MATHEMATISCH CENTRUM
2e Boerhaavestraat 49
Amsterdam - 0

Statistische Afdeling
Memorandum S 190 (M 50A)

Toets voor een generalisatie van het probleem van m[1]) rangschikkingen.

Gegeven is een systeem van $m \times k$ onderling onafhankelijke stochastische grootheden $\underline{x}_1^{(\alpha)}, \underline{x}_2^{(\alpha)}, \dots, \underline{x}_k^{(\alpha)}$ $(\alpha=1,2,\dots,m)$. Er zijn m waarnemers, waarnemer No α verricht $n_i^{(\alpha)}$ waarnemingen van de variabele $\underline{x}_i^{(\alpha)}$. De grootheden hebben betrekking op k verschillende objecten of omstandigheden; de index i correspondeert met object (omstandigheid) .

Op grond van de waarnemingen wensen we nu de hypothese H_0 te toetsen, die inhoudt dat voor iedere α $(\alpha=1,\dots,m)$ de variabelen $\underline{x}_1^{(\alpha)}, \underline{x}_2^{(\alpha)}, \dots \underline{x}_k^{(\alpha)}$ dezelfde kansverdeling hebben. Men wil verder dat deze toets onderscheidend is tegen het alternatief dat inhoudt dat er een overeenstemming bestaat tussen de volgorden toegekend aan de objecten door de waarnemers, op grond van de gemiddelden der door hen waargenomen stochastische grootheden.

Het verdient de voorkeur bij het opzetten van het experiment $n_i^{(\alpha)} = \frac{\sum_i n_i^{(\alpha)}}{k}$ $(i=1,2\dots\dots k\,;\,\alpha=1,2\dots\dots m)$ te kiezen. Dit komt het onderscheidingsvermogen van de toets ten goede en vereenvoudigt het rekenwerk.

De toetsingsgrootheid $\underline{S}$ waarmee H_0 tegen H wordt getoetst wordt nu op de volgende wijze gevormd:
We vergelijken bij waarnemer α de i^e en de j^e steekproef $(1 \leqq i < j \leqq k)$ door de grootheid $\underline{\tilde{W}}_{i,j}^{(\alpha)}$ te berekenen, welke wordt gedefiniëerd als het aantal paren (λ,ν) $(1 \leqq \lambda \leqq n_i^{(\alpha)}; 1 \leqq \nu \leqq n_j^{(\alpha)})$ waarvoor $\underline{x}_{i\lambda}^{(\alpha)} > \underline{x}_{j\nu}^{(\alpha)}$ is, verminderd met het aantal paren (λ,ν) $(1 \leqq \lambda \leqq n_i^{(\alpha)}; 1 \leqq \nu \leqq n_j^{(\alpha)})$ waarvoor $\underline{x}_{i,\lambda}^{(\alpha)} < \underline{x}_{j\nu}^{(\alpha)}$ is.
$\underline{\tilde{W}}_{i,j}^{(\alpha)}$ is dus de gereduceerde Wilcoxon-grootheid (zie Mem. S47 (M7)) voor de i^e en de j^e steekproef van waarnemer α.
Indien de steekproeven groot zijn kan men $\underline{\tilde{W}}_{i,j}^{(\alpha)}$ sneller op de volgende wijze berekenen:

We kennen aan de $n_i^{(\alpha)} + n_j^{(\alpha)}$ waarnemingen uit de i^e en j^e steekproef van waarnemer α de rangnummers $1,2\dots\dots(n_i^{(\alpha)} + n_j^{(\alpha)})$ toe naar opklimmende grootte van de waarnemingen. Indien een

---

1) Zie voor m=2 de memoranda S190 (M49) en S168 (M61).

aantal waarnemingen uit deze twee steekproeven dezelfde waarde heeft die verschilt van de waarden der andere waarnemingen wordt aan elk van deze waarnemingen hetzelfde rangnummer toegekend, en wel het gemiddelde van de rangnummers die ze gekregen zouden hebben als ze alle verschillend waren, doch in de rangschikking naar opklimmende grootte dezelfde plaatsen tussen de andere waarnemingen behouden hadden.
Noemen we nu de som der rangnummers van de waarnemingen uit de i^e steekproef $\underline{R}_{ij}^{(\alpha)}$, dan is

$$\underline{\tilde{W}}_{ij}^{(\alpha)} = 2\,\underline{R}_{i,j}^{(\alpha)} - n_i^{(\alpha)}(n_i^{(\alpha)} + n_j^{\alpha} + 1). \qquad (1 \leqq \alpha \leqq m)$$

De toetsings-grootheid $\underline{S}$ wordt nu gegeven door:

$$\underline{S} = \sum_{\alpha < \beta} \underline{S}_{\alpha,\beta},$$

waarin

$$\underline{S}_{\alpha,\beta} = \sum_{i<j} \frac{\underline{\tilde{W}}_{i,j}^{(\alpha)} \cdot \underline{\tilde{W}}_{i,j}^{(\beta)}}{n_i^{(\alpha)} n_j^{(\alpha)} n_i^{(\beta)} n_j^{(\beta)}}$$

Onder de nulhypothese is de verwachting van $\underline{S}$:

$$\mathcal{E}(\underline{S}|H_o) = 0.$$

Stel dat de waarnemingen van waarnemer α bestaan uit τ_α groepen van $t_h^{(\alpha)}$ $(h = 1,2,\ldots\ldots \tau)$ onderling gelijke waarnemingen $(\alpha = 1,2,\ldots\ldots m)$ 2)

We definiëren nu de volgende grootheden:

$$T_2^{(\alpha)} = 1 - \frac{\sum_1^{\tau_\alpha} h\; t_h^{(\alpha)}(t_h^{(\alpha)} - 1)}{n^{(\alpha)}(n^{(\alpha)} - 1)} \qquad \text{voor} \quad n^{(\alpha)} = \sum_1^{K} i\; n_i^{(\alpha)} \geqq 2$$

$$T_2^{(\alpha)} = 0 \quad \text{voor} \quad n^{(\alpha)} < 2.$$

$$T_3^{(\alpha)} = 1 - \frac{\sum_1^{\tau_\alpha} h\; t_h^{(\alpha)}(t_h^{(\alpha)} - 1)(t_h^{(\alpha)} - 2)}{n^{(\alpha)}(n^{(\alpha)} - 1)(n^{(\alpha)} - 2)} \qquad \text{voor} \quad n^{(\alpha)} \geqq 3.$$

$$T_3^{(\alpha)} = 0 \quad \text{voor} \quad n^{(\alpha)} < 3.$$

$$T_{2,3}^{(\alpha)} = \tfrac{1}{2} T_2^{(\alpha)} - \tfrac{1}{3} T_3^{(\alpha)}.$$

Zij voorts voor elk paar (α,β) $(1 \leqq \alpha < \beta \leqq m)$ het aantal waarden van i $(i = 1,\ldots\ldots K)$ waarvoor $n_i^{(\alpha)}, n_i^{(\beta)} > 0$ is, $K_{\alpha,\beta}$, en laat het symbool $\sum i'$ een sommatie over deze waarvan van i. aangeven.

----

2) De volgorde van deze getallen $t_h^{(\alpha)}$ bij rangschikking naar opklimmende grootte van de waarnemingen uit de groepen is van invloed op de exacte verdeling onder H van $\underline{S}$, doch niet op verwachting en variantie van $\underline{S}$.

De variantie van $\underline{S}$ onder de nulhypothese wordt dan gegeven door:

$$\mathrm{Var.}(\underline{S}\mid H_0) = \sum_{\alpha<\beta} \mathrm{Var.}(\underline{S}_{\alpha,\beta}\mid H_0),$$

waarin

$$\mathrm{Var.}(\underline{S}_{\alpha,\beta}\mid H_0) = 0 \quad \text{voor } k_{\alpha,\beta} < 2$$

en

$$\mathrm{Var}(\underline{S}_{\alpha,\beta}\mid H_0) = \Big\{ 2T_{2,3}^{(\alpha)} \cdot T_{2,3}^{(\beta)} \cdot \sum_i{}' (n_i^{(\alpha)} \cdot n_i^{(\beta)})^{-1} + \frac{2}{3} T_{2,3}^{(\alpha)} T_3^{(\beta)} \sum_i{}' (n_i^{(\alpha)})^{-1} + \frac{2}{3} T_3^{(\alpha)} T_{2,3}^{(\beta)} \sum_i (n_i^{(\beta)})^{-1} +$$

$$+ \frac{1}{9} k_{\alpha,\beta}(k_{\alpha,\beta}-2) T_3^{(\alpha)} T_3^{(\beta)} \Big\} \sum_i{}' (n_i^{(\alpha)} n^{(\beta)})^{-1} +$$

$$+ \frac{1}{9} T_3^{(\alpha)} T_3^{(\beta)} \sum_i{}' (n_i^{(\alpha)})^{-1} \sum_i{}' (n_i^{(\beta)})^{-1} - \frac{2}{3} \sum_i{}' \Big\{ T_3^{(\alpha)} T_{2,3}^{(\beta)} n_i^{(\alpha)} + T_{2,3}^{(\alpha)} T_3^{(\beta)} n_i^{(\beta)} + 3 T_{2,3}^{(\alpha)} T_{2,3}^{(\beta)} \Big\} (n_i^{(\alpha)} n^{(\beta)})^{-2}$$

voor $k_{\alpha,\beta} \geqq 2$

Voor grote waarden van k heeft $\underline{S}$ onder de nulhypothse bij benadering een normale verdeling.
Onder de alternatieve hypothesen H zal $\underline{S}$ over het algemeen grotere waarden aannemen dan onder de nulhypothse.
De kritieke zone krijgt dus de vorm

$$S \geqq S_\alpha$$

waarin S_α de kleinste waarde is die voldoet aan

$$P[\underline{S} \geqq S_\alpha \mid H_0] \leqq \alpha$$

Voor grote waarden van k geldt bij benadering

$S_\alpha = \xi_\alpha \sqrt{\mathrm{Var}(\underline{S}\mid H_0)}$ waarin ξ_α gedefiniëerd wordt door

$\frac{1}{\sqrt{2\pi}} \int_{\xi_\alpha}^{\infty} e^{-\frac{1}{2}t^2} \cdot dt = \alpha$, zodat ξ_α in een tabel van de normale verdelingsfunctie kan worden opgezocht.

De overschrijdingskans k^* van de gevonden waarde S van $\underline{S}$ is per definitie $k^* = P[\underline{S} \geqq S \mid H_0]$, en voor grote waarden van k is

$$k^* = \frac{1}{\sqrt{2\pi}} \int_u^{\infty} e^{-\frac{1}{2}t^2} dt \quad \text{waarin } u = \frac{S}{\sqrt{\mathrm{Var}(\underline{S}\mid H_0)}}$$

Men kan dan k^* met behulp van een tabel van de normale verdelingsfunctie bepalen.

Litteratuur:

[1.] M.G. Kendall, Rank Correlation methods, London 1952.

[2.] W.J. Dixon and F.J. Massey Jr., Introduction to statistical analysis, Mc Graw-Hill Book Comp., 1951.

[3.] T.J. Terpstra, A generalization of Kendall's rank correlation statistic

I. Proc.Kon.Ned.Akad.v.Wetensch.A58, 690-696; Indagationes Mathematicae 17, 690-696.

II. Proc.Kon.Ned.Akad.v.Wetensch.A59, 59-66; Indagationes Mathematicae 18, 59-66.

[4.] Math.Centrum, Een generalisatie van Kendalls rangcorrelatie-toets, Rapport S190 (M49).

[5.] Math.Centrum, Een parametervrije toets tegen verloop van groepen waarnemingen, Rapport S168 (M61).

[6.] Math.Centrum, Toets voor het probleem van m rangschikkingen, Rapport S190 (M50).

[7.] Math.Centrum, Methode der m rangschikkingen, Rapport S47 (M14).

[8.] Math.Centrum, De toets van Wilcoxon, Rapport S47 (M7).