}) b^{-1} \sum_{h'} (a_{ijh'} - a_{...})]$$

$$= (N-1)N^{-1}b^{-1}\sigma_a^{-2} \sum_i \sum_j E[\sum_h (a_{ijh} - a_{...}) \sum_{h'} (a_{ijh'} - a_{...})]$$

$$= (N-1)N^{-1}b^{-1}\sigma_a^{-2} \sum_i \sum_j \mathrm{var}(t'_{ij}) \qquad \text{omdat } E(t'_{ij}) = 0$$

$$= (N-1)N^{-1}b^{-1}\sigma_a^{-2} \sum_i \sum_j b(N-b)(N-1)^{-1}\sigma_a^2 \qquad \text{uit (5.17)}$$

$$= mk(mkb-b)/N$$

$= (mk-1) =$ aantal g.v.v. van T_s, soos wat die geval behoort te wees.

Definieer vervolgens

$$(5.20) \quad t_{ij} = \sum_h (a_{ijh} - a_{i..} - a_{.j.} + a_{...})$$

$$(5.21) \quad t_{i.} = \sum_j \sum_h (a_{ijh} - a_{...}) \quad \text{en}$$

$$(5.22) \quad t_{.j} = \sum_i \sum_h (a_{ijh} - a_{...}) \ .$$

Lemma 5.2. Die variante t_{ij}, $t_{i.}$ en $t_{.j}$ is, vir enige vaste i en j, onderling ongekorreleerd.

Bewys: $\text{kov}(t_{i.},t_{.j}) = E[\sum_{j'}\sum_{h}(a_{ij'h}-a_{...})\sum_{i'}\sum_{h'}(a_{i'jh'}-a_{...})]$

$$= \sum_{i'}\sum_{j'} E[\sum_{h}(a_{ij'h}-a_{...})\sum_{h'}(a_{i'jh'}-a_{...})]$$

$$= E[\sum_{h}(a_{ijh}-a_{...})\sum_{h'}(a_{ijh'}-a_{...})] +$$

$$+ \sum_{(i,j')\neq(i',j)}\sum\ \sum\sum E[\sum_{h}(a_{ij'h}-a_{...})\sum_{h'}(a_{i'jh'}-a_{...})] \ . \quad 4)$$

Neem die terme afsonderlik.

$$(5.23)\quad E[\sum_{h}(a_{ijh}-a_{...})\sum_{h'}(a_{ijh'}-a_{...})] = E(t_{ij}'^{2})$$

$$= b(N-b)(N-1)^{-1}\sigma_a^2$$

uit (5.17),

$$(5.24)\quad E[\sum_{h}(a_{ij'h}-a_{...})\sum_{h'}(a_{i'jh'}-a_{...})] = -\ b^2(N-1)^{-1}\sigma_a^2 \quad \text{uit}$$

(5.18) waarby $i'\neq i$ en/of $j'\neq j$ in laasgenoemde geval.

Dus $\text{kov}(t_{i.},t_{.j}) = b(N-b)(N-1)^{-1}\sigma_a^2 - (mk-1)b^2(N-1)^{-1}\sigma_a^2$

$$= 0 \quad \text{want } N = mkb.$$

$t_{i.}$ en $t_{.j}$ is dus ongekorreleerd. Dit geld vir enige moontlike waardes van i en j.

Verder is:

$$\text{kov}(t_{i.},t_{ij}) = E[\sum_{j'}\sum_{h}(a_{ij'h}-a_{...})\sum_{h'}(a_{ijh'}-a_{i..}-a_{.j.}+a_{...})]$$

$$= E[\sum_{j'}\sum_{h}(a_{ij'h}-a_{...})\sum_{h'}(a_{ijh'}-a_{...})-\sum_{j'}\sum_{h}(a_{ij'h}-a_{...})\sum_{h'}(a_{i..}-a_{...})$$

$$-\sum_{j'}\sum_{h}(a_{ij'h}-a_{...})\sum_{h'}(a_{.j.}-a_{...})] \ .$$

Neem die terme afsonderlik.

$$E[\sum_{j'}\sum_{h}(a_{ij'h}-a_{...})\sum_{h'}(a_{ijh'}-a_{...})] =$$

$$= E[\sum_{h}(a_{ijh}-a_{...})\sum_{h'}(a_{ijh'}-a_{...})] +$$

$$+ \sum_{j'(\neq j)} E[\sum_{h}(a_{ij'h}-a_{...})\sum_{h'}(a_{ijh'}-a_{...})]$$

$$= b(N-b)(N-1)^{-1}\sigma_a^2 - (k-1)b^2(N-1)^{-1}\sigma_a^2$$

$$= b^2k(m-1)(N-1)^{-1}\sigma_a^2 \ .$$

---

4). Hierdie sommasie beteken $i'\neq i$ en/of $j'\neq j$.

$$E[\sum_{j'}\sum_{h}(a_{ij'h} - a_{...})\sum_{h'}(a_{i..} - a_{...})]$$

$$= b\sum_{j'} E[\sum_{h}(a_{ij'h} - a_{...})(a_{i..} - a_{...})]$$

$$= b\sum_{j'} E[\sum_{h}(a_{ij'h} - a_{...})k^{-1}b^{-1}\sum_{j''}\sum_{h'}(a_{ij''h'} - a_{...})]$$

$$= k^{-1}\sum_{j'}\sum_{j''} E[\sum_{h}(a_{ij'h} - a_{...})\sum_{h'}(a_{ij''h'} - a_{...})]$$

$$= k^{-1}\sum_{j''} E[\sum_{h}(a_{ij''h} - a_{...})\sum_{h'}(a_{ij''h'} - a_{...})] +$$

$$+ k^{-1}\sum_{j'\neq j''}\sum E[\sum_{h}(a_{ij'h} - a_{...})\sum_{h'}(a_{ij''h'} - a_{...})]$$

$$= k^{-1}kb(N-b)(N-1)^{-1}\sigma_a^2 - k^{-1}k(k-1)b^2(N-1)^{-1}\sigma_a^2$$

$$= b^2k(m-1)(N-1)^{-1}\sigma_a^2 .$$

$$E[\sum_{j'}\sum_{h}(a_{ij'h} - a_{...})\sum_{h'}(a_{.j.} - a_{...})]$$

$$= b\sum_{j'} E[\sum_{h}(a_{ij'h} - a_{...})m^{-1}b^{-1}\sum_{i'}\sum_{h'}(a_{i'jh'} - a_{...})]$$

$$= m^{-1}\sum_{j'}\sum_{i'} E[\sum_{h}(a_{ij'h} - a_{...})\sum_{h'}(a_{i'jh'} - a_{...})]$$

$= 0$ soos in $\text{kov}(t_{i.},t_{.j})$.

Ons kry dus:

$$\text{kov}(t_{i.},t_{ij}) = b^2k(m-1)(N-1)^{-1}\sigma_a^2 - b^2k(m-1)(N-1)^{-1}\sigma_a^2$$

$$= 0.$$

$t_{i.}$ en t_{ij} is dus ongekorreleerd.

Op analoë wyse as in die vorige geval, is $\text{kov}(t_{.j},t_{ij}) = 0$, d.w.s. $t_{.j}$ en t_{ij} is ongekorreleerd.

Hiermee is die lemma bewys.

Lemma 5.3. Die (mk+m+k) variante t''_{ij}, $t'_{i.}$ en $t'_{.j}$ vir 5) $i=1,2,...,m$; $j=1,2,...,k$ is gesamentlik asimptoties normaal verdeel onder die voorwaardes (5.13) en (5.14) indien $N\to\infty$ (d.i. $b\to\infty$).

Bewys: Ons maak gebruik van stelling 2.1, naamlik dat $S_N = \sum_p b_{Np}d_p$ asimptoties normaal verdeel is as $N\to\infty$ indien die voorwaarde

---

5). Hier is $t''_{ij} = b^{-\frac{1}{2}}t_{ij}$, $t'_{i.} = b^{-\frac{1}{2}}t_{i.}$ en $t'_{.j} = b^{-\frac{1}{2}}t_{.j}$.

$$(2.29)\quad \lim_{N\to\infty} \frac{N \operatorname*{maks}_p (a_{Np}-\bar{a}_N)^2 \operatorname*{maks}_p (b_{Np}-\bar{b}_N)^2}{\sum_p (a_{Np}-\bar{a}_N)^2 \sum_p (b_{Np}-\bar{b}_N)^2} = 0$$

bevredig word (sien stelling 2.1).

Stel $a_{Np} = a_p$ vir $p=1,2,\ldots,N$ en

$$b^{\frac{1}{2}} b_{Np} = (mk)^{-1}[mkc_{i'j'} - m\sum_j c_{i'j} - k\sum_i c_{ij'} + \sum_i\sum_j c_{ij}] +$$
$$+ km^{-1}(mc_{i'.} - \sum_i c_{i.}) + mk^{-1}(kc_{.j'} - \sum_j c_{.j}) \quad \text{vir alle}$$

p waarvoor z_p in die sel (i',j') lê. Hier word ook veronderstel dat die c_{ij} 's, $c_{i.}$'s en $c_{.j}$'s $(mk+m+k)$ willekeurige eindige konstantes is. (Die $c_{i.}$'s en $c_{.j}$'s is willekeurige konstantes onafhanklik van die c_{ij} 's).

Nou is:

$$(5.25)\quad b^{\frac{1}{2}} S_N = b^{\frac{1}{2}} \sum_p b_{Np} d_p$$

$$= \sum_{i'}\sum_{j'}\sum_h [(mk)^{-1}(mkc_{i'j'} - m\sum_j c_{i'j} - k\sum_i c_{ij'} + \sum_i\sum_j c_{ij})$$
$$+ km^{-1}(mc_{i'.} - \sum_i c_{i.}) + mk^{-1}(kc_{.j'} - \sum_j c_{.j})]\, a_{i'j'h}$$

$$= \sum_{i'}\sum_{j'}\sum_h c_{i'j'} a_{i'j'h} - k^{-1}\sum_{i'}\sum_{j'}\sum_h\sum_j c_{i'j} a_{i'j'h} - m^{-1}\sum_{i'}\sum_{j'}\sum_h\sum_i c_{ij'} a_{i'j'h} +$$
$$+ (mk)^{-1}\sum_{i'}\sum_{j'}\sum_h\sum_i\sum_j c_{ij} a_{i'j'h} + km^{-1}\sum_{i'}\sum_{j'}\sum_h (mc_{i'.} - \sum_i c_{i.}) a_{i'j'h}$$
$$+ mk^{-1}\sum_{i'}\sum_{j'}\sum_h (kc_{.j'} - \sum_j c_{.j}) a_{i'j'h}$$

$$= \sum_i\sum_j\sum_h c_{ij} a_{ijh} - k^{-1}\sum_i\sum_j\sum_h\sum_{j'} c_{ij} a_{ij'h} - m^{-1}\sum_i\sum_j\sum_h\sum_{i'} c_{ij} a_{i'jh} +$$
$$+ (mk)^{-1}\sum_i\sum_j\sum_h\sum_{i'}\sum_{j'} c_{ij} a_{i'j'h} + k\sum_{i'} c_{i'.} \sum_{j'}\sum_h a_{i'j'h} -$$
$$- km^{-1}\sum_i c_{i.} \sum_{i'}\sum_{j'}\sum_h a_{i'j'h} + m\sum_{i'}\sum_{j'}\sum_h c_{.j'} a_{i'j'h} -$$
$$- mk^{-1}\sum_{i'}\sum_{j'}\sum_h\sum_j c_{.j} a_{i'j'h}$$

$$= \sum_i\sum_j c_{ij}\sum_h a_{ijh} - b\sum_i\sum_j c_{ij} a_{i..} - b\sum_i\sum_j c_{ij} a_{.j.} + b\sum_i\sum_j c_{ij} a_{...} +$$
$$+ k\sum_i c_{i.} \sum_{j'}\sum_h a_{ij'h} - bk^2\sum_i c_{i.} a_{...} + m\sum_j c_{.j}\sum_{i'}\sum_h a_{i'jh} -$$
$$- bm^2\sum_j c_{.j} a_{...}$$

$$= \sum_i \sum_j c_{ij} [\sum_h (a_{ijh} - a_{i..} - a_{.j.} + a_{...})] + k \sum_i c_{i.} \sum_{j'} \sum_h (a_{ij'h} - a_{...})$$

$$+ m \sum_j c_{.j} \sum_{i'} \sum_h (a_{i'jh} - a_{...})$$

$$= \sum_i \sum_j c_{ij} t_{ij} + k \sum_i c_{i.} t_{i.} + m \sum_j c_{.j} t_{.j} \quad \text{sodat}$$

$$S_N = \sum_i \sum_j [c_{ij} t''_{ij} + c_{i.} t'_{i.} + c_{.j} t'_{.j}].$$

Die variante t''_{ij}, $t'_{i.}$ en $t'_{.j}$ vir $i=1,2,\ldots,m$; $j=1,2,\ldots,k$ is dus gesamentlik asimptoties normaal verdeel onder voorwaarde (2.29) as $N \to \infty$.[6] Ons toon vervolgens aan dat (2.29) bevredig word.

Sonder verlies aan algemeenheid kan veronderstel word dat die c_{ij} 's nie almal gelyk is nie, en netso vir die $c_{i.}$'s en die $c_{.j}$'s.

Omdat bogenoemde konstantes almal eindig is, is

$$b \cdot \max_p (b_{Np} - \bar{b}_N)^2 \leq K < \infty.$$

Verder is $\lim_{N\to\infty} b N^{-1} \sum_p (b_{Np} - \bar{b}_N)^2 \geq \varepsilon > 0$ omdat die getalle $b^{\frac{1}{2}} b_{Np}$ vir die verskillende selle nie almal gelyk kan wees as alle c's groepsgewyse nie almal gelyk is nie.

Onder voorwaarde (5.14) kry ons dan as voldoende voorwaarde vir die geldigheid van (2.29) die voorwaarde

(5.13) $\max_p (a_p - a_{...})^2 = o(N)$.

Lemma 5.4. Onder voorwaardes (5.13) en (5.14) besit elk van die drie groepe variante t''_{ij}, $t'_{i.}$ en $t'_{.j}$ $(i=1,2,..,m;$ $j=1,2,\ldots,k)$ asimptoties, vir $N\to\infty$, gesamentlike normaalverdelings wat onderling onafhanklik is.

Bewys: Stel $\vec{M} = \{m_1, m_2, \ldots, m_n\}$ en $\vec{t} = \{t_1, t_2, \ldots, t_n\}$ met $n = mk+m+k$, $m_g = E(t_g)$, $g=1,2,\ldots,n$, en

$$(5.26) \quad t_g = \begin{cases} t''_{ij}, \ 1 \leq i \leq m, \ 1 \leq j \leq k & \text{vir } 1 \leq g \leq mk \\ t'_{i.}, \ 1 \leq i \leq m & \text{vir } mk+1 \leq g \leq mk+m \\ t'_{.j}, \ 1 \leq j \leq k & \text{vir } mk+m+1 \leq g \leq n \end{cases}$$

sodanig dat $t_1 = t''_{11}$, $t_2 = t''_{12}$, .. $t_k = t''_{1k}$, $t_{k+1} = t''_{21}$, $t_{k+2} = t''_{22}$, ens.

6). Die bewering geld eintlik alleen as aan 'n ekstra voorwaarde voldoen word. Sien bylaag C.

Omdat die variante t''_{ij}, $t'_{i.}$ en $t'_{.j}$ vir alle i en j gesamentlik asimptoties 'n normaalverdeling besit (lemma 5.3), word die karakteristieke funksie van hierdie gesamentlike verdeling gegee deur (sien CRAMéR(1946) p. 310):

$$(5.27)\quad \phi(\vec{u}) = \text{eksp}\Big(i \sum_{g=1}^{n} m_g u_g - \tfrac{1}{2} \sum_{f=1}^{n} \sum_{g=1}^{n} \lambda_{fg} u_f u_g\Big)$$

waar λ_{fg} die kovariansie tussen t_f en t_g aandui en $\vec{u} = \{u_1, u_2, \ldots, u_n\}$.

Maar

$$\sum\sum_{f,g=1}^{n} \lambda_{fg} u_f u_g = \sum_{f=1}^{n} \sum_{g=1}^{mk} \lambda_{fg} u_f u_g + \sum_{f=1}^{n} \sum_{g=mk+1}^{mk+m} \lambda_{fg} u_f u_g +$$

$$+ \sum_{f=1}^{n} \sum_{g=mk+m+1}^{n} \lambda_{fg} u_f u_g$$

$$= \sum\sum_{f,g=1}^{mk} \lambda_{fg} u_f u_g + \sum\sum_{f,g=mk+1}^{mk+m} \lambda_{fg} u_f u_g + \sum\sum_{f,g=mk+m+1}^{n} \lambda_{fg} u_f u_g$$

omdat die variante t''_{ij}, $t'_{i.}$ en $t'_{.j}$ onderling ongekorreleerd is (lemma 5.2) en dus

$\lambda_{fg} = 0$ is vir:

i) $1 \leq f \leq mk$ en $mk+1 \leq g \leq n$

ii) $mk+1 \leq f \leq mk+m$ en $1 \leq g \leq mk$ of $mk+m+1 \leq g \leq n$

iii) $mk+m+1 \leq f \leq n$ en $1 \leq g \leq mk+m$.

Ons kry dus

$$(5.28)\quad \phi(\vec{u}) = \text{eksp}\Big(i \sum_{g=1}^{n} m_g u_g - \tfrac{1}{2} \sum\sum_{f,g=1}^{mk} \lambda_{fg} u_f u_g - \tfrac{1}{2} \sum\sum_{f,g=mk+1}^{mk+m} \lambda_{fg} u_f u_g$$

$$- \tfrac{1}{2} \sum\sum_{f,g=mk+m+1}^{n} \lambda_{fg} u_f u_g\Big)$$

$$= \text{eksp}\Big(i \sum_{g=1}^{mk} m_g u_g - \tfrac{1}{2} \sum\sum_{f,g=1}^{mk} \lambda_{fg} u_f u_g\Big) \cdot$$

$$\cdot\ \text{eksp}\Big(i \sum_{g=mk+1}^{mk+m} m_g u_g - \tfrac{1}{2} \sum\sum_{f,g=mk+1}^{mk+m} \lambda_{fg} u_f u_g\Big) \cdot$$

$$\cdot\ \text{eksp}\Big(i \sum_{g=mk+m+1}^{n} m_g u_g - \tfrac{1}{2} \sum\sum_{f,g=mk+m+1}^{n} \lambda_{fg} u_f u_g\Big)$$

$= \phi_1(\vec{u}_1)\phi_2(\vec{u}_2)\phi_3(\vec{u}_3)$, die produk van drie

karakteristieke funksies wat elkeen die karakteristieke funksie is van 'n normaal-verdeelde groep variante. (Hierby is $\vec{u}_1 = \{u_1, u_2, \ldots, u_{mk}\}$, $\vec{u}_2 = \{u_{mk+1}, \ldots, u_{mk+m}\}$ en $\vec{u}_3 = \{u_{mk+m+1}, \ldots, u_n\}$). Verder is
$\phi_1(\vec{u}_1)$ die karakteristieke funksie van die variante t''_{ij},
$\phi_2(\vec{u}_2)$ die karakteristieke funksie van die variante $t'_{i.}$ en
$\phi_3(\vec{u}_3)$ die karakteristieke funksie van die variante $t'_{.j}$.

Die gesamentlike verdeling van die variante t''_{ij} is asimptoties normaal (CRAMéR(1946) p. 310). Soortgelyk vir die gesamentlike verdelings van die variante $t'_{i.}$ en $t'_{.j}$. Hierdie drie normaalverdelings is verder ook onderling onafhanklik omdat die karakteristieke funksie van die gesamentlike verdeling geskryf kan word as die produk van die afsonderlike karakteristieke funksies en gevolglik is die drie groepe variante onderling onafhanklik.

Opmerking.

Uit die voorgaande bewys volg dat enige funksie van die variante t''_{ij} asimptoties onafhanklik is van enige funksie van die variante $t'_{i.}$ en van enige funksie van die variante $t'_{.j}$ (vergelyk CRAMéR(1962) p. 16 stelling 3).

Definieer nou:

(5.29) $\hat{t}_{ij} = (N-1)^{\frac{1}{2}} t_{ij} N^{-\frac{1}{2}} \sigma_a^{-1}$

(5.30) $\hat{t}_{i.} = (N-kb)^{\frac{1}{2}} t_{i.} [N \text{ var}(t_{i.})]^{-\frac{1}{2}}$ en

(5.31) $\hat{t}_{.j} = (N-mb)^{\frac{1}{2}} t_{.j} [N \text{ var}(t_{.j})]^{-\frac{1}{2}}$.

Vir $t_{i.} = \sum_j \sum_h (a_{ijh} - a_{...})$ kan soos in lemmas 2.2 en 2.3 bewys word dat

(5.32) $\text{var}(t_{i.}) = kb(N-kb)(N-1)^{-1} \sigma_a^2$ en

(5.33) $\text{kov}(t_{i.}, t_{i'.}) = -k^2 b^2 (N-1)^{-1} \sigma_a^2$ vir $i' \neq i$.

Vir $t_{.j} = \sum_i \sum_h (a_{ijh} - a_{...})$ is

(5.34) $\text{var}(t_{.j}) = mb(N-mb)(N-1)^{-1} \sigma_a^2$ en

(5.35) $\text{kov}(t_{.j}, t_{.j'}) = -m^2 b^2 (N-1)^{-1} \sigma_a^2$ vir $j' \neq j$.

STELLING 5.2.

Onder voorwaardes (5.13) en (5.14) besit T_m onder H_o asimptoties, vir $N\to\infty$, 'n χ^2-verdeling met (m-1) grade van vryheid.

Bewys: Omdat die gesamentlike verdeling van die variante $t'_{i.}$ asimptoties, vir $N\to\infty$, normaal is (lemma 5.4), is die gesamentlike verdeling van die variante $\hat{t}_{i.}$ ook asimptoties normaal.

Met behulp van vergelykings (5.30), (5.32) en (5.33) volg:

$$\text{var}(\hat{t}_{i.}) = 1 - m^{-1} \quad \text{en}$$
$$\text{kov}(\hat{t}_{i.},\hat{t}_{i'.}) = - m^{-1} .$$

Die variante $\hat{t}_{i.}$ besit dus gesamentlik asimptoties 'n normaalverdeling as $N\to\infty$ met momentematriks

$$M_6 = \begin{Bmatrix} 1 - m^{-1}, & - m^{-1}, & \dots, & - m^{-1} \\ - m^{-1}, & 1 - m^{-1}, & \dots, & - m^{-1} \\ \cdots & \cdots & \cdots & \cdots \\ - m^{-1}, & - m^{-1}, & \dots, & 1 - m^{-1} \end{Bmatrix}$$

$$= I - \vec{m}\vec{m}'$$

waar I die eenheidsmatriks is, $\vec{m}'$ die ryvektor $\{m^{-\frac{1}{2}},m^{-\frac{1}{2}},\dots,m^{-\frac{1}{2}}\}$ en $\vec{m}$ die ooreenkomstige kolomvektor.

Uit lemma 1.2 volg dat M_6 van rang (m-1) is omdat $\sum\limits_i m^{-1} = 1$. Uit stelling 1.1 opmerking 2 (vergelyk ook CRAMÉR(1946) p. 419) volg dan dat $\sum\limits_i \hat{t}^2_{i.}$ asimptoties, vir $N\to\infty$, χ^2 verdeel is met (m-1) g.v.v.

Hierby is

$$\sum_i \hat{t}^2_{i.} = (N-kb)N^{-1}\sum_i t^2_{i.}[\text{var}(t_{i.})]^{-1} \qquad \text{uit (5.30)}$$

$$= (N-1)[Nkb\sigma_a^2]^{-1}\sum_i[\sum_j\sum_h(a_{ijh} - a_{...})]^2 \quad \text{uit (5.21) en (5.32)}$$

$$= (N-1)kbN^{-1}\sigma_a^{-2}\sum_i(a_{i..} - a_{...})^2$$

$$= T_m \quad \text{uit (5.10)} .$$

Hiermee is die stelling bewys.

Opmerking.

$$(5.36)\quad E(T_m) = E[(N-1)kbN^{-1}\sigma_a^{-2}\sum_i (a_{i..} - a_{...})^2]$$

$$= (N-1)kbN^{-1}\sigma_a^{-2}\sum_i E(a_{i..} - a_{...})^2$$

$$= \frac{(N-1)kb}{N\sigma_a^2}\sum_i E\left[\frac{1}{kb}\sum_j\sum_h (a_{ijh} - a_{...})\frac{1}{kb}\sum_{j'}\sum_{h'}(a_{ij'h'} - a_{...})\right]$$

$$= \frac{(N-1)}{kbN\sigma_a^2}\sum_i\Big\{\sum_j E[\sum_h (a_{ijh} - a_{...})\sum_{h'}(a_{ijh'} - a_{...})] +$$

$$+ \sum_{j\neq j'}\sum E[\sum_h (a_{ijh} - a_{...})\sum_{h'}(a_{ij'h'} - a_{...})]\Big\}$$

$$= \frac{(N-1)}{kbN\sigma_a^2}\sum_i [\sum_j b(N-b)(N-1)^{-1}\sigma_a^2 - \sum_{j\neq j'}\sum b^2(N-1)^{-1}\sigma_a^2] \text{ uit (5.23) en (5.24)}$$

$$= mk^{-1}N^{-1}[k(N-b) - k(k-1)b]$$

$$= (m-1) \quad \text{omdat } N = mkb$$

= aantal g.v.v. van T_m, in ooreenstemming met die teorie van die χ^2-verdeling.

STELLING 5.3.

Onder voorwaardes (5.13) en (5.14) besit T_k onder H_0 asimptoties, vir $N\to\infty$, 'n χ^2-verdeling met $(k-1)$ grade van vryheid.

Bewys: Op analoë wyse as in stelling 5.2 kan met behulp van lemmas 1.2 en 5.4 en stelling 1.1 opmerking 2 bewys word dat $\sum_j \hat{t}^2_{.j}$ asimptoties, vir $N\to\infty$, 'n χ^2-verdeling besit met $(k-1)$ g.v.v.

Hier is

$$\sum_j \hat{t}^2_{.j} = (N-mb)N^{-1}\sum_j t^2_{.j}[\mathrm{var}(t_{.j})]^{-1}$$

$$= (N-1)[Nmb\sigma_a^2]^{-1}\sum_j [\sum_i\sum_h (a_{ijh} - a_{...})]^2 \quad \text{uit (5.22) en (5.34)}$$

$$= (N-1)mbN^{-1}\sigma_a^{-2}\sum_j (a_{.j.} - a_{...})^2$$

$$= T_k \quad \text{uit (5.11).}$$

Hiermee is die stelling bewys.

Opmerking.

Soos in die opmerking na stelling 5.2 kan bewys word dat

(5.37) $E(T_k) = (k-1) =$ aantal g.v.v. van T_k .

STELLING 5.4.

Onder voorwaardes (5.13) en (5.14) besit T_I onder H_o asimptoties 'n χ^2-verdeling met $(m-1)(k-1)$ grade van vryheid as $N \to \infty$.

Bewys: T_s, T_m en T_k besit asimptoties, vir $N \to \infty$, χ^2-verdelings met respektiewelik $(mk-1)$, $(m-1)$ en $(k-1)$ grade van vryheid (stellings 5.1, 5.2 en 5.3).

Verder is $T_s = T_m + T_k + T_I$ (lemma 5.1) waarby T_m, T_k en T_I onderling onafhanklik is (lemma 5.4).

Die aantal grade van vryheid van T_s is dus hoogstens gelyk aan die som van die aantal grade van vryheid van T_m, T_k en T_I (SCHEFFé(1959) p. 422).

$(mk-1) \leq (m-1) + (k-1) + a$ waar a die rang van die momentematriks van die variante t_{ij} voorstel.

Dus $a \geq (m-1)(k-1)$.

Beskou

(5.20) $t_{ij} = \sum_h (a_{ijh} - a_{i..} - a_{.j.} + a_{...})$

$= b(a_{ij.} - a_{i..} - a_{.j.} + a_{...})$.

Daar is km variante t_{ij} waartussen daar minstens $(k+m-1)$ onafhanklike lineêre verbande bestaan, want:

(5.38) $\sum_i \sum_j t_{ij} = 0$ lê een lineêre verband tussen die t_{ij}'s vas.

Verder is

(5.39) $\sum_j t_{ij} = 0$ vir $i=1,2,\ldots,m$.

Hierdeur word $(m-1)$ verdere lineêre verbande tussen die t_{ij}'s vasgelê. (Indien $\sum_j t_{ij} = 0$ vir alle waardes van i behalwe één, dan volg die m^{de} betrekking uit (5.38)).

Soortgelyk lê

(5.40) $\sum_{i} t_{ij} = 0$ vir $j=1,2,\ldots,k$

nog (k-1) onafhanklike lineêre verbande vas. Daar bestaan dus minstens (m+k-1) onafhanklike lineêre betrekkings tussen die t_{ij}'s, sodat die rang a van die momentematriks van die t_{ij}'s hoogstens gelyk is aan

$$mk - (m+k-1) = (m-1)(k-1).$$ (Sien CRAMÉR(1946) pp. 110 en 312).

Maar $a \geq (m-1)(k-1)$. Dus

(5.41) $a = (m-1)(k-1)$.

Omdat die variante t''_{ij} asimptoties, vir $N \to \infty$, gesamentlik 'n normaalverdeling besit met momentematriks van rang $a = (m-1)(k-1)$, volg uit stellings 5.1, 5.2 en 5.3 en lemma 5.4 dat T_I asimptoties 'n χ^2-verdeling met (m-1)(k-1) g.v.v. besit. (Let op dat

(5.42) $T_I = T_s - T_k - T_m$ uit lemma 5.1, sodat

$$\begin{aligned} E(T_I) &= E(T_s) - E(T_k) - E(T_m) \\ &= (mk-1) - (k-1) - (m-1) \\ &= (m-1)(k-1) \ ). \end{aligned}$$

Hiermee is die stelling bewys.

5.3.5. Die Betekenis van T_s, T_m, T_k en T_I.

T_s kan beskou word as 'n toets vir „sel-effekte" omdat dit meet in watter mate die selgemiddeldes $a_{ij.}$ van die totaalgemiddeld $a_{...}$ verskil.

Die toetsingsgrootheid T_m kan gebruik word as 'n toets in hoeverre die ry-gemiddeldes $a_{i..}$ van die totaalgemiddeld $a_{...}$ verskil. T_m meet dus die sogenaamde „ry-effekte".

Analoog aan T_m meet T_k die „kolom-effekte".

Elke selgemiddeld $a_{ij.}$ is onderhewig aan i) ry-effekte; ii) kolom-effekte; iii) ry-kolom interaksie.

In die toetsingsgrootheid

$$(5.43)\quad T_I = (N-1)bN^{-1}\sigma_a^{-2}\sum_i\sum_j(a_{ij.} - a_{i..} - a_{.j.} + a_{...})^2$$

$$= (N-1)bN^{-1}\sigma_a^{-2}\sum_i\sum_j[(a_{ij.}-a_{...})-(a_{i..} - a_{...})-(a_{.j.} - a_{...})]^2$$

word getrag om ry- en kolomeffekte uit te skakel deur aftrekking van $a_{i..}$ en $a_{.j.}$ sodat T_I beskou kan word as 'n maat van die interaksie tussen rye en kolomme. Skematies kry ons:

TABEL 5.1.

Opsomming van Toetsingsgroothede.

Variasiebron	Toetsingsgrootheid	G.v.V.
Ry-effekte	$T_m = \frac{(N-1)kb}{N\sigma_a^2}\sum_i(a_{i..} - a_{...})^2$	$(m-1)$
Kolomeffekte	$T_k = \frac{(N-1)mb}{N\sigma_a^2}\sum_j(a_{.j.} - a_{...})^2$	$(k-1)$
Interaksie	$T_I = \frac{(N-1)b}{N\sigma_a^2}\sum_i\sum_j(a_{ij.}-a_{i..}-a_{.j.}+a_{...})^2$	$(m-1)(k-1)$
Sel-effekte	$T_s = \frac{(N-1)b}{N\sigma_a^2}\sum_i\sum_j(a_{ij.} - a_{...})^2$	$(mk-1)$

5.3.6. Spesiale gevalle van T_s, T_m, T_k en T_I.

Deur $\Psi_N(\delta)$ gelyk te stel aan spesifieke funksies, kan verskillende spesiale gevalle van die toetsingsgroothede verkry word (sien byvoorbeeld §2.5). Ons bespreek slegs twee gevalle.

Eenvoudigheidshalwe neem ons aan dat in geval van knope daar deur die lotingsmeganisme rangnommers toegeken word, sodat ons die voorkoms van knope verder kan ignoreer.

5.3.6.1. $\Psi_N(\delta_{ijh}) \equiv \delta_{ijh}$ (sien §2.5.2.1).

$$(5.44)\quad T_m = (N-1)kbN^{-1}\sigma_a^{-2}\sum_i(a_{i..} - a_{...})^2 \qquad \text{uit } (5.10)$$

$$= \frac{12}{Nkb(N+1)}\sum_i[\sum_j\sum_h r_{ijh} - \tfrac{1}{2}kb(N+1)]^2 .$$

Netso is

$$(5.45)\quad T_k = \frac{12}{Nmb(N+1)} \sum_j [\sum_i \sum_h r_{ijh} - \tfrac{1}{2}mb(N+1)]^2 \qquad \text{en}$$

$$(5.46)\quad T_I = \frac{12}{N(N+1)bk^2m^2} \sum_i \sum_j [mk \sum_h r_{ijh} - m \sum_j \sum_h r_{ijh} - k \sum_i \sum_h r_{ijh} + \\ + \tfrac{1}{2}mkb(N+1)]^2 .$$

5.3.6.2. $\Psi_N(\delta_{ijh}) \equiv Q(\delta_{ijh})$ (sien §2.5.2.3).

In hierdie geval is:

$$(5.47)\quad T_m = (N-1) \sum_i [\sum_j \sum_h Q(\delta_{ijh})]^2 / [kb \sum_p Q^2(\delta_p)] ,$$

$$(5.48)\quad T_k = (N-1) \sum_j [\sum_i \sum_h Q(\delta_{ijh})]^2 / [mb \sum_p Q^2(\delta_p)] \qquad \text{en}$$

$$(5.49)\quad T_I = (N-1)[bk^2m^2 \sum_p Q^2(\delta_p)]^{-1} \sum_i \sum_j [mk \sum_h Q(\delta_{ijh}) - \\ -m \sum_j \sum_h Q(\delta_{ijh}) - k \sum_i \sum_h Q(\delta_{ijh})]^2 .$$

Soortgelyk kan daar nog ander toetsingsgroothede gekonstrueer word (sien vorige hoofstukke se spesiale gevalle van $\Psi_N(\delta)$).

5.4. VERGELYKING TUSSEN DIE PARAMETRIESE EN NIE-PARAMETRIESE TOETSE.

By die gewone tweefaktor analise van variansie vir homogeniteit van gemiddeldes met b waarnemings per sel, het ons die volgende skema:

TABEL 5.2.

Variansietabel.

Variasiebron	Som van Vierkante	G.v.V
Ry-effekte	$q_1 = kb \sum_i (x_{i..} - x_{...})^2$	$(m-1)$
Kolom-effekte	$q_2 = mb \sum_j (x_{.j.} - x_{...})^2$	$(k-1)$
Interaksie	$q_3 = b \sum_i \sum_j (x_{ij.} - x_{i..} - x_{.j.} + x_{...})^2$	$(m-1)(k-1)$
Subtotaal	$Q_s = b \sum_i \sum_j (x_{ij.} - x_{...})^2$	$(mk-1)$
Residu	$q_4 = \sum_i \sum_j \sum_h (x_{ijh} - x_{ij.})^2$	$mk(b-1)$
Totaal	$Q = \sum_i \sum_j \sum_h (x_{ijh} - x_{...})^2$	$(mkb-1)$

Hierby stem q_1 ooreen met T_m (sien tabel 5.1) in soverre dat albei ry-effekte meet, q_2 met T_k, q_3 met T_I en Q_s met T_s. Let op dat

$$T_Q \overset{\text{def}}{=} (N-1)N^{-1}\,\sigma_a^{-2} \sum_i \sum_j \sum_h (a_{ijh} - a_{\dots})^2$$

$$= (N-1)\,\sigma_a^{-2} \,.\, \sigma_a^2 \qquad \text{uit (5.8)}$$

$$= (N-1)$$

= 'n bepaalde waarde vir vaste m, k en b wat dus geen waarskynlikheidsverdeling besit nie. Die rede hiervoor is dat $\sum_i \sum_j \sum_h a_{ijh} \equiv$ konstant.

Vir ry-effekte toets ons in die geval van tabel 5.2 met

(5.50) $F_1 = mk(b-1)q_1/[(m-1)q_4]$.

F_1 besit 'n F-verdeling met (m-1) en mk(b-1) g.v.v.

As $b \to \infty$, gaan die verdeling van $(m-1)F_1$ oor in 'n χ^2-verdeling met (m-1) g.v.v. (KENNEY en KEEPING(1951) vol. II p. 183) en kry ons presies dieselfde toestand as in die verdelingsvrye geval hierbo waar ons vir ry-effekte toets met T_m wat asimptoties χ^2 verdeel is met (m-1) g.v.v. as $b \to \infty$.

Soortgelyk toets ons vir kolom-effekte met

(5.51) $F_2 = mk(b-1)q_2/[(k-1)q_4]$

waarby $(k-1)F_2$ asimptoties, vir $b \to \infty$, 'n χ^2-verdeling met (k-1) g.v.v. besit (vergelyk T_k).

Vir interaksie toets ons met

(5.52) $F_3 = mk(b-1)q_3/[(m-1)(k-1)q_4]$

waarby $(m-1)(k-1)F_3$ asimptoties, vir $b \to \infty$, 'n χ^2-verdeling besit met (m-1)(k-1) g.v.v. (vergelyk T_I).

Asimptoties, as $b \to \infty$, toon ons toetse en die gewone variansie-analise toetse 'n groot ooreenkoms. Hierdie ooreenkoms blyk verder duidelik indien ons die verskillende toetsingsgroothede se formules met mekaar vergelyk. Byvoorbeeld vir ry-effekte vergelyk ons $(m-1)F_1$ met T_m.

$$(5.53)\quad (m-1)F_1 = mk(b-1)q_1/q_4 \quad \text{uit } (5.50)$$
$$= mk(b-1)kb\sum_i(x_{i..}-x_{...})^2/\sum_i\sum_j\sum_h(x_{ijh}-x_{ij.})^2$$
$$= (1-\tfrac{1}{b})bk\sum_i(x_{i..}-x_{...})^2/N^{-1}\sum_i\sum_j\sum_h(x_{ijh}-x_{ij.})^2 .$$

Uit (5.10) volg:

$$(5.54)\quad T_m = (1-\tfrac{1}{N})kb\sum_i(a_{i..}-a_{...})^2/\sigma_a^2 .$$

Stel

$$(5.55)\quad \sigma_{ij}^2 = b^{-1}\sum_h(x_{ijh}-x_{ij.})^2 \quad \text{en}$$

$$(5.56)\quad \sigma_N^2 = (mk)^{-1}\sum_i\sum_j\sigma_{ij}^2 .$$

Dan is

$$(5.57)\quad (m-1)F_1 = (1-\tfrac{1}{b})\, kb\sum_i(x_{i..}-x_{...})^2/\sigma_N^2 .$$

Neem ons byvoorbeeld $\Psi_N(\delta) = Q(\delta) \equiv \Phi^{-1}(\delta)$, dan sal σ_a^2 asimptoties gelyk wees aan 1 (STOKER(1955) p. 57).

Die toepassing van die F-toets is gebaseer op die veronderstelling dat die x 'e normaalvariante is. Gevolglik is σ_N^2 asimptoties in waarskynlikheid gelyk aan σ^2, die variansie van die normaalvariante. Dit skyn voor die hand liggend te wees dat T_m en $(m-1)F_1$ asimptoties ekwivalent is onder H_0 met F die normaal(0,1) v.f.

'n Belangrike vraag is: hoe groot moet b wees alvorens die T-toetse (d.w.s. T_s, T_m, T_k en T_I) toegepas kan word. Hierdie vraag is moeilik te beantwoord. Vir die geval van 2 rye, 2 kolomme en 10 waarnemings per sel met $\Psi_N(\delta)=Q(\delta)$ is vir 'n aantal gevalle die T- en F-toetse se oorskrydingswaarskynlikhede bereken waarby die waarnemings uit WOLD, H(1954): „Tracts for Computers: Random normal deviates" getrek is. Die toepassing van die F-toetse is alleen geregverdig indien ons te doen het met normaal-variante met gelyke variansies.

Die berekende oorskrydingswaarskynlikhede van die T- en F-toetse word in tabelvorm saamgevat waarby die frekwensies vir die toetse vir ry-effekte deur {·}, vir kolomeffekte deur (·) en vir interaksie deur [·] in elke sel aangedui word.

TABEL 5.3.

Oorskrydingswaarskynlikhede van die T- en F-toetse.

		Oorskrydingswaarskynlikhede						
	F \ T	0 - 0.025	0.025- 0.05	0.05- 0.10	0.10- 0.20	0.20- 0.40	0.40- 0.60	0.60- 0.80
Oorskrydingswaarskynlikhede	0 - 0.025	{2},[1]						
	0.025- 0.05		{2},[2]		[1]			
	0.05 - 0.10			{1},[2] (1)	(1)			
	0.10 - 0.20				{2},[1] (1)	{1},[1] (1)		
	0.20 - 0.40				(1)	{3},[2] (3)		
	0.40 - 0.60						{1},[2] (2)	{2} (1)
	0.60 - 0.80						(1)	{3} (1)

Dit lyk asof die T- en F-toetse in die geval onder beskouing min of meer ooreenstemmende oorskrydingswaarskynlikhede lewer, maar geen besliste uitspraak kan gelewer word alvorens baie meer berekeninge gedoen is nie.

5.5. ASIMPTOTIESE RELATIEWE DOELTREFFENDHEID.

5.5.1. Inleiding.

Vervolgens wil ons die asimptotiese relatiewe doeltreffendheid bereken van die T-toetse met betrekking tot die F-toets vir alternatiewe van verskuiwing en met betrekking tot Bartlett se toets vir alternatiewe van verspreiding. Vir eers word aangeneem dat daar k_{ij} waarnemings in die sel (i,j) is (ter wille van aansluiting by die notasie van hoofstuk III).

5.5.2. Toetse vir Ry-effekte.

Beskou alle waarnemings binne 'n bepaalde ry as een steekproef, of k steekproewe uit dieselfde populasie. Dan is daar m sulke steekproewe van grootte $k_{i.}$, $i=1,2,\ldots,m$. Dui die verdelingsfunksie van die i^{de} ry aan deur $F_{i.}$, $i=1,2,\ldots,m$. Ons wil nou die nulhipotese H_o: $F_{1.} \equiv F_{2.} \equiv \ldots \equiv F_{m.} \equiv F$ toets onder die aanname dat daar geen kolom- of interaksie-effekte is nie. H_o word getoets teen 'n alternatiewe hipotese H, naamlik dat nie alle $F_{i.}$ identies is nie.

Stel nou

(5.58) $\nu_{Ni.} = k_{i.}/N$,

(5.59) $\nu_{i.} = \lim\limits_{N\to\infty} \nu_{Ni.}$,

(5.60) $F_{i.k_{i.}}(x) = (\text{aantal } x_{ijh} \leq x)/k_{i.}$ vir vaste i en vir $j=1,2,\ldots,k$; $h=1,2,\ldots,k_{ij}$,

(5.61) $H_N(x) = \sum\limits_i \nu_{Ni.} F_{i.k_{i.}}(x)$,

(5.62) $H(x) = \sum\limits_i \nu_{Ni.} F_{i.}(x)$ en

(5.63) $t_{i.} = \int\limits_{-\infty}^{\infty} J_N[H_N(x)]dF_{i.k_{i.}}(x)$ waarby J_N 'n willekeurige funksie is wat voldoen aan soortgelyke voorwaardes as in hoofstuk III (voorwaardes (3.10)-(3.14) en (3.93)).

Soos in lemma 3.14 volg dat

(5.64) $\lim\limits_{N \to \infty} E(t_{i.} | H) = \int\limits_{I} J(H) dF_{i.}$.

Dui ons die variansie van $t_{i.}$ onder H aan deur var($t_{i.}$ |H), kan daar op analoë wyse as in lemma 3.15 'n uitdrukking vir var($t_{i.}$ |H) afgelei word.

Laat

(5.65) $\hat{t}_{i.} = (N-k_{i.})^{\frac{1}{2}}[t_{i.} - E(t_{i.} | H)][N \text{ var}(t_{i.} | H)]^{-\frac{1}{2}}$.

Onder H_o reduseer $\hat{t}_{i.}$ na die vorm aangedui in (5.30).

Nou definieer ons as toetsingsgrootheid

(5.66) $T_m = \sum\limits_{i=1}^{m} \hat{t}_{i.}^{2}$.

In notasie is daar nou aangesluit by dié van hoofstuk III. Alle resultate van hoofstuk III kan aangepas word vir hierdie geval. Soos in stelling 3.6 volg onder analoë voorwaardes dat T_m onder H asimptoties, vir $N \to \infty$, 'n χ^2-verdeling besit met (m-1) grade van vryheid.

Op soortgelyke wyse as in §e 3.6.2 en 3.6.3 kan toetse vir verskuiwing en verskil in verspreiding gedefinieer word met a.r.d.-eienskappe wat ooreenstem met dié van hoofstuk III.

Analoog aan die metode van hierdie paragraaf, kan toetse vir kolomeffekte gekonstrueer word onder die aanname dat daar geen ry-effekte en interaksie voorkom nie.

5.5.3. Toetse vir Sel-effekte.

Beskou die waarnemings binne elke sel as één steekproef uit 'n verdeling met verdelingsfunksie F_{ij}. Dan is daar mk steekproewe van grootte k_{ij}, i=1,2,..,m; j=1,2,...,k. Die nulhipotese H_o: $F_{11} \equiv F_{12} \equiv \ldots \equiv F_{mk} \equiv F$ moet getoets word onder die aanname dat daar geen ry-of kolomeffekte of ry-kolom interaksie bestaan nie.

Stel

(5.67) $\nu_{Nij} = k_{ij}/N$,

(5.68) $\nu_{ij} = \lim_{N\to\infty} \nu_{Nij}$,

(5.69) $F_{ijk_{ij}}(x) = (\text{aantal } x_{ijh} \leq x)/k_{ij}$ vir vaste i en j en vir $h=1,2,\ldots,k_{ij}$,

(5.70) $H_N(x) = \sum_i \sum_j \nu_{Nij} F_{ijk_{ij}}(x)$,

(5.71) $H(x) = \sum_i \sum_j \nu_{Nij} F_{ij}(x)$ en

(5.72) $t'_{ij} = \int_{-\infty}^{\infty} J_N[H_N(x)]dF_{ijk_{ij}}(x)$ waar J_N voldoen aan soortgelyke voorwaardes as in hoofstuk III.

Dui die verwagtingswaarde en variansie van t'_{ij} onder H aan deur $E(t'_{ij}|H)$ en $\text{var}(t'_{ij}|H)$.

Laat

(5.73) $\hat{t}'_{ij} = (N-k_{ij})^{\frac{1}{2}}[t'_{ij} - E(t'_{ij}|H)][N\,\text{var}(t'_{ij}|H)]^{-\frac{1}{2}}$.

Dan definieer ons as toetsingsgrootheid

(5.74) $T_s = \sum_i \sum_j \hat{t}'^{\,2}_{ij}$.

Soos in stelling 3.6 kan aangetoon word dat T_s asimptoties, vir $N\to\infty$, 'n χ^2-verdeling besit met (mk-1) grade van vryheid en kan soos voorheen a.r.d.-uitdrukkings afgelei word vir alternatiewe van verskuiwing en verskil in verspreiding.

5.6. SLOTOPMERKINGS.

1). Die toetse gebaseer op T_m, T_k en T_s in die geval $\Psi_N(\delta) = Q(\delta)$ byvoorbeeld vir verskuiwing, is (onder die genoemde aannames, waaronder dat geen ander effek as dié een waarvoor getoets word, teenwoordig is nie) asimptoties ten minste net so goed soos die F-toets ($E_{T_m,F} \geq 1$).

2). Met behulp van tabelle in VAN DER WAERDEN(1957) is die toepassing van bogenoemde T-toetse (opmerking 1 hierbo) dikwels makliker as dié van die F-toetse.

3). Die vraag oor hoe groot b moet wees alvorens die T-toetse toegepas kan word, is nog nie bevredigend beantwoord nie.

4). Die volgende probleme is onder andere nog onopgelos:

a). Die uitbreiding van die voorgaande teorie na die geval waar die aantal waardes in die onderskeie selle verskillend is.

b). A.r.d.-uitdrukkings vir interaksie-toetse.

c). 'n Volledige verdelingsvrye ekwivalent van die parametriese variansie-analise, naamlik waar byvoorbeeld die toetse vir ry- en kolomeffekte (asimptoties) onafhanklik bly wanneer daar ry- en/of kolomeffekte of interaksie bestaan (m.a.w. H_0 nie geld nie).

5). 'n Ander interessante benadering tot die probleem van 'n verdelingsvrye variansie-analise word gevind in die volgende twee artikels:

HODGES, J.L. en LEHMANN, E. L.(1962) „Rank methods for combination of independent experiments in the analysis of variance", Ann. Math. Stat. 33, pp. 482-497, en LEHMANN, E. L.(1963) „Asymptotically nonparametric inference: An laternative approach to linear models", Ann. Math. Stat. 34, pp. 1494-1506.

BYLAAG A.

Indien

(2.40') $\lim\limits_{N\to\infty} N^{-1}\sum\limits_h (a_h - \bar{\Psi}_N)^2 \geq \varepsilon > 0$ en

(2.42') $N^{-1}\sum\limits_h (\Psi_{Nh} - \bar{\Psi}_N)^4 = O(N)$

dan voldoen die variante $\hat{t}_i$ aan die voorwaarde

(2.50) $\lim\limits_{N\to\infty} E|\hat{t}_i|^{2+\delta} < \infty$ vir $i=1,2,\ldots,k$.

Bewys: Stel $M_N = \sum\limits_h (a_h - \bar{\Psi}_N)^2$ en

$$P_N = \sum_h (a_h - \bar{\Psi}_N)^4 \leq \sum_h (\Psi_{Nh} - \bar{\Psi}_N)^4 .$$

Op analoë wyse as in lemma 2.2 volg dat

$$E(\underline{t}_i - E\underline{t}_i)^4 = \frac{n_i(N-n_i)(N^2+N-6Nn_i+6n_i^2)}{N(N-1)(N-2)(N-3)}\cdot P_N + $$
$$+ \frac{3n_i(n_i-1)(N-n_i)(N-n_i-1)}{N(N-1)(N-2)(N-3)}\cdot M_N^2 .$$

Uit (2.4) volg

$$\hat{t}_i = (N-n_i)^{\frac{1}{2}}[\underline{t}_i - E\underline{t}_i][n_i(N-n_i)N^{-1}(N-1)^{-1}M_N]^{-\frac{1}{2}}$$
$$= [\underline{t}_i - E\underline{t}_i][n_i(N-1)^{-1}M_N]^{-\frac{1}{2}}.$$

$$E|\hat{\underline{t}}_i|^4 = E[\underline{t}_i - E\underline{t}_i]^4[n_i(N-1)^{-1}M_N]^{-2}$$
$$= \frac{(N-1)(N-n_i)(N^2+N-6Nn_i+6n_i^2)}{n_iN(N-2)(N-3)M_N^2}\cdot P_N +$$
$$+ \frac{3(N-1)(N-n_i)(N-n_i-1)(n_i-1)}{n_iN(N-2)(N-3)}$$
$$= O(1)$$

want $P_N M_N^{-2} \leq N^{-1}\sum\limits_h (\Psi_{Nh} - \bar{\Psi}_N)^4\, N^{-1}[N^{-1}\sum\limits_h (a_h - \bar{\Psi}_N)^2]^{-2}$

$= O(1)$ uit (2.40') en (2.42').

Gevolglik geld $E|\hat{t}_i|^{2+\delta} = O(1)$ met $0 < \delta \leq 2$.

Opmerking.

Onder voorwaarde (2.40) geld voorwaarde (2.40') met waarskynlikheid een.

BYLAAG B.

Onder die voorwaardes (3.10) - (3.13), (3.28) en

$$(3.14') \quad \begin{array}{l} |J(H)| \le K[H(1-H)]^{-\frac{1}{4}+\delta} \\ |J^{(m)}(H)| \le K[H(1-H)]^{-m-\frac{1}{2}+\delta} \quad , \; m=1,2 \quad \text{en } \delta > 0 \end{array}$$

geld die voorwaarde

$$(3.76) \quad \lim_{N\to\infty} E|\hat{t}_i|^{2+\delta} < \infty \quad \text{vir } i=1,2,\ldots,k \text{ en } 0 < \delta \le 2.$$

Bewys: $E|\hat{t}_i|^{2+\delta} = \left[\dfrac{N-n_i}{N\mathrm{var}(t_i)}\right]^{1+\frac{1}{2}\delta} E|t_i-Et_i|^{2+\delta}$ uit (3.56).

Aangesien $0 < \lim\limits_{N\to\infty} N\mathrm{var}(t_i) < \infty$ uit (3.68), is dit voldoende om te bewys dat $N^{1+\frac{1}{2}\delta}E|t_i-Et_i|^{2+\delta} = O(1)$.

Uit §3.4 opmerking 3, (3.60) en (3.62) volg:
$E(t_i) = A^{(i)} + o(N^{-\frac{1}{2}})$ met $A^{(i)}$ gedefinieer deur (3.61). Nou is

$$N^{1+\frac{1}{2}\delta}E|t_i-Et_i|^{2+\delta} = N^{1+\frac{1}{2}\delta}E|B_{1N}^{(i)}+B_{2N}^{(i)}+o_p(N^{-\frac{1}{2}})|^{2+\delta}$$

$$= N^{1+\frac{1}{2}\delta}E|\sum_g \nu_{Ng}[n_i^{-1}\sum_{j=1}^{n_i}\tilde{B}_g(x_{ij})-n_g^{-1}\sum_{j=1}^{n_g}\tilde{B}_i(x_{gj})] + o_p(N^{-\frac{1}{2}})|^{2+\delta}.$$

Die resultaat volg m.b.v. lemma 3.11 saam met (3.1) en (3.3) indien bewys kan word dat

$$E|n_i^{-\frac{1}{2}}\sum_{j=1}^{n_i}\tilde{B}_g(x_{ij})|^{2+\delta} = O(1).$$

Onder voorwaarde (3.14') word hieraan voldoen vir $\delta = 2$, want:

$$E|n_i^{-\frac{1}{2}}\sum_j \tilde{B}_g(x_{ij})|^4$$

$= E[n_i^{-2}\sum_j\sum_{j'}\sum_{j''}\sum_{j'''}\tilde{B}_g(x_{ij})\tilde{B}_g(x_{ij'})\tilde{B}_g(x_{ij''})\tilde{B}_g(x_{ij'''})]$ waar alle j 's die waardes $1,2,\ldots,n_i$ deurloop

$= n_i^{-2}[\sum_j\sum_{j'}E\tilde{B}_g^2(x_{ij})E\tilde{B}_g^2(x_{ij'}) + \sum_j E\,\tilde{B}_g^4(x_{ij})]$ want x_{ij} en $x_{ij'}$ is onderling onafhanklik vir $j \ne j'$ en $E\tilde{B}_g(x_{ij})=0$

$= O(1)$

aangesien:

$n_i^{-2}\sum_j\sum_{j'} E\,\tilde{B}_g^2(x_{ij})E\,\tilde{B}_g^2(x_{ij'}) = O(1)$ m.b.v. (3.27) en onder op bladsy 72, en

$n_i^{-2}\sum_j E\,\tilde{B}_g^4(x_{ij}) = O(1)$ soos blyk uit die volgende:

$$|B_g(x_{ij})| = \left|\int_{x_0}^{x_{ij}} J'[H(x)]dF_g(x)\right|$$

$$\leq K\left|\int_{x_0}^{x_{ij}} J'[H(x)]dH(x)\right|$$

$$\leq K\,[\,|J[H(x_{ij})]| + |J[H(x_0)]|\,]$$

$$\leq K'[H(x_{ij})\{1-H(x_{ij})\}]^{\delta-\frac{1}{4}} + O(1) \text{ uit } (3.14').$$

$$E[B_g(x_{ij})]^4 \leq K''\int_0^1 [H(1-H)]^{4(\delta-\frac{1}{4})}dH + O(1)$$

$$= K''\int_0^1 [H(1-H)]^{-1+4\delta}dH + O(1)$$

$$= O(1), \text{ sodat}$$

$$E[\tilde{B}_g(x_{ij})]^4 = O(1).$$

Hiermee is die bewys voltooi.

Opmerkings.

1). Waarskynlik is die meer beperkende voorwaarde (3.14') in vergelyking met (3.14) onnodig. Let in hierdie verband op dat $\lim_{N\to\infty} E\tilde{B}_g^2(x_{ij}) < \infty$ onder (3.14) - sien C+S(1958). Verder is nodig vir bogenoemde afleiding dat $\lim_{N\to\infty} n_i^{-2}\sum_j E\tilde{B}_g^4(x_{ij}) < \infty$ waarby reeds

$\lim_{N\to\infty} n_i^{-1}\sum_j E\tilde{B}_g^4(x_{ij}) < \infty$ as (3.14') geld.

Daar kan ook op gewys word dat

$|B_g(x_{ij})| < K[J\{H(x_{ij})\}]$ en dat

(3.13) $J_N(1) = o(N^{\frac{1}{2}})$.

2). Alle spesiale gevalle van die algemene toetsingsgrootheid wat in hierdie hoofstuk beskou word, voldoen aan die meer beperkende voorwaarde (3.14').

BYLAAG C.

Aan die voorwaardes

i) $\lim_{N \to \infty} E|t'_{i.}|^{2+\delta} = O(1)$,

ii) $\lim_{N \to \infty} E|t'_{.j}|^{2+\delta} = O(1)$ en

iii) $\lim_{N \to \infty} E|t''_{ij}|^{2+\delta} = O(1)$ met $\delta > 0$

word voldoen onder voorwaardes (5.13), (5.14) en

(5.13') $N^{-1} \sum_{p} (a_p - a_{...})^4 = O(N)$.

Bewys: Die bewys dat aan voorwaardes i) en ii) voldoen word, geskied soos in bylaag A.

Deur te skrywe

$a_{ijh} - a_{i..} - a_{.j.} + a_{...} = (a_{ijh} - a_{...}) - (a_{i..} - a_{...}) - (a_{.j.} - a_{...})$,

kan geredelik aangetoon word dat (5.13') voldoende is vir die geldigheid van voorwaarde iii) deur op 'n soortgelyke wyse as in bylaag A te werk te gaan met gebruikmaking van die tegniek wat in die bewys van lemma 5.2 ontwikkel is.

Die geldigheid van voorwaarde (5.13') word verder implisiet aanvaar.

SUMMARY.

DISTRIBUTION-FREE TESTS FOR THE PROBLEM OF TWO OR MORE SAMPLES.

Distribution-free tests are constructed for the k-sample problem ($k \geq 2$). Some of these tests are extensions of two-sample tests proposed by WILCOXON(1945), TERRY(1952), MOOD(1954), VAN DER WAERDEN(1957) and LEMMER and STOKER(1960).

The test statistics have asymptotic χ^2 distributions with (k-1) degrees of freedom under the null hypothesis H_o (chapter II), and also under the general hypothesis H subject to certain general conditions (chapter III - see CHERNOFF and SAVAGE(1958)). A number of these k-sample distribution-free tests for difference in location are compared with the parametric F test and in the case of tests for dispersion with the Bartlett test for homogeneity of variance by means of their asymptotic relative efficiencies. Several of the k-sample tests compare favourably with the parametric tests and are highly recommended for practical application. Necessary and sufficient conditions are given for the consistency of these tests.

Different tests are suggested for the problem of m rankings (chapter IV) and their asymptotic properties derived.

The analysis of variance in the two-factor case is studied in chapter V. Different tests for row-, column- and interaction effects are constructed, which, under H_o, possess asymptotic χ^2 distributions and are distributed independent of each other.

LITERATUURLYS.

1. ANDREWS, F. C.(1954) „Asymptotic behavior of some rank order tests for analysis of variance", Ann. Math. Stat. 25, pp. 724-736.

2. ANSARI, A. R. en BRADLEY, R. A.(1960) „Rank-sum tests for dispersions", Ann. Math. Stat. 31, pp. 1174-1189.

3. BENARD, A en VAN ELTEREN, P.(1953) „A generalization of the method of m rankings", Indag. Math. 15, pp. 358-369.

4. CAPON, J.(1961) „Asymptotic efficiency of certain locally most powerful rank tests", Ann. Math. Stat. 32, pp. 88-100.

5. CHACKO, V. J.(1963) „Testing homogeneity against ordered alternatives", Ann. Math. Stat. 34, pp. 945-956.

6. CHERNOFF, H. en SAVAGE, I. R. (1958) „Asymptotic normality and efficiency of certain nonparametric test statistics", Ann. Math. Stat. 29, pp. 972-994.

7. CRAMéR, H.(1946) „Mathematical Methods of Statistics", Princeton University Press, Princeton.

8. CRAMéR, H.(1962) „Random Variables and Probability Distributions", 2de uitgawe, Cambridge University Press.

9. CROUSE, C. F.(1960) „Combinatorial tests, for differences in location and dispersion, for the case of m samples", Proefskrif, Universiteit van London.

10. DWASS, M.(1956) „The large-sample power of rank order tests in the two-sample problem", Ann. Math. Stat. 27, pp. 352-374.

11. FRASER, D. A. S.(1957) „Nonparametric Methods in Statistics", New York.

12. FREUND, J. E. en ANSARI, A. R.(1957) „Two-way rank-sum tests for variances", Technical Report No. 34 to Office of Ordnance Research and National Science Foundation, Virginia Polytechnic Institute.

13. FRIEDMAN, M.(1937) „The use of ranks to avoid the assumption of normality implicit in the analysis of variance", Jour. Am. Stat. Ass. 32, pp. 675-699.

14. GIBSON, G. A.(1954) „Advanced Calculus", MacMillan and Co. Ltd., London.

15. HODGES, J. L. en LEHMANN, E. L.(1956) „The efficiency of some nonparametric competitors of the t-test", Ann. Math. Stat. 27, pp. 324-335.

16. HODGES, J. L. en LEHMANN, E. L.(1961) „Comparison of the normal scores and Wilcoxon tests", Proc. 4 th. Berkeley Symp. on Math. Stat. and Prob. 1, pp. 307-318.

17. HOEFFDING, W.(1948) „A class of statistics with asymptotically normal distribution", Ann. Math. Stat. 19, pp. 293-325.

18. HOEFFDING, W.(1951) „A combinatorial central limit theorem", Ann. Math. Stat. 22, pp. 558-566.

19. HOEFFDING, W.(1952) „The large sample power of tests based on permutations of observations", Ann. Math. Stat. 23, pp. 169-192.

20. KENDALL, M. G.(1948) „Rank Correlation Methods", London.

21. KENDALL, M. G. en STUART, A.(1958) „The advanced Theory of Statistics", vol.I, Griffin and Co. Ltd., London.

22. KENDALL, M. G. en STUART, A.(1961) „The advanced Theory of Statistics", vol. II, Griffin and Co., Ltd., London.

23. KENNEY, J. F. en KEEPING, E. S.(1951) „Mathematics

of Statistics", vol. II, 2de uitgawe, D Van Nostrand, New York.

24. KLOTZ, J.(1962) „Nonparametric tests for scale", Ann. Math. Stat. 33, pp. 498-512.

25. KRUSKAL, W. H.(1952) „A non-parametric test for the several sample problem", Ann. Math. Stat. 23, pp. 525-540.

26. LEMMER, H. H. en STOKER, D. J.(1960) „'n Ondersoek na die doeltreffendheid van verdelingsvrye toetse vir die probleem van twee klein steekproewe om vir verskil in spreiding te toets", Tydskrif vir Wetenskap en Kuns, Okt. 1960, pp. 177-191.

27. LEMMER, H. H. en STOKER, D. J.(1961) „'n Klas van verdelingsvrye toetsingsgroothede vir die probleem van k onafhanklike steekproewe", Tydskrif vir Natuurwetenskappe, S. A. Akademie vir Wetenskap en Kuns, I, No. 4, Des. 1961, pp. 231-246.

28. LOèVE, M.(1955) „Probability Theory", 2de uitgawe, D. Van Nostrand, New Jersey.

29. MOOD, A. M.(1950) „Introduction to the Theory of Statistics", McGraw-Hill Book Co., New York.

30. MOOD, A. M.(1954) „On the asymptotic efficiency of certain nonparametric two-sample tests", Ann. Math. Stat. 25, pp. 514-522.

31. NOETHER, G. E.(1949) „On a theorem by Wald and Wolfowitz", Ann. Math. Stat. 20, pp. 455-458.

32. PITMAN, E. J. G.(1937) „Significance tests which may be applied to samples from any populations", I, Jour. Roy. Stat. Soc. Suppl. 4, pp. 119-130.

33. RIJKOORT, P. J.(1952) „A generalisation of Wilcoxon's test", Indag. Math. 14, p. 394-404.

34. SCHEFFé, H.(1959) „The Analysis of Variance", New York.

35. SCHUTTE, H. J. en VAN ROOY, D. J.(1961) „Lineêre Algebra en Meetkunde", Pro Rege-Pers Bpk., Potchefstroom.

36. STOKER, D. J.(1955) „Oor 'n klas van toetsingsgroothede vir die probleem van twee steekproewe", Proefskrif, Universiteit van Amsterdam.

37. SUKHATME, B. V.(1957) „On certain two-sample nonparametric tests for variances", Ann. Math. Stat. 28, pp. 188-194.

38. SUKHATME, B. V.(1958) „Testing the hypothesis that two populations differ only in location", Ann. Math. Stat. 29, pp. 60-78.

39. TANG, P. C.(1938) „The power function of the analysis of variance tests with tables and illustrations of their use", Statistical Research Memoirs, vol. II, University College, London, pp. 126-149.

40. TATE, M. W. en CLELLAND, R. C.(1957) „Nonparametric and shortcut statistics in the social, biological and medical sciences", Interstate Printers and Publishers, Danville, Illinois.

41. TERPSTRA, T. J.(1955) „A generalization of Kendall's rank correlation statistic", I, Indag. Math. 17, pp. 690-696.

42. TERPSTRA, T. J.(1956) „A generalization of Kendall's rank correlation statistic", II, Indag. Math. 18, pp. 59-66.

43. TERPSTRA, T. J.(1962) „Interpretations and generalizations of Kendall's rank correlation test", Proefskrif, Rijksuniversiteit, Groningen.

44. TERRY, M. E.(1952) „Some rank order tests, which are most powerful against spesific parametric alternatives", Ann. Math. Stat. 23, pp. 346-366.

45. TURNBULL, W. H.(1947) „Theory of Equations", Oliver and Boyd, London.

46. USPENSKY, J. V.(1937) „Introduction to Mathematical Probability", McGraw-Hill, New York.

47. VAN DANTZIG, D.(1947-1950) „Kadercursus Mathematische Statistiek", Amsterdam, Mathematisch Centrum.

48. VAN DER WAERDEN, B. L.(1957) „Mathematische Statistik", Springer-Verlag, Berlyn.

49. VAN ELTEREN, P.(1957) „The asymptotic distribution for large m of Terpstra's statistic for the problem of m rankings", Indag. Math. 19, pp. 522-534.

50. VAN ELTEREN, P. en NOETHER, G. E.(1959) „The asymptotic efficiency of the χ_r^2-test for a balanced incomplete block design", Biometrika 46, p. 475.

51. WILCOXON, F.(1945) „Individual comparisons by ranking methods", Biometrics 1, pp. 80-83.

52. WILKS, S. S.(1962) „Mathematical Statistics", Wiley and Sons, New York.