Het Behrens-Fisher probleem

TECHNISCHE UNIVERSITEIT EINDHOVEN

Faculteit Wiskunde en Informatica

Memorandum COSOR 87 - 02

HET BEHRENS-FISHER PROBLEEM

by

H.N. LINSSEN

Eindhoven, Nederland

Januari 1987

HET BEHRENS-FISHER PROBLEEM

H.N. Linssen

Januari 1987

Inhoud

HET BEHRENS-FISHER PROBLEEM

1 Probleemstelling en notatie-afspraken

Bij het Behrens-Fisher probleem gaat het om betrouwbaarheidsuitspraken over het verschil in lokatie van twee ongespecificeerde normale populaties. Steekproeven $(x_1,\dots,x_n)$ en $(y_1,\dots,y_n)$ zijn gegeven. De interessante parameter is $\mu_x - \mu_y$.

De kansdichtheid respectievelijk kansverdeling van x in t wordt genoteerd als $p_x(t)$ respectievelijk $P_x(t)$.
De fiduciedichtheid respectievelijk verdeling van θ in t wordt genoteerd als $f_\theta(t)$ respectievelijk $F_\theta(t)$.
Kortheidshalve wordt ook gedefinieerd:

$$p(x) := p_x(x) \quad \text{en} \quad f(\theta) := f_\theta(\theta) \ .$$

$P(x)$ en $F(\theta)$ worden analoog gedefinieerd. Parameters worden zonodig achter ';' opgenomen. Voorbeelden: $p_x(t;\theta)$ en $f_\theta(t;x)$. Een of meer van de argumenten kan worden weggelaten omwille van de eenvoud. Zo kan $f_q(q;\hat{q})$ worden genoteerd als $f_q(q)$ of als $f(q)$ of als f_q, afhankelijk van de context.

2 Waarnemingsreductie

Voldoende is $\bar{x}$, $\bar{y}$, s_x^2, s_y^2. Omdat alleen het lokatie-*verschil* interessant is, is $\bar{x} - \bar{y}$ relevant, maar $\bar{x} + \bar{y}$ niet. Er geldt:

$$\bar{x} - \bar{y} \sim N\left(\mu_x - \mu_y \, , \, \frac{\sigma_x^2 + \sigma_y^2}{n}\right) ,$$

zodat *bij vaste* s_x^2 *en* s_y^2

$$t := \frac{\bar{x} - \bar{y} - (\mu_x - \mu_y)}{\sqrt{(s_x^2 + s_y^2)/n}} \sim N\left(0 \, , \, \frac{\sigma_x^2 + \sigma_y^2}{s_x^2 + s_y^2}\right) .$$

De 'spil' t kan ook opgevat worden als fiducievariabele bij vaste σ_x^2 en σ_y^2. Als we definiëren

$$\sigma_1^2 = \min(\sigma_x^2,\sigma_y^2) \quad \text{en} \quad \sigma_2^2 = \max(\sigma_x^2,\sigma_y^2)$$

en s_1^2 en s_2^2 analoog, dan geldt dus fiducieel:

$$t \sim N\left(0\,,\frac{\sigma_1^2+\sigma_2^2}{s_1^2+s_2^2}\right) .$$

Samenvattend:

$$p_{t|s_1^2,s_2^2}(t;\sigma_1^2,\sigma_2^2) = f_{t|\sigma_1^2,\sigma_2^2}(t;s_1^2,s_2^2) .$$

Van deze voorwaardelijke fiducieverdeling kan een onvoorwaardelijke (marginale) worden gemaakt door de verwachting ervan te bepalen met betrekking tot de simultane fiducieverdeling van (σ_1^2,σ_2^2).

3 De simultane verdeling van (σ_1^2,σ_2^2)

Zij $\nu = n-1$. De simultane kansverdeling van $\nu s_x^2/\sigma_x^2$ en $\nu s_y^2/\sigma_y^2$ is het produkt van twee χ_ν^2-verdelingen en wordt genoteerd als $\mathbb{C}(\cdot,\cdot)$, zodat:

$$p_{\nu s_x^2,\nu s_y^2}(t_x,t_y) = \frac{1}{\sigma_x^2\,\sigma_y^2}\,\mathbb{C}\left(\frac{t_x}{\sigma_x^2}\,,\frac{t_y}{\sigma_y^2}\right) \qquad (0 \leq t_x,t_y) .$$

Daaruit volgt meteen:

$$p_{\nu s_1^2,\nu s_2^2}(t_1,t_2) = \frac{1}{\sigma_1^2\,\sigma_2^2}\left\{\mathbb{C}\left(\frac{t_1}{\sigma_1^2}\,,\frac{t_2}{\sigma_2^2}\right) + \mathbb{C}\left(\frac{t_2}{\sigma_1^2}\,,\frac{t_1}{\sigma_2^2}\right)\right\} ,$$

waarbij $0 \leq t_1 \leq t_2$ en gebruik gemaakt is van de eigenschap:

$$\mathbb{C}(x,y) = \mathbb{C}(y,x) .$$

De simultane kansverdeling van $\hat{q} := s_1^2/s_2^2$ en $h := \nu s_2^2/\sigma_2^2$ wordt gegeven door:

$$p(h,\hat{q}) = \frac{h}{q}\left\{\mathfrak{C}\left(\frac{\hat{q}h}{q}, h\right) + \mathfrak{C}\left(\frac{h}{q}, \hat{q}h\right)\right\}$$

met $q := \sigma_1^2/\sigma_2^2$.

De reden om over te gaan op deze nieuwe variabelen is dat de marginale verdeling van $\hat{q}$ alleen van q afhangt en dat de fiducieverdeling van q daarmee kan worden bepaald (zie verder). De kansverdeling van h gegeven $\hat{q}$ kan dan dienen voor de bepaling van de fiducieverdeling van h gegeven q.

Opmerking. Als kansvariabele opgevat varieert h in de steekproefruimte via s_2^2 bij vaste σ_2^2. Als fiducievariabele opgevat varieert h in de parameterruimte bij vaste s_2^2. De grootheid h is een spil (pivot) van de fiduciële inversie. De kansverdeling is voorwaardelijk $\hat{q}$ en de fiducieverdeling is voorwaardelijk q. Er geldt:

$$f_{h|q}(h;\hat{q}) = p_{h|\hat{q}}(h;q) \ . \qquad \square$$

Er geldt:

$$\begin{aligned} p_{h|\hat{q}}(h) &= p_{h,\hat{q}}(h,\hat{q})/p_{\hat{q}}(\hat{q}) \\ &\propto h^{\nu-1}(e^{-ha/2} + e^{-hb/2}) \ , \end{aligned}$$

met

$$a = 1 + \hat{q}/q \quad \text{en} \quad b = 1/q + \hat{q} \ .$$

Er geldt ook (zie eerder):

$$f_{t|q,h} = N(0,c\nu/h) \quad \text{met} \quad c = (1+q)/(1+\hat{q}) \ ,$$

zodat

$$\begin{aligned} f_{t|q}(t) &= \int_0^\infty f_{t|q,x}(t)\ f_{h|q}(x)\,dx \\ &\propto \int_0^\infty x^{\frac{1}{2}} \exp\frac{-xt^2}{2c\nu} \cdot x^{\nu-1}\left(\exp\frac{-ax}{2} + \exp\frac{-bx}{2}\right)dx \end{aligned}$$

$$\propto a^{-\nu-\frac{1}{2}}\left(1+\frac{t^2}{ac\nu}\right)^{-\nu-\frac{1}{2}} + b^{-\nu-\frac{1}{2}}\left(1+\frac{t^2}{bc\nu}\right)^{-\nu-\frac{1}{2}}$$

$$\propto a^{-\nu}\, t_{2\nu,a_1} + b^{-\nu}\, t_{2\nu,a_2} ,$$

waarbij $a_1 = \sqrt{ac/2}$ en $a_2 = \sqrt{bc/2}$ en $t_{\nu,\sigma}(t)$ de dichtheid in t is van een geschaalde Student-verdeelde grootheid met ν vrijheidsgraden en schaalparameter σ, met andere woorden: Als t standaard-Student verdeeld is dan is $\sigma.t$ geschaald Student-verdeeld met schaalparameter σ.
De fiducieverdeling is compleet (integreert tot 1), dus:

$$f_{t|q}(t) = \frac{b^{\nu}\, t_{2\nu,a_1}(t) + a^{\nu}\, t_{2\nu,a_2}(t)}{a^{\nu} + b^{\nu}} .$$

De fiducieverdeling van t gegeven q is dus een mengsel van twee geschaalde Student-verdelingen.
Als q = 1 geldt $a = b = 1+\hat{q}$ en $c = 2/(1+\hat{q})$ en dus

$$f_{t|q=1} = t_{2\nu,1} .$$

Om tenslotte de marginale verdeling van t te vinden is het nodig de fiducieverdeling van q te kennen.

4 De verdeling van $q = \sigma_1^2/\sigma_2^2$

Er geldt:

$$P(\hat{q}) = P(s_x^2/s_y^2 < \hat{q} \text{ of } s_x^2/s_y^2 > 1/\hat{q}) = P_{f_\nu}(q\hat{q}) + P_{f_\nu}(\hat{q}/q) ,$$

waarbij f_ν een F-verdeelde kansvariabele is met ν vrijheidsgraden voor teller en noemer.
De afgeleide naar $\hat{q}$ levert de met q geparametriseerde kansdichtheid van $\hat{q}$ op, terwijl (min de) afleide naar q de met $\hat{q}$ geparametriseerde fiduciedichtheid van q oplevert:

$$f(q) = \frac{\hat{q}}{q^2}\, p_{f_\nu}\!\left(\frac{\hat{q}}{q}\right) - \hat{q}p_{f_\nu}(q\hat{q}) \qquad (0 \le q \le 1) .$$

<u>Let op</u>! Als $\hat{q} = 1$ dan is $f(q) = 0$.

De fiducieverdeling is dus niet compleet. Er geldt

$$\text{fid} := \int_0^1 f_q(x)dx = \int_{\hat{q}}^{1/\hat{q}} p_{f_\nu}(x)dx < 1 \quad !!$$

De fiducieverdeling voor q is compleet te maken door de niet-toegewezen fiducie (gelijk aan 1 - fid) toe te wijzen aan q = 1. De fiducieverdeling is dan singulier in q = 1 met singuliere fiducie 1 - fid.
Met betrekking tot de aldus gedefinieerd fiducieverdeling wordt nu de marginale verdeling van t bepaald. Tenslotte dus:

$$f(t) = \int_0^1 f_{t|q}(t)f(q)dq + (1-\text{fid})\ t_{2\nu,1}(t) \ .$$

Deze verdeling geven we de naam '*Wilkinson-verdeling*', omdat G.N. Wilkinson deze oplossing van het Behrens-Fisher probleem op verschillende plaatsen gesuggereerd heeft, o.a. in [1] en in een brief aan de auteur, echter zonder een gedetailleerde uitwerking te publiceren.

Speciale gevallen

- Als $\hat{q} = 1$ dan is $a = b = 1 + \hat{q}$, $c = 2/(1+\hat{q})$ en dus $f(t) = t_{2\nu,1}(t)$.
- Als $\hat{q} \downarrow 0$ dan fid $\uparrow$ 1 en

$$f_{t\ q} \to t_{2\nu,a_1} \ , \qquad \text{met } a_1^2 = a/2 \ .$$

Omdat $f_{\hat{q},q}(x) \to p_{f_\nu}(x)$ geldt

$$f(t) \to \int_{\hat{q}}^{\infty} \left(1 + \frac{t^2}{\nu(1+x)}\right)^{-\nu-\frac{1}{2}} \frac{x^{\frac{1}{2}\nu-1}}{(1+x)^\nu}\, dx \ .$$

Substitutie van $y = x/(z+x)$ met $z = 1 + t^2/\nu$ leert:

$$f(t) \to t_{\nu,1}(t) \ .$$

5 Resumé

Waarnemingsmodel:

$$x_i \sim N(\mu_x,\sigma_x^2) \;;\quad y_i \sim N(\mu_y,\sigma_y^2) \qquad (i = 1,\dots,n)\ .$$

Voldoende grootheden: $\bar{x}$, $\bar{y}$, s_x^2 en s_y^2.
De parameter van interesse is $\mu_x - \mu_y$. Daarom is relevant:

$$\bar{x} - \bar{y}\ ,\quad s_1^2 \ \text{ en } \ s_2^2\ .$$

Irrelevant is:

$$\bar{x} + \bar{y} \quad \text{en} \quad \operatorname{sgn}\{\ln(s_x^2/s_y^2)\}\ .$$

Voor de fiducievariabele

$$t := \frac{\mu_x - \mu_y - (\bar{x} - \bar{y})}{\sqrt{(s_1^2 + s_2^2)/n}}$$

geldt:

$$t \sim N\left(0\,, \frac{(\sigma_1^2 + \sigma_2^2)}{(s_1^2 + s_2^2)}\right)$$

bij gegeven σ_1^2 en σ_2^2.
De steekproefdichtheid wordt gefactoriseerd volgens:

$$p(\hat{q},s_2^2;q,\sigma_2^2) = p(\hat{q};q)p(s_2^2|\hat{q};q,\sigma_2^2)\ .$$

Uit $p_{\hat{q}}$ en de waarde van $\hat{q}$ volgt door inversie de (incomplete) fiducieverdeling voor q. Deze wordt compleet gemaakt door de niet-toegewezen fiducie toe te wijzen aan q = 1. Uit $p_{s_2^2|\hat{q}}$ en s_2^2 volgt de fiducieverdeling voor σ_2^2 gegeven q. Het produkt van deze twee verdelingen is de simultane verdeling van q en σ_2^2. De marginale verdeling van t is nu de verwachting van de verdeling van t gegeven q en σ_2^2 met betrekking tot deze simultane verdeling.

6 De Behrens-verdeling

Indien simultane uitspraken vereist zijn voor μ_x en μ_y dan is de simultane verdeling van μ_x en μ_y nodig. De minimaal voldoende grootheid is $(\bar{x},\bar{y},s_x^2,s_y^2)$, die tevens relevant is voor de simultane uitspraken.
Het is bekend dat:

$$\mu_x \sim \bar{x} + t_\nu \sqrt{s_2^2/n} \quad \text{en} \quad \mu_y \text{ analoog verdeeld.}$$

De simultane dichtheid is het produkt van de marginale dichtheden.
De marginale verdeling daaruit bepaald van

$$t = (\mu_x - \mu_y - (\bar{x} - \bar{y}))/\sqrt{(s_x^2 + s_y^2)/n}$$

wordt de Behrens-verdeling genoemd en is getabelleerd in New Cambridge elementary statistical Tables (1984) [2]. De verdeling kan worden geschreven als:

$$F'(t;k) = \int_{-\infty}^{\infty} t_{\nu,1}(x)\, P_{t_{\nu,1}}(t\sqrt{1+k} - x\sqrt{k})dx$$

met $k = s_x^2/s_y^2$ of s_y^2/s_x^2 naar believen.

De dichtheid is een mengsel van (standaard) Student verdelingen.
De Behrens-verdeling geeft onnodig conservatieve intervallen voor $\mu_x - \mu_y$ omdat irrelevante waarnemingscomponenten niet zijn geëlimineerd. Enkele percentagepunten van de Behrens-verdeling worden in onderstaande tabel naast de overeenkomstige percentagepunten van de in deze notitie afgeleide Wilkinson-verdeling.

Tabel

97.5% punten van de
Behrens (B) en Wilkinson (W) verdelingen

<table>
<tr><td></td><td colspan="7">arctan s_1/s_2</td></tr>
<tr><td></td><td>0°</td><td colspan="2">15°</td><td colspan="2">30°</td><td colspan="2">45°</td></tr>
<tr><td>$\nu_{\downarrow}$</td><td>B = W</td><td>B</td><td>W</td><td>B</td><td>W</td><td>B</td><td>W</td></tr>
<tr><td>1</td><td>12.71</td><td>15.56</td><td>13.57</td><td>17.36</td><td>9.50</td><td>17.97</td><td>4.30</td></tr>
<tr><td>2</td><td>4.30</td><td>4.41</td><td>4.39</td><td>4.56</td><td>4.11</td><td>4.62</td><td>2.78</td></tr>
<tr><td>3</td><td>3.18</td><td colspan="2">3.19</td><td>3.23</td><td>3.13</td><td>3.24</td><td>2.45</td></tr>
<tr><td>4</td><td>2.78</td><td colspan="2">2.77</td><td>2.78</td><td>2.75</td><td>2.79</td><td>2.31</td></tr>
<tr><td>5</td><td>2.57</td><td colspan="2">2.56</td><td>2.56</td><td>2.55</td><td>2.57</td><td>2.23</td></tr>
<tr><td>6</td><td>2.45</td><td colspan="2">2.44</td><td>2.44</td><td>2.43</td><td>2.44</td><td>2.18</td></tr>
<tr><td>7</td><td>2.36</td><td colspan="2">2.36</td><td colspan="2">2.35</td><td>2.35</td><td>2.14</td></tr>
<tr><td>8</td><td>2.31</td><td colspan="2">2.30</td><td colspan="2">2.29</td><td>2.29</td><td>2.12</td></tr>
<tr><td>10</td><td>2.23</td><td colspan="2">2.22</td><td colspan="2">2.22</td><td>2.21</td><td>2.09</td></tr>
<tr><td>12</td><td>2.18</td><td colspan="2">2.17</td><td colspan="2">2.17</td><td>2.17</td><td>2.06</td></tr>
<tr><td>24</td><td>2.06</td><td colspan="2">2.06</td><td colspan="2">2.06</td><td>2.06</td><td>2.01</td></tr>
<tr><td>∞</td><td>1.96</td><td colspan="2">1.96</td><td colspan="2">1.96</td><td colspan="2">1.96</td></tr>
</table>

7 Voorwaardelijke eigenschappen

Het is niet moeilijk te zien dat $p_{t|\hat{q}}(\cdot;q) = f_{t|q}(\cdot;\hat{q})$. Immers:

$$p_{t|\hat{q}}(\cdot;q) = \int_0^\infty p_{t|\hat{q},h}(\cdot;q,x)\,p_{h|\hat{q}}(x;q)\,dx =$$

$$= \int_0^\infty f_{t|q,h}(\cdot;\hat{q},x)\,f_{h|q}(x;\hat{q})\,dx = f_{t|q}(\cdot;\hat{q}) \;.$$

De steekproefverdeling van t voorwaardelijk $\hat{q}$ is dus ook een mengsel van twee geschaalde Student-verdelingen met 2ν vrijheidsgraden. Er geldt

$$P_{t|\hat{q}}(t) = wP_{t_{2\nu}}(t/a_1) + (1-w)P_{t_{2\nu}}(t/a_2) ,$$

waarbij

$$w = (1 + (a/b)^{\nu})^{-1} .$$

Bij gegeven $\hat{q}$ zijn de (on)waarheidskansen van betrouwbaarheidsintervallen ongelijk aan de nominale (on)betrouwbaarheid en afhankelijk van q.
Voor Wilkinson- en Behrens-intervallen zijn in de onderstaande tabel de op $\hat{q}$ geconditioneerde onwaarheidskansen van tweezijdige 95%-betrouwbaarheidsintervallen voor ν = 1 weergegeven.

<u>onwaarheidskansen</u> % (voorwaardelijk $\hat{q}$, afhankelijk q)
van op de Wilkinson- en Behrens-verdeling gebaseerde
tweezijdige betrouwbaarheidsintervallen
(nominale onbetrouwbaarheid = 5%)

Wilkinson

$\hat{q}\downarrow$ \ q →	1	5^{-1}	5^{-2}	5^{-3}
1	5.0	8.5	24.0	52.5
5^{-1}	0.8	1.1	2.9	10.3
5^{-2}	0.5	0.6	0.9	2.4
5^{-3}	0.5	0.5	0.6	0.9

Behrens

$\hat{q}\downarrow$ \ q →	1	5^{-1}	5^{-2}	5^{-3}
1	0.3	0.6	2.0	8.6
5^{-1}	0.4	0.5	1.4	5.6
5^{-2}	0.4	0.4	0.6	1.7
5^{-3}	0.5	0.5	0.6	0.9

Behrens-intervallen zijn soms aanzienlijk conservatiever dan Wilkinson-intervallen. De deelverzameling van de steekproefruimte gegeven door $\hat{q}$ = 1 is negatief semi-relevant voor Wilkinson-intervallen.
Bij iedere $\hat{q}$ is de verwachte onwaarheidskans met betrekking tot de fiducieverdeling van q gelijk aan de nominale onbetrouwbaarheid.
Welch en Aspin [3] construeerden betrouwbaarheidsintervallen zodanig dat, voor iedere q, de verwachte onwaarheidskans met betrekking tot de steekproefverdeling van $\hat{q}$ gelijk is aan de nominale onbetrouwbaarheid. Dan ontstaat de artefact dat $\hat{q}$ = 1 een negatief relevante deelverzameling van de teekproefruimte definieert (zie tabel).

onwaarheidskansen % voor Wilkinson- en Welch-intervallen

(nominale onbetrouwbaarheid 5%, $\nu = 8$, $\hat{q} = 1$)

$q \rightarrow$	1	2^{-1}	2^{-2}	2^{-3}
Wilkinson	5.0	6.2	10.9	20.1
Welch-Aspin	5.4	6.8	11.6	21.0

Voor $\nu = 8$ zijn de verschillen niet groot. Voor kleinere aantal vrijheidsgraden zijn de Welch-waarden helaas niet getabelleerd. Tabellen in [4].
In het geval dat $\hat{q} = 1$, zijn de Welch-Aspin intervallen altijd bevat in de intervallen die je krijgt als je mag aannemen dat $q = 1$, hetgeen strijdig is met de intuïtie.
Een approximatie voor de Welch-Aspin-tabelwaarden is gegeven door Welch [5]. Daarbij wordt de gewone Student-verdelingstabel gebruikt met een benaderend aantal vrijheidsgraden $\hat{\nu}$ gegeven door

$$1 \leq \frac{\hat{\nu}}{\nu} = \frac{(1+\hat{q})^2}{1+\hat{q}^2} \leq 2 .$$

Er geldt: $\hat{\nu} = \nu \Leftrightarrow \hat{q} = 0$ en $\hat{\nu} = 2\nu \Leftrightarrow \hat{q} = 1$.

Als $\hat{\nu}$ niet geheel is kan er lineair geïnterpoleerd worden tussen de Student-tabelwaarden voor de naburige gehele aantallen vrijheidsgraden.
De voorwaardelijke eigenschappen van op de lineair geïnterpoleerde benaderende t-waarde van Welch gebaseerde tweezijdige betrouwbaarheidsintervallen is vergeleken met die van de Wilkinson-intervallen (zie tabel).

onwaarheidskansen in % voor Wilkinson- en benaderende Welch-intervallen
(nominale onbetrouwbaarheid = 5%, ν = 1)

Wilkinson

$\hat{q}\downarrow$ \ $q\rightarrow$	1	5^{-1}	5^{-2}	5^{-3}
1	5.0	8.5	24.0	52.5
5^{-1}	0.8	1.1	2.9	10.3
5^{-2}	0.5	0.6	0.9	2.4
5^{-3}	0.5	0.5	0.6	0.9

Welch-benadering

$\hat{q}\downarrow$ \ $q\rightarrow$	1	5^{-1}	5^{-2}	5^{-3}
1	5.0	8.5	24.0	52.5
5^{-1}	1.1	1.6	4.2	14.1
5^{-2}	0.7	0.8	1.2	3.1
5^{-3}	0.6	0.6	0.7	1.0

Het oordeel of de benadering van Welch een bevredigende approximatie levert van de Wilkinson-waarden is niet zonder meer te geven en hangt af van het criterium dat gekozen wordt.

8 Referenties

[1] G.N. Wilkinson, 1977, On Resolving the Controversy in Statistical Interference, J.R. Statist. Soc. B, 39, 119-171

[2] New Cambridge Elementary Statistical Tables, 1984, D.V. Lindley & W.F. Scott, Cambridge University Press.

[3] A.A. Aspin, 1949, Tables for use in comparisons whose accuracy involves two variances, separately estimated. Biometrika 36, 1949, 290-293.

[4] Biometrika Tables for Statisticians, 1970, E.S. Pearson & H.O. Hartley, Cambridge University Press.

[5] B.L. Welch, 1949, Appendix by [3], 293-296.