STICHTING
MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM

S 178 (M 67)

Verdelingsvrije toetsen voor twee steekproeven en de methode der 2 × 2-tabel

Constance van Eeden

Verdelingsvrije toetsen voor twee steekproeven en de methode der 2×2 -tabel *)

door Constance van Eeden

S u m m a r y

D i s t r i b u t i o n f r e e two sample tests and the method of the 2×2 - table.

The method of the 2×2 *- table may be used for testing the hypothesis that two probabilities* p_A *and* p_B *are equal. This test is based on two independent series A and B of independent trials, each trial resulting in a success or a failure with probabilities* p_A *and* $q_A = 1 - p_A$ *for series A,* p_B *and* $q_B = 1 - p_B$ *for series B.*

Each of these trials may be considered as an observation of a random variable which may take the two „values": success and failure, which may be denoted by two symbols e.g. z_1 *and* z_2*. Then the two series A and B are two independent series of observations of two random variables* $\underline{x}$ *(for A) and* $\underline{y}$ *(for B), taking the value* z_1*, with probability* p_A *(respectively* p_B*) and the value* z_2 *with probability* q_A *(respectively* q_B*). The hypothesis* $p_A = p_B$ *is then identical with the hypothesis that* $\underline{x}$ *and* $\underline{y}$ *possess the same probability distribution. This hypothesis may be tested by means of a distributionfree two sample test.*

A class of distributionfree two sample tests mentioned in [1] *(cf. p. 251) is considered. It contains i.a. the tests of E. J. G. P i t m a n, F. W i l c o x o n, M. E. T e r r y and B. L. v a n d e r W a e r d e n. In this mainly expository paper it is shown that each test belonging to this class, when applied to the series of observations mentioned above, is identical with the method of the* 2×2 *- table.*

1. Inleiding

Indien men de hypothese wil toetsen, dat twee kansen p_A en p_B aan elkaar gelijk zijn, gebruikt men gewoonlijk de methode der 2×2 - tabel. Men heeft daarvoor twee reeksen (A en B) van experimenten nodig, alle experimenten moeten onafhankelijk zijn en bij ieder experiment van reeks A is de kans op „succes" p_A, bij B is deze p_B.

Bestaat A uit m experimenten met a_1 successen en $a_2 = m - a_1$ mislukkingen en B uit n experimenten met b_1 successen en $b_2 = n - b_1$ mislukkingen, dan kan men de resultaten als volgt samenvatten:

*) Rapport S 178 (M 67) van de Statistische Afdeling van het Mathematisch Centrum, Amsterdam.

TABEL I. *Waarnemingsschema*

reeks	aantallen		
	successen	mislukkingen	experimenten
A	a_1	a_2	m
B	b_1	b_2	n
totaal	t_1	t_2	N

Op deze tabel kan men nu, als de aantallen groot zijn, een benaderende toets toepassen, b.v. de χ^2-toets; zijn de aantallen klein, dan kan men exact toetsen. De uitvoering van deze toetsen is voor de hier te behandelen kwestie niet van belang; men vindt deze b.v. beschreven in [2].Wel van belang is het feit, dat deze toetsen *voorwaardelijke* toetsen zijn, onder de voorwaarde dat er in totaal t_1 successen (en dus $t_2 = N - t_1$ mislukkingen) gevonden zijn. Men redeneert als het ware als volgt: er zijn t_1 successen gevonden, laten wij nu nagaan of er hiervan, rekening houdend met de verhouding m/n, niet te veel of te weinig in reeks A gevonden zijn. Verder merken wij op, dat, onder de voorwaarde $\underline{t}_1$ [1]) $= t_1$, door de voor $\underline{a}_1$ gevonden waarde ook de waarden van a_2, b_1 en b_2 bepaald zijn, zodat a_1 als de toetsingsgrootheid van deze toets beschouwd kan worden, ook indien men deze niet expliciet als zodanig gebruikt.

Nu kan men ieder experiment ook beschouwen als een waarneming van een stochastische grootheid, die twee „waarden", nl. „succes" of „mislukking", aan kan nemen en deze waarden kan men desgewenst aangeven met de symbolen z_1 en z_2. Men kan voor z_1 en z_2 ook twee getallen nemen, b.v. 0 en 1. De bovenbeschreven methode ondervindt hiervan geen enkele invloed. Men beschouwt dan dus de reeksen A en B als twee onafhankelijke reeksen van onafhankelijke waarnemingen van twee grootheden $\underline{x}$ (voor A) en $\underline{y}$ (voor B), die ieder de waarden z_1 en z_2 kunnen aannemen, met kans p_A (resp. p_B) op z_1 en $q_A = 1 - p_A$ (resp. $q_B = 1 - p_B$) op z_2. De getoetste hypothese, $p_A = p_B$, kan dan ook geformuleerd worden als: $\underline{x}$ en $\underline{y}$ bezitten dezelfde waarschijnlijkheidsverdeling. Dit is echter een hypothese, die men ook kan toetsen met behulp van een of andere toets voor twee steekproeven. Daar de onderzochte verdelingen verre van normaal zijn (alleen de waarden z_1 en z_2 kunnen worden aangenomen) komt de toets van Student niet in aanmerking; wel echter verdelingsvrije toetsen.

De vraag doet zich nu voor, of toepassing van een dergelijke toets voor twee steekproeven dezelfde oplossing van het probleem geeft als de methode der 2×2 -tabel. Uit M. G. Kendall [3] p. 43—44 en 165 blijkt reeds dat dit

1) Stochastische grootheden worden onderscheiden van getallen (b.v. van de waarden die zij bij een experiment aannemen) door hun symbolen te onderstrepen.

voor de toets van Wilcoxon inderdaad het geval is. In de volgende paragraaf zullen wij nu laten zien, dat dit voor een grote klasse van toetsen, waartoe behalve de toets van F. Wilcoxon [4], [7], [9] ook die van E. J. G. Pitman [5], M. E. Terry [6] en B. L. van der Waerden [8] behoren, het geval is.

2. Verdelingsvrije toetsen voor twee steekproeven

Beschouw twee onderling onafhankelijke steekproeven $x_1, x_2, \ldots, x_m$ en $y_1, y_2, \ldots, y_n$ $(m + n = N)$, die waarnemingen zijn van de stochastische grootheden $\underline{x}$ resp. $\underline{y}$. Stel dat de twee steekproeven tezamen genomen bestaan uit k verschillende waarden $z_1 < z_2 < \ldots < z_k$ en dat $\underline{x}$ (resp. $\underline{y}$) $\underline{a}_\nu$ (resp. $\underline{b}_\nu$) maal de waarde z_ν aanneemt. Laat verder $\underline{a}_\nu + \underline{b}_\nu = \underline{t}_\nu$ zijn ($\nu = 1, 2, \ldots, k$). De grootheden $\underline{t}_1, \underline{t}_2, \ldots, \underline{t}_k$ zijn dus de grootten der groepen gelijke waarnemingen en als er geen gelijke waarnemingen zijn dan is $\underline{t}_\nu = 1$ voor iedere ν en dan is $k = N$.

Wij beschouwen nu een klasse van verdelingsvrije toetsen voor twee steekproeven, waarvan de toetsingsgrootheid, voor het geval dat er geen gelijke waarnemingen zijn, geschreven kan worden in de vorm:

$$\underline{u} = \sum_{i=1}^{m} \varphi(\underline{r}_i, \underline{x}_i) \qquad \text{(zie [1], blz. 251).} \tag{2.1}$$

Hierin is $\underline{r}_i$ het rangnummer van $\underline{x}_i$ als de twee steekproeven tezamen genomen naar opklimmende grootte gerangschikt worden.

Verder beschouwen wij uitsluitend het geval, dat de kritieke zones op de gebruikelijke wijze worden opgebouwd uit grote en kleine waarden van $\underline{u}$.

De toetsen van Pitman, Wilcoxon, Terry en van der Waerden behoren hier alle toe. Voor deze toetsen geldt nl.:

(2.2) Pitman: $\varphi(r_i, x_i) = x_i$,

(2.3) Wilcoxon: $\varphi(r_i, x_i) = r_i$,

(2.4) Terry: $\varphi(r_i, x_i) = E\,\underline{Z}_{N,\,r_i}$,

(2.5) van der Waerden: $\varphi(r_i, x_i) = \psi\left(\dfrac{r_i}{N+1}\right)$,

waarin $E\,\underline{Z}_{N,r_i}$ de verwachting is van de r_i^e waarneming van een naar opklimmende grootte gerangschikte aselecte steekproef ter grootte van N uit een normale verdeling met gemiddelde 0 en spreiding 1 en $\psi(\epsilon)$ het ϵ-quantile van de normale verdeling met gemiddelde 0 en spreiding 1 is, dus

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\psi(\epsilon)} e^{-\frac{1}{2}x^2}\, dx = \epsilon. \tag{2.6}$$

Bij de laatstgenoemde drie toetsen hangt $\varphi(r_i, x_i)$ dus niet van x_i af, bij de eerste hangt $\varphi(r_i, x_i)$ niet van r_i af.

In die gevallen, waarin $\varphi(r_i, x_i)$ van x_i afhangt wordt de toets steeds voorwaardelijk uitgevoerd onder de voorwaarde dat de twee steekproeven tezamen bestaan uit de waarden $z_1, z_2, \ldots, z_k$. Alleen onder deze voorwaarde nl. is de verdeling van $\underline{u}$ onder de hypothese H_0, inhoudende dat $\underline{x}$ en $\underline{y}$ dezelfde verdeling bezitten, onafhankelijk van de verdeling van $\underline{x}$ en $\underline{y}$.

Indien er gelijke waarnemingen voorkomen wordt de toets voor alle gevallen voorwaardelijk uitgevoerd onder de voorwaarde $\underline{t}_\nu = t_\nu$ $(\nu = 1, 2, \ldots, k)$. Voor die gevallen, waarin $\varphi(r_i, x_i)$ niet van r_i afhangt kan (2.1) ongewijzigd gehandhaafd worden. Men kan dan dus schrijven

$$\underline{u} = \sum_{\nu=1}^{k} \underline{a}_\nu \varphi(z_\nu), \tag{2.7}$$

waarbij in het rechterlid dus alleen de aantallen $\underline{a}_1, \underline{a}_2, \ldots, \underline{a}_k$ (zie boven) stochastisch zijn.

Voor de overige gevallen, dus de gevallen waarin $\varphi(r_i, x_i)$ wel van r_i afhangt, gaat men als volgt te werk: de twee steekproeven worden tezamen naar opklimmende grootte gerangschikt (gelijke waarnemingen worden hierbij in een willekeurige volgorde geplaatst) en in deze volgorde worden de rangnummers $1, 2, \ldots, N$ toegekend. De rangnummers van de waarnemingen die gelijk zijn aan z_ν (dat zijn er t_ν) zijn dan

$$h_\nu + 1,\ h_\nu + 2, \ldots, h_\nu + t_\nu \qquad \left(h_\nu = \sum_{\mu=1}^{\nu-1} t_\mu\right). \tag{2.8}$$

Men kan nu verschillende methoden volgen, b.v.:

a. voor iedere ν het gemiddelde $\bar{r}_\nu$ berekenen van de rangnummers (2.8), dus

$$\bar{r}_\nu = h_\nu + \tfrac{1}{2}(t_\nu + 1). \tag{2.9}$$

Als bijdrage tot $\underline{u}$ van iedere waarneming van $\underline{x}$, die gelijk is aan z_ν neemt men dan $\varphi(\bar{r}_\nu, z_\nu)$ dus

$$\underline{u} = \sum_{\nu=1}^{k} \underline{a}_\nu\ \varphi(\bar{r}_\nu, z_\nu), \tag{2.10}$$

b. voor iedere ν het gemiddelde $\bar{\varphi}_\nu$ berekenen van

$$\varphi(h_\nu + 1, z_\nu),\ \varphi(h_\nu + 2, z_\nu), \ldots, \varphi(h_\nu + t_\nu, z_\nu).$$

Als bijdrage tot $\underline{u}$ van iedere waarneming van $\underline{x}$, die gelijk is aan z_ν neemt men dan $\bar{\varphi}_\nu$, dus

$$\underline{u} = \sum_{\nu=1}^{k} \underline{a}_\nu\ \bar{\varphi}_\nu \text{ [1]}. \tag{2.11}$$

1) Voor de toets van Wilcoxon zijn de methoden *a* en *b* identiek, daar bij deze toets $\varphi(r_i, x_i) = r_i$ is. De hier gedefiniëerde toetsingsgrootheid voor deze toets werd door Wilco-

Uit het bovenstaande volgt dat, voor de hier behandelde verdelingsvrije toetsen voor twee steekproeven, de toetsingsgrootheid geschreven kan worden in de vorm

$$\underline{u} = \sum_{\nu=1}^{k} \underline{a}_\nu g_\nu, \tag{2.12}$$

waarin $g_1, g_2, \ldots, g_k$ gegeven getallen zijn, die niet alle aan elkaar gelijk zijn. Passen wij zo'n toets toe op het waarnemingsschema van tabel I, dus op twee steekproeven die bestaan uit twee groepen gelijke waarnemingen, dan wordt de toetsingsgrootheid

$$\underline{u} = \underline{a}_1 g_1 + \underline{a}_2 g_2 = \underline{a}_1 (g_1 - g_2) + m g_2. \tag{2.13}$$

Hierin zijn alleen $\underline{u}$ en $\underline{a}_1$ stochastisch zodat $\underline{u}$ een lineaire functie is van $\underline{a}_1$, de toetsingsgrootheid van de methode der 2×2 -tabel. Dit betekent echter, dat grote waarden van $\underline{u}$ corresponderen met grote waarden van $\underline{a}_1$ en kleine waarden van $\underline{u}$ met kleine waarden van $\underline{a}_1$ (of omgekeerd, dat hangt af van het teken van $g_1 - g_2$).

Verder zijn de voorwaarden, waaronder de toetsen worden uitgevoerd identiek, nl. $\underline{t}_1 = t_1$ en $\underline{t}_2 = t_2$. Daar tenslotte in alle gevallen de kritieke zones op dezelfde wijze uit grote en kleine waarden van de toetsingsgrootheid worden opgebouwd, volgt hieruit, dat de toetsen voor het beschouwde geval alle identiek zijn.

Ter toelichting beschouwen wij de toets van Wilcoxon iets nader. Voor deze toets geldt (zie (2.9)):

$$g_1 = \tfrac{1}{2}(t_1 + 1), \quad g_2 = \tfrac{1}{2}(t_2 + 1) + t_1, \quad g_1 - g_2 = -\tfrac{1}{2}N, \tag{2.14}$$

dus (zie (2.13)):

$$\underline{u} = -\tfrac{1}{2} N\underline{a}_1 + \tfrac{1}{2} m(t_2 + 1) + mt_1. \tag{2.15}$$

Voor de grootheid $\underline{W}$ (zie voetnoot 1 blz. 160) geldt dus

$$\underline{W} - mn = -(\underline{a}_1 N - mt_1). \tag{2.16}$$

Uit het feit dat de toets, gebaseerd op $\underline{u}$, identiek is met de methode der 2×2 -tabel volgt dat de momenten van de verdeling van $\underline{u}$ gevonden kunnen worden uit die van $\underline{a}_1$ en omgekeerd.

---

xon ingevoerd in [9]. Men gebruikt voor deze toets ook wel de grootheid $\underline{U}$, die gedefiniëerd wordt als de som van het aantal paren waarnemingen (x_i, y_j) met $x_i > y_j$ en de helft van het aantal paren waarnemingen (x_i, y_j) met $x_i = y_j$ $(i = 1, 2, \ldots, m; j = 1, 2, \ldots, n)$. Hiervoor geldt

$$\underline{U} = \underline{u} - \tfrac{1}{2} m(m+1).$$

In [7] wordt voor deze toets de grootheid $\underline{W} = 2\underline{U}$ gebruikt.

Uit (2.16) volgt b.v.:

$$(2.17)\quad \sigma^2(\underline{a}_1 \mid t_1, H_0) = \frac{1}{N^2}\,\sigma^2\,(\underline{W} \mid t_1, H_0) = \frac{1}{N^2}\cdot\frac{mn\,\{N^3 - (t_1^3 + t_2^3)\}}{3N\,(N-1)} =$$

$$= \frac{m\,n\,t_1 t_2}{N^2\,(N-1)}. \qquad \text{(zie [2] en [7]).}$$

Opmerking:

Behalve de twee boven vermelde behandelingsmethoden a en b van gelijke waarnemingen zijn er nog andere denkbaar. Zolang deze ertoe leiden, dat aan gelijke waarnemingen gelijke bijdragen tot $\underline{u}$ worden toegekend, terwijl deze bijdragen voor de verschillende groepen gelijke waarnemingen niet alle aan elkaar gelijk zijn en bovendien niet stochastisch zijn, blijft het bovenstaande onveranderd geldig. Indien hieraan echter niet voldaan is, hetgeen b.v. optreedt, indien men binnen iedere groep van gelijken de beschikbare rangnummers of waarden van φ onder de leden der groep verloot — een methode die ook wel eens aanbevolen wordt — gaat de identiteit met de methode der 2×2 -tabel verloren. Deze methode verdient echter geen aanbeveling daar het onderscheidingsvermogen der toets er nadeling door beïnvloed wordt.

Literatuur

[1] van Dantzig, D. and J. Hemelrijk, Statistical methods based on few assumptions, Bull. of the Intern. Stat. Inst. **34** (1954), deel 2, 239—267.

[2] van Eeden, Constance, Methoden voor het vergelijken, toetsen en schatten van onbekende kansen, Statistica 7 (1953), 141—162.

[3] Kendall, M. G., Rank Correlation Methods, 2e druk, London (1955).

[4] Mann, H. B. and D. R. Whitney, On a test of whether one of two random variables is stochastically larger than the other, Ann. Math. Stat. **18** (1947), 50—60.

[5] Pitman, E. J. G., Significance tests which may be applied to samples from any populations I, Journ. Roy. Stat. Soc. Suppl. 4 (1937), 119—129.

[6] Terry, M. E., Some rank order tests, which are most powerful against specific parametric alternatives, Ann. Math. Stat. **23** (1952), 346—366.

[7] Wabeke, Ir Doraline en Constance van Eeden, Handleiding voor de toets van Wilcoxon, Rapport S 176 (M 65) van de Statistische Afdeling van het Mathematisch Centrum, Amsterdam, 1955.

[8] van der Waerden, B. L., Ein neuer Test für das Problem der zwei Stichproben, Mathem. Annalen **126** (1952), 93—107.

[9] Wilcoxon, F., Individual comparisons by ranking methods, Biometrics **1** (1945), 80—82.