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. $Z = (z_1, z_2, \ldots, z_N) \varepsilon E_N$.

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 $Z \varepsilon E_N$ 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

$$(1.5) \quad \lambda^2 = \sum_{j=1}^{k} \sum_{j'=1}^{k} \frac{|\Lambda_{jj'}|}{|\Lambda|} E(x_j)E(x_{j'}).$$ 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,...,k en ζ,ζ' die waardes 1,2,...,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) $\phi_{x_1,\ldots x_k}(\vec{t}) = e^{-\frac{1}{2}\vec{t}'\Lambda\vec{t}}$

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

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

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

$\sum_{i=1}^{k} y_i^2$ 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

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

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

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

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:

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

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

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:

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

$= 0$, 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

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

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

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

Bewys: Dit is voldoende om aan te toon dat $A_1A_1^{-1} = I$, die eenheidsmatriks.

Laat $A_1A_1^{-1} = \{\alpha_{\zeta\zeta'}\}$, '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 momente-matriks Λ.

STELLING 1.1.

Indien

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

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

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

dan besit $T = \sum\limits_{i=1}^{k} x_i^2$ '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 $\text{var}(\underline{x}_i) = 1-\nu_i$ vir $i=1,2,\ldots,k$ en $\text{kov}(\underline{x}_i,\underline{x}_j) = -\sqrt{\nu_i\nu_j}$ vir $i \neq j$.

Gevolglik is

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

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

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

Uit (1.27) is dus

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

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

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

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

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

$$= \nu_k^{-1}[\nu_k\sum_\zeta x_\zeta^2 + (-\sqrt{\nu_k}\ x_k)^2]$$ 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 $\sum\limits_{i=1}^{k} x_i^2$ 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) $\sum_i \sqrt{\nu_i}\ \mu_i = 0$,

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

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 ($0\leq\delta\leq1$) 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 $\varepsilon\leq\delta\leq1-\varepsilon$. 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

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

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_0 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_0.

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:

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

Bewys (sien STOKER (1955)):

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

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

waar Σ' 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

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

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

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

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

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

Bewys (sien STOKER(1955)):

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

waarby Σ' dieselfde beteken as in lemma 2.1

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

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

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

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

Opmerkings.

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

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

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

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

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

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

---

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

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

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:

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

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

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

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

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

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) $$\frac{\text{maks}_h (a_h-\bar{a}_N)^2}{N^{-1}\sum_h (a_h-\bar{a}_N)^2} = o_p(N) \text{ vir } N\to\infty$$ 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 $\text{maks}_h (b_{Nh}-\bar{b}_N)^2$ altyd eindig (vir alle N), en wel

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

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

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 terminste 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:

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

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

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

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

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 $\lim_{N\to\infty} \bar{\Psi}_N$, waarby $\bar{\Psi}_N = N^{-1}\sum_h \Psi_{Nh}$, 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) $\liminf_{N\to\infty} \sigma^2_{N,a}(g_1, g_2, \ldots, g_\lambda) \geq c^2$ 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:

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

(2.42) $\max_h [\Psi_N(\delta_h)-\bar{\Psi}_N]^2 = o(N)$.

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

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

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

verdeel is as $N\to\infty$ waarby

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

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

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

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

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

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

(2.51) $\hat{\underline{S}} \overset{\text{def}}{=} \sum\limits_i e_i \hat{\underline{t}}_i$ (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) $\max\limits_h(\Psi_{Nh}-\bar{\Psi}_N)^2 = o(N)$ en

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

is voldoende vir die geldigheid van die voorwaardes

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

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

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: $\lim_{N\to\infty} E(T_k) = (k-1)$ en

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

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

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

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

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

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

Gevolglik is

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

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.

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

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

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

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

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

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

(2.60) $E(t_i) = n_i N^{-1}\sum\limits_h h/(N+1)$ 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) $\sigma_a^2 = [N(N^2-1)+\sum\limits_\alpha g_\alpha-\sum\limits_\alpha g_\alpha^3][12N(N+1)^2]^{-1}$

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

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

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

Uit (2.18) volg nou dat

(2.63) $\mathrm{var}(t_i) = \dfrac{n_i(N-n_i)}{12(N+1)}[1-\dfrac{\gamma}{N(N^2-1)}]$

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

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

Die toetsingsgrootheid T_k word nou

(2.65) $T_W = \sum_i \frac{12(N+1)}{Nn_i\beta}(t_i - \frac{1}{2}n_i)^2$ 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 $\max_h(\Psi_{Nh} - \bar{\Psi}_N)^2 = O(1)$.

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 $i=1,2,\ldots,k$,

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

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) $\mathrm{var}(R_i) = \frac{n_i(N-n_i)(N+1)}{12}[1 - \frac{\gamma}{N(N^2-1)}]$

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

Vervolgens herlei (2.65) na:

$-3(N+1)[1 - \frac{\gamma}{N(N^2-1)}]^{-1}$

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

$= H^{*}$ 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) $t_i = \sum_j \Xi_{r_{ij}}$

(2.69) $E(t_i) = n_i N^{-1} \sum_h \Xi_h$

$= 0$ weens simmetrie.

Verder is

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

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

As toetsingsgrootheid volg nou uit (2.58):

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

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 strengstygende 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) $\text{var}(t_i) = n_i(N-n_i)N^{-1}(N-1)^{-1} \sum_h a_h^2$ 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) $$\begin{aligned}\text{maks}_h(\Psi_{Nh}-\bar{\Psi}_N)^2 &= \text{maks}_h[Q(\delta_h)-\bar{Q}]^2 \\ &= \text{maks}_h \; Q^2(\delta_h) \\ &= O(2 \log_e(N+1))\end{aligned}$$

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 $\lim\limits_{N\to\infty} \Psi_N(\delta) = \Psi(\delta)$ 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 $\delta = \delta_{N_o}$, $0 < \delta_{N_o} < 1$, die „keerpunt" (maksimumpunt of minimumpunt) bereik en $\Psi(\delta)$ in die punt $\delta = \delta_o$, $0<\delta_o<1$ die keerpunt bereik, dat $|\delta_{N_o} - \delta_o| \xrightarrow[\infty]{N} 0$ en wel gelykmatig in N.

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

$$\liminf_{N\to\infty} \sigma^2_{F,a}(\underline{g}_1,\underline{g}_2,\ldots,\underline{g}_\lambda) \geq c^2 \quad \text{spr } 0$$

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

a). N ewe.

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

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

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

Vir var(t_i) bereken ons eers:

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

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

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

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

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

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 $\frac{1}{2}N$ 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 $\Psi_N(\delta_h)\equiv\delta_h$

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

In hierdie geval is

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

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

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

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

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

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

waar

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

Verder is

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

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

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

Hier is

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

(2.102) $$\begin{aligned} E(t_i) &= n_i N^{-1} \sum_h \Psi_N(\delta_h) \\ &= n_i N^{-1} \sum_h [h/(N+1) - \tfrac{1}{2}]^2 \\ &= \frac{n_i(N-1)}{12(N+1)} . \end{aligned}$$

Uit (2.20) volg:

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

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

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:

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

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

met

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

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

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

(2.109) $T_M = \dfrac{180(N+1)^3}{N(N^2-4)} \sum\limits_i \dfrac{t_i^2}{n_i} - \dfrac{5(N-1)^2(N+1)}{4(N^2-4)}$.

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

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

$\mathrm{var}(M) = \frac{1}{180}\, mn(N^2-4)(N+1)$ 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) $t_i = \sum\limits_j [r_{ij}/(N+1) -\frac{1}{2}]^2$.

Ons onderskei twee gevalle.

a). N ewe.

(2.111) $E(t_i) = n_i N^{-1} \sum\limits_h [r_h/(N+1) -\frac{1}{2}]^2$

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

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

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

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

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

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

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

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

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

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

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

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

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

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.

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

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

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

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

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

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

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

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

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

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

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 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

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

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

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

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

$$(2.124)\quad \sigma_a^2=\frac{(2N+1)(8N+11)(N-1)}{180(N+1)^3}-\frac{1}{180N(N+1)^4}\sum_{\alpha=1}^{\lambda} g_\alpha(g_\alpha^2-1)\ x$$
$$x\ [(8g_\alpha+11)(2g_\alpha+1)+60G_{\alpha-1}(g_\alpha+1)+60G_{\alpha-1}^2]\ ,$$

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

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) $$\sigma_a^2 = \frac{(N^2-4)(4N+11)}{720(N+1)^3} - \frac{1}{180N(N+1)^4}\Big\{\sum_{\alpha=1}^{\beta} g_\alpha(g_\alpha^2-1) \times$$
$$\times[(8g_\alpha+11)(2g_\alpha+1)+60G_{\alpha-1}(g_\alpha+1)+60G_{\alpha-1}^2] +$$
$$+ \sum_{\alpha=\beta+1}^{\lambda} g_\alpha(g_\alpha^2-1)[(8g_\alpha+11)(2g_\alpha+1)+60G'_{\alpha+1}(g_\alpha+1)+60G'^2_{\alpha+1}]\Big\} .$$

Nou is

(2.134) $\mathrm{var}(t_i) = \frac{n_i(N-n_i)\sigma_a^2}{(N-1)}$ 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) $$\operatorname{var}(t_i) = \frac{n_i(N-n_i)(N^2-4)(4N+11)}{720(N+1)^3(N-1)}$$ 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) $$\sigma_c^2 = \frac{1}{N}\sum_{h=1}^{N}\left[\frac{r_h^2}{(N+1)^2} - \frac{N^2+2N+3}{12N(N+1)}\right]^2$$

$$= \frac{1}{N(N+1)^4}\left[2\sum_{h=1}^{\frac{1}{2}(N-1)} h^4 + \left(\frac{N+1}{2}\right)^4 - \frac{(N+1)(N^2+2N+3)}{3N}\sum_{h=1}^{\frac{1}{2}(N-1)} h^2 - \frac{(N+1)^3(N^2+2N+3)}{24N} + \frac{(N+1)^2(N^2+2N+3)^2}{144N}\right]$$

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

$(\sigma_c^2-\sigma_a^2)$ 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

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

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

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

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

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

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

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

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

Die toetsingsgrootheid is

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

Uit STOKER (1955) p. 57 volg:

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

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) $E(t_i) = n_i N^{-1} \sum_h E(\xi^2_h)$ 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 $\Psi_N(\delta_h) = E(\xi^2_h)$.

Die toetsingsgrootheid T_k is in hierdie geval:

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

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

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

$$\xi^2_N = O_p(2 \log_e N),$$

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

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

sodat voorwaarde (2.42) dus bevredig word.

Onder die voorwaardes

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

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

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

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

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:

$$(2.160)\quad E(t_i|H_{0\sigma}; Z) = n_i N^{-1} \sum_h a_h .$$

Die verwagtingswaarde is voorwaardelik.

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

$$(2.161)\quad \mathrm{var}(t_i|H_{0\sigma}; Z) = n_i(N-n_i)N^{-1}(N-1)^{-1} \sum_h (a_h-\bar{\Psi}_N)^2$$

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

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

Soortgelyk:

$$(2.163)\quad \mathrm{kov}(t_i,t_{i'}|H_{0\sigma}; Z) = -n_i n_{i'}(N-1)^{-1}\sigma_a^2 .$$

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

$$(2.35) \quad \frac{\underset{h}{\text{maks}}(a_h-\bar{\Psi}_N)^2}{N^{-1}\sum\limits_h(a_h-\bar{\Psi}_N)^2} = o_p(N) \text{ vir } N\to\infty,$$

(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 $\sum\limits_i c_i t_i$ 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 $\sum\limits_i c_i t_i$ asimptoties normaal verdeel is as $N\to\infty$ (sien HOEFFDING (1952) en STOKER (1955) p. 34).

Opmerkings.

1). Onder die voorwaarde

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

is die voorwaarde

$$(2.165) \quad \underset{h}{\text{maks}}(a_h-\bar{\Psi}_N)^2 = o_p(N)$$

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:

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

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

STELLING 2.7.

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

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

Bewys: Soos stelling 2.4.

Opmerking.

Hierdie stelling geld voorwaardelik vir gegewe Z.

STELLING 2.8.

Indien

(2.164) $\liminf_{N\to\infty} \sigma_a^2 \geq \varepsilon > 0$ 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) $a_h = g_\alpha^{-1} \sum_{\mu=1}^{g_\alpha} z_{G_{\alpha-1}+\mu}$ 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) $\lim\limits_{N\to\infty} N^{-1} \sum\limits_{h}(z_h-\bar{z})^2 = \text{var}(\underline{z})$ (in waarskynlikheid)

>0 (sodat voorwaarde (2.164)

bevredig word), en

(2.174) $\lim\limits_{N\to\infty} N^{-1} \sum\limits_{h}|z_h-\bar{z}|^{2+\delta} \leq E|\underline{z}-\mu|^{2+\delta}$ (in waarskynlikheid)

$< \infty$ vir $\delta>0$ 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) $T_p' = \frac{(kb-1)}{k\, S_z^2} \sum\limits_{i=1}^{k} (b^{-1} \sum\limits_{j=1}^{b} x_{ij}-\bar{z})^2$.

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

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

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

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.