S // S-ARCHIEF

AW

AFDELING BEWERKING WAARNEMINGSUITKOMSTEN
VAN DE
CENTRALE ORGANISATIE VOOR TOEGEPAST NATUURWETENSCHAPPELIJK ONDERZOEK

Antwoord aangehecht

BUREAU: Koningskade 12
Telefoon 776090*
Postcheque- en Gironummer ten name van
de Centrale Organisatie T. N. O.: 55305

's-GRAVENHAGE, 7 Maart 1949
Verzoeke bij beantwoording datum en
nummer van dit schrijven aan te halen.

No. 49 S 92

Bericht op schrijven van:

Onderwerp:

Bijlagen:

Aan de Hooggeleerde Heer Prof. Dr D. van Dantzig,
Mathematisch Centrum,
2e Boerhaavestraat 49,
A M S T E R D A M.

Hooggeachte Professor van Dantzig,

De Heer Drion meent de oplossing gevonden te hebben voor het probleem van de "ijklijn", dat hij U een paar weken geleden mocht voorleggen. Ik zou het zeer op prijs stellen, indien U Uw oordeel hierover zoudt willen geven. Hieronder volgt de uiteenzetting van de Heer Drion.

De oplossing is gebaseerd op de t-test, zoals deze door CRAMÉR op blz. 237 wordt uiteengezet; dit wijkt dus enigszins af van WELCH's ideeën.

Om de gevolgde methode duidelijk te doen uitkomen, moge eerst het overeenkomstige één-dimensionale probleem worden behandeld.

Uit een normaal universum met gemiddelde μ en spreiding σ is een steekproef van n individuen getrokken. Het gemiddelde van de steekproef bedraagt $\frac{1}{n}\Sigma x = \bar{x}$ en de variance $\frac{1}{n-1}\Sigma(x - \bar{x})^2 = s^2$. Gevraagd wordt de waarschijnlijkheid dat een nieuwe onafhankelijke waarneming aan één individu uit dit universum een waarde ξ zal geven, zodanig dat $\bar{x} - \xi \geq \partial$. De grootheden $\bar{x} - \mu$ en $\xi - \mu$ zijn onafhankelijk verdeeld met gemiddelde nul en met spreiding $\frac{\sigma}{\sqrt{n}}$ respectievelijk σ. Het verschil $(\bar{x} - \mu) - (\xi - \mu) = \bar{x} - \xi = \partial$ is derhalve normaal verdeeld met gemiddelde nul en spreiding $\sigma\sqrt{1 + \frac{1}{n}}$. De waarde $\frac{s^2}{\sigma^2}$ is verdeeld volgens een χ^2-verdeling met (n - 1) graden van vrijheid.

Typ.:
Coll.:
Contr.:

-2-

Het quotient

$$t' = \frac{\dfrac{(\bar{x} - \mu) - (\xi - \mu)}{\sigma\sqrt{1 + 1/n}}}{\dfrac{s}{\sigma}} = \frac{\lambda}{s}\sqrt{\frac{n}{n + 1}}$$

is derhalve verdeeld volgens de t-verdeling met $(n - 1)$ graden van vrijheid.

Het twee-dimensionale probleem, het eigenlijke probleem van de ijklijn, kan op analoge wijze behandeld worden. Stel dat de ware vergelijking van de ijklijn $y = \alpha + \beta\,(x - \bar{x})$ is. Hierbij zijn de x-waarden nauwkeurig bekende grootheden, de y-waarden de corresponderende aflezingen van het te ijken instrument. Aangenomen wordt, dat de y-waarden bij vaste x normaal verdeeld zijn met gemiddelde $\alpha + \beta\,(x - \bar{x})$ en met spreiding σ. Door de vergelijking in bovenstaande vorm te schrijven wordt bereikt, dat de functies van x, waarmede de coëfficienten α en β vermenigvuldigd zijn, nl. 1 en $(x - \bar{x})$, orthogonaal zijn. Dientengevolge is de covariance tussen de "maximum-likelihood"- schattingen voor α en β gelijk aan nul.

asymptotisch

De ijklijn wordt vastgelegd door telkens enige bepalingen te doen bij een aantal verschillende waarden van x. Het totaal aantal waarnemingen bedraagt n. Door middel van de methode der kleinste kwadraten worden de beste schattingen (maximum-likelihood schattingen) $\hat{a}$ en $\hat{b}$ voor α en β gevonden, benevens de schatting $s^2 = \frac{1}{n - 2}\sum\left\{y - \hat{a} - \hat{b}(x - \bar{x})\right\}^2$ voor de variance σ^2.

$\frac{1}{n}$

De vormen $(\hat{a} - \alpha)$ en $(\hat{b} - \beta)$ zijn beide onafhankelijk en normaal verdeeld met gemiddelde nul en variance $\frac{\sigma^2}{n}$ respectievelijk $\frac{\sigma^2}{\sum(x - \bar{x})^2}$. De waarde $\frac{s^2}{\sigma^2}$ is nu verdeeld volgens een χ^2 verdeling met $(n - 2)$ graden van vrijheid, onafhankelijk van $(\hat{a} - \alpha)$ en $(\hat{b} - \beta)$.

asymptotisch

Het probleem is nu, om de waarschijnlijkheid te bepalen, dat indien bij een waarde X een nieuwe bepaling gedaan wordt, hierbij een uitkomst $\geq \eta$ gevonden wordt.

De verdeling van $y - \alpha - \beta\,(X - \bar{x})$ is normaal met gemiddelde nul en variance σ^2. De verdeling van $(\hat{a} - \alpha) + (\hat{b} - \beta)(X - \bar{x})$ is normaal met gemiddelde nul en variance $\sigma^2\left(\frac{1}{n} + \frac{(X - \bar{x})^2}{\sum(x - \bar{x})^2}\right)$.

De verdelingen van $(\hat{a} - \alpha) + (\hat{b} - \beta)(X - \bar{x})$, van $\eta - \alpha - \beta\,(X - \bar{x})$ en van $s^2 = \frac{1}{n - 2}\sum(y - \hat{a} - \hat{b}x)^2$ zijn onderling onafhankelijk. Derhalve is de verdeling van $\left[(\hat{a} - \alpha) + (\hat{b} - \beta)(X - \bar{x})\right] - \left[\eta - \alpha - \beta\,(X - \bar{x})\right] = \hat{a} + \hat{b}\,(X - \bar{x}) - \eta$ normaal met gemiddelde nul en variance

asympt.

$$\sigma^2\left(\frac{1}{n} + \frac{(X - \bar{x})^2}{\sum(x - \bar{x})^2} + 1\right)$$

$\hat{\alpha} = \bar{y}$

$\hat{\beta} = \frac{s_y}{s_x} r_{xy}$

$\hat{\sigma}^2 = \hat{Q} = s_y^2(1 - r_{xy}^2)$

$Q = \frac{1}{n}\sum(y_i - \alpha - \beta(x_i - \bar{x}))^2 = (\bar{y} - \alpha)^2 + s_y^2 - 2\beta s_x s_y r_{xy} + \beta^2 s_x^2$

$L = -\frac{n}{2}\left(\frac{Q}{\sigma^2} + \ln \sigma^2\right)$

$-\frac{\partial^2 L}{\partial \alpha^2} = \frac{n}{\sigma^2}$ $\quad -\frac{\partial^2 L}{\partial \beta^2} = \frac{n}{\sigma^2} s_x^2$ $\quad -\frac{\partial^2 L}{\partial \alpha \partial \beta} = 0$

Het quotient

$$t'' = \frac{\dfrac{\left[(\hat{a} - \alpha) + (\hat{b} - \beta)(X - \bar{x})\right] - \left[\eta - \alpha - \beta\,(X - \bar{x})\right]}{\sigma\sqrt{\left(\dfrac{1}{n} + \dfrac{(X - \bar{x})^2}{\Sigma(x - \bar{x})^2} + 1\right)}}}{\dfrac{s^2}{\sigma^2}}$$

$$= \frac{\hat{a} + \hat{b}\,(X - \bar{x}) - \eta}{s\sqrt{\left(\dfrac{1}{n} + \dfrac{(X - \bar{x})^2}{\Sigma(x - \bar{x})^2} + 1\right)}}$$

zal derhalve verdeeld zijn volgens de t-verdeling met $(n - 2)$ graden van vrijheid.

Indien men met behulp van de ijklijn de bij η behorende waarde van X bepalen wil, zal men hiervoor in de bovenstaande formule die waarde van t" moeten invullen, die een kans P = 0,5 heeft om door het toeval overschreden te worden, d.w.z. t" = 0. Men vindt dan $(X - \bar{x}) = \frac{\eta - \hat{a}}{\hat{b}}$, d.w.z. de mediane waarde van X wordt gevonden bij het snijpunt van $y = \eta$ en de ijklijn. Confidence-intervals voor X vindt men, door de waarde van t", behorende bij $P = \varepsilon$ en $P = 1 - \varepsilon$ te bepalen en door uit de kwadratische vergelijking X op te lossen. Hierbij zal alleen de wortel overeenkomende met de positieve waarde van $\sqrt{\frac{1}{n} + \frac{(X - \bar{x})^2}{\Sigma(x - \bar{x})^2} + 1}$ betekenis hebben.

Indien de ijklijn kromlijnig is en men geen aanleiding heeft om op theoretische gronden aan een bepaalde vorm de voorkeur te geven, dan kan men proberen een tweede of eventueel een hogere graads parabool aan te passen. Indien men de formule in de vorm $y = \alpha_o P_o + \alpha_1 P_1 + \alpha_2 P_2 + \ldots + \alpha_k P_k$ schrijft, waarbij de P_i's orthogonale polynomia zijn, dan zal als beste schatting $\hat{a}_i$ voor α_i een in y lineaire vorm gevonden worden; de vormen $\hat{a}_i$ zijn normaal verdeeld en onderling onafhankelijk. De variance van $\hat{a}_o P_o(X) + \hat{a}_1 P_1(X) + \hat{a}_2 P_2(X) + \ldots + \hat{a}_k P_k(X)$ kan dan gevonden worden als de som Var. $(\hat{a}_o)\{P_o(X)\}^2$ + Var. $(\hat{a}_1)\{P_1(X)\}^2 + \ldots +$ Var. $(\hat{a}_k)\{P_k(X)\}^2$; deze variance gedeeld door σ^2 is verdeeld als een χ^2 verdeling met $(n - k - 1)$ graden van vrijheid, onafhankelijk van $(\hat{a}_i - \alpha_i)$.

Men kan derhalve op geheel analoge wijze te werk gaan als in het lineaire geval.

Met bij voorbaat onze dank voor Uw hulp, verblijf ik met de meeste

hoogachting
De Afdeling Bewerking
Waarnemingsuitkomsten
het Hoofd

(Th.J.D.Erlee).

30 Maart 1949.

Aan: T.N.O.,
Afd. Bewerking Waarnemingsuitkomst
Koningskade 12,
's-GRAVENHAGE.

Mijne Heren,

Tot mijn spijt had ik, mede tengevolge van mijn verhuizing naar Valeriusstr.58, niet eerder gelegenheid Uw schrijven dd 7 dezer in behandeling te nemen. Ik heb thans de beschouwingen van Dr. Erion nagegaan en ik kan daarover het volgende mededelen:

De methode van aannemelijkste parameterschatting is correct toegepast met dien verstande dat zij in de noemers van de voor s^2 gevonden uitdrukkingen op bladzijde 1 en 2 $\frac{1}{n}$ in plaats van $\frac{1}{n-1}$ resp. $\frac{1}{n-2}$ geeft. Dit is op zichzelf geen bezwaar daar de methode overigens ook resultaten geeft die slechts asymptotisch voor $n \rightarrow$ oneindig juist zijn. Dit dient echter wel in aanmerking genomen te worden, daar dientengevolg (zoals bij iedere toepassing van genoemde methode) in het eind resultaat, bijv. in de berekening van t'' afwijkingen van de relatieve orde $\frac{1}{n}$ kunnen optreden, die somtijds tot nog grotere afwijkingen in de berekende waarschijnlijkheden kunnen leiden. Het is mogelijk dat deze afwijkingen bij de gebruikte noemers $\frac{1}{n-1}$ resp. $\frac{1}{n-2}$ iets kleiner zijn dan zij bij de noemer $\frac{1}{n}$ zouden zijn, maar dit zou afzonderlijk dienen te worden nagega
In verband met het bovenstaande dient enkele malen, bijv. op blz.2 alinea 2 laatste regel, alinea 4 eerste regel, alinea derde regel en blz. 3 alinea 1 voorlaatste regel het woord <u>asymptotisch</u> te worden tussengevoegd. De eerste zin van pag.3 aline 2 is niet geheel duidelijk.

De voorgaande opmerkingen zijn in het bijzonder van beteken met betrekking tot de toepassing der methode van de betrouwbaarheidsgrens op pag.3, daar tengevolge van de ~~enontbeerlijkheid~~ onnauwkeurigheid, van de orde $\frac{1}{n}$ de onbetrouwbaarheid te geflatteerd kan zijn voorgesteld. Een nauwkeuriger taxatie van de hiervoor aan te brengen correcties zou echter meer tijd vergen.

In de hoop U hiermede van dienst te zijn geweest en steeds gaarne tot verdere adviezen bereid, teken ik

Hoogachtend,

Prof. Dr. D. van Dantzig

Opmerking: zie ook Cramér, blz. 549-551.
Blijkbaar is de methode <u>wel</u> voor kleine n juist, als de goede factoren $\frac{1}{n-1}$ en $\frac{1}{n-2}$ gebruikt worden. J. Th. Runn.
Mei 1955

Antwoord-rapport aangehecht. S 12 ARCHIEF

AFDELING BEWERKING WAARNEMINGSUITKOMSTEN
VAN DE
CENTRALE ORGANISATIE VOOR TOEGEPAST NATUURWETENSCHAPPELIJK ONDERZOEK

BUREAU: Koningskade 12
Telefoon 776090*
Postcheque- en Gironummer ten name van
de Centrale Organisatie T. N. O.: 55305

'S-GRAVENHAGE, 15 Maart 1949
Verzoeke bij beantwoording datum en nummer van dit schrijven aan te halen.

No. 49 S 108

Bericht op schrijven van:

Onderwerp:

Bijlagen:

Aan de Hooggeleerde Heer Prof. Dr D. van Dantzig,
Mathematisch Centrum,
2e Boerhaavestraat 49,
A M S T E R D A M.

Zeer geachte Professor van Dantzig,

Ten vervolge op mijn brief van 7 Maart jl. No. 49 S 92 zend ik U hierbij een beschouwing van de Heer Drion over het berekenen van een ijklijn, indien aangenomen moet worden, dat in beide variabelen een fout voorkomt, zonder dat de verhouding tussen de twee standaarddeviaties bekend is. Wij zouden het zeer op prijs stellen Uw oordeel over deze oplossing te mogen vernemen. Hier volgt thans het betoog van de Heer Drion.

Voor het aanpassen van een rechte lijn aan een reeks punten, welke zowel in de abscis als in de ordinaat met een fout belast zijn, is, voor zover kon worden nagegaan, nog steeds geen bevredigende oplossing gevonden, tenzij à priori ten minste de verhouding tussen de varianties voor beide richtingen bekend is. De beste oplossing is wellicht die, welke is gegeven door A. WALD in "The fitting of straight lines if both variables are subject to error", Annals of Mathematical Statistics 1940. Vol. XI blz. 284. Een ernstig bezwaar van de oplossing lijkt echter, dat de waargenomen grootheden in twee groepen verdeeld moeten worden en dat de uitkomst, tenzij het aantal punten zeer groot is, dus in het algemeen van deze verdeling zal afhangen. De oplossing, die hieronder gegeven wordt, berust in wezen op het beginsel der maximum-likelihood.

Typ.:
Coll.:
Contr.:

11712 · '48

Gegeven zij een aantal van n punten (x_1,y_1) - - - (x_n,y_n); de grootheden x_i en y_i zijn waarnemingen van (onbekende) grootheden ξ_i en η_i en de verschillen $(\xi_i - x_i)$ en $(\eta_i - y_i)$ zijn onderling onafhankelijk en normaal verdeeld met spreiding σ_1 resp. σ_2. Noch σ_1, noch σ_2, noch hun verhouding zijn bekend. Tussen de grootheden ξ_i, en η_i bestaat een lineaire betrekking $\eta_i = \alpha + \beta\xi_i$ met constante, doch niet bekende coëfficiënten α en β. Gevraagd wordt een schatting voor α, β, σ_1, en σ_2.

De waarschijnlijkheid een uitkomst (x_i,y_i) te vinden, als de ware waarden ξ_i en η_i zijn, is evenredig met:

2π

$$F_i = \frac{1}{\sigma_1\sigma_2\pi} \exp. - \tfrac{1}{2}\left\{\frac{(\xi_i - x_i)^2}{\sigma_1^2} + \frac{(\eta_i - y_i)^2}{\sigma_2^2}\right\} \qquad \mathrm{I}$$

Het is duidelijk, dat alle punten, die op de ellips E

$$\frac{(\xi_i - x)^2}{\sigma_1^2} + \frac{(\eta_i - y)^2}{\sigma_2^2} = k_i^2$$

liggen een gelijke waarschijnlijkheid hebben. Deze waarschijnlijkheid zal het grootst zijn indien k_i^2 zo klein mogelijk is.

De waarschijnlijkheid van het gelijktijdig optreden der waarden (x_1,y_1) - - - (x_n,y_n) is evenredig met:

$(2\pi)^n$

$$F = \frac{1}{\sigma_1^n\sigma_2^n\pi^n} \exp. - \tfrac{1}{2}\Sigma\left\{\frac{(\xi_i - x_i)^2}{\sigma_1^2} + \frac{(\eta_i - y_i)^2}{\sigma_2^2}\right\}, \qquad \mathrm{II}$$

waarbij tussen ξ_i en η_i de betrekking $\eta_i = \alpha + \beta\xi_i$ bestaat.

De moeilijkheid is, dat niet de "ware" waarden ξ_i en η_i bekend zijn, doch slechts de waargenomen waarden x_i en y_i. Om de ware waarden van ξ_i en η_i, behorende bij de waarneming x_i, y_i, te vinden kan nu evenwel worden gebruik gemaakt van een soort maximum-likelihood redenering; het punt (x_i,y_i) zal daarbij gerekend worden een waarneming te zijn van dát punt (ξ_i,η_i) van de rechte lijn $\eta = \alpha + \beta\xi$, dat de grootste waarde aan de waarschijnlijkheid F_i van dit gebeuren geeft. Dit zal plaatsvinden, indien k_i^2 zo klein mogelijk is, dus als het middelpunt (ξ_i,η_i) van de ellips E zodanig op de rechte gekozen wordt, dat de ellips raakt aan een lijn door (x_i,y_i) evenwijdig aan $\eta = \alpha + \beta\xi$. Hierbij moet in het oog gehouden worden, dat de assen der ellips een constante, zij het nog onbekende, verhouding $\frac{\sigma_1}{\sigma_2}$ hebben.

Het is duidelijk, dat tegen een dergelijke maximum-likelihood redenering bezwaren aangevoerd kunnen worden. Bij de normale vorm van het maximum-likelihood beginsel kiest men de onbekende parameters dusdanig, dat het samengestelde gebeuren bestaande uit de n waarnemingen, zo waarschijnlijk mogelijk is. Bij de onderhavige oplossing wordt evenwel de verdergaande eis gesteld, dat elke deelgebeurtenis zo waarschijnlijk mogelijk was, weliswaar gezien in combinatie met alle andere waargenomen deel-gebeurtenissen, zoals uit het vervolg zal blijken. In elk geval zou echter, indien de punten (x_i, y_i) niet de waarnemingen waren van de volgens bovenstaande redenering berekende punten (ξ_i, η_i), de plaatsgevonden samengestelde gebeurtenis een minder grotere waarschijnlijkheid gehad hebben dan bij deze aanname, zoals ook het geval zou zijn, indien de constanten α en β en σ_1 en σ_2 een andere waarde hebben dan uit de onderstaande berekening volgt. Dit rechtvaardigt, in de sfeer van de maximum-likelihood hypothese, de methode waarschijnlijk voldoende.

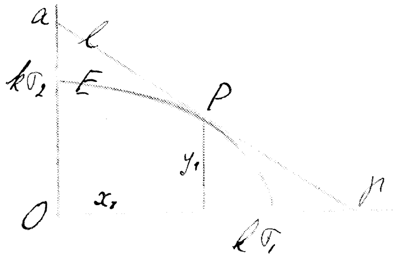
Voor de verdere ontwikkeling moet gebruik gemaakt worden van de volgende hulpstelling uit de analytische meetkunde.

<u>Hulpstelling:</u>

Indien de ellips $\frac{x^2}{k^2\sigma_1^2} + \frac{y^2}{k^2\sigma_2^2} = 1$ raakt aan de rechte lijn $y = a + bx$ dan is:

$$a^2 = k^2(\sigma_2^2 + b^2\sigma_1^2)$$

<u>Bewijs:</u>



Stel de ellips E met formule $\frac{x^2}{k^2\sigma_1^2} + \frac{y^2}{k^2\sigma_2^2} = 1$ raakt in het punt $P(x_1, y_1)$ aan de rechte lijn ℓ met formule $y = a + bx$.

Kan eenvoudiger
door gewoon a+bx voor y in te vullen
en de discriminant = 0 te stellen.

De tangens van de raaklijn in P is $-\frac{k^2\sigma_2^2x_1}{k^2\sigma_1^2y_1} = -\frac{\sigma_2^2x_1}{\sigma_1^2y_1} = b(=\mathrm{tg}\gamma)$.

Derhalve is $x_1 = \frac{-b\sigma_1^2y_1}{\sigma_2^2}$. Ingevuld in de vergelijking voor ℓ, waarop het punt $P(x_1, y_1)$ óók ligt, geeft dit:

$$y_1 = a + \left(b \cdot \frac{-b\sigma_1^2y_1}{\sigma_2^2}\right), \qquad y_1\left(1 + \frac{b^2\sigma_1^2}{\sigma_2^2}\right) = a$$

$$y_1 = \frac{a\sigma_2^2}{\sigma_2^2 + b^2\sigma_1^2}, \qquad x_1 = \frac{-a.b\,\sigma_1^2}{\sigma_2^2 + b^2\sigma_1^2}$$

Invullen van x_1 en y_1 in de vergelijking der ellips levert na vereenvoudiging:

$$\frac{a^2b^2\sigma_1^{\ 2} + a^2\sigma_2^{\ 2}}{k^2\,(\sigma_2^{\ 2} + b^2\sigma_1^{\ 2})^2} = 1 \text{ of } a^2 = k^2(\sigma_2^{\ 2} + b^2\sigma_1^{\ 2}).$$

Het betoog loopt verder als volgt:

Teneinde de maximum waarde van de likelihood-functie F (formule II) te vinden, moeten eerst met behulp van het boven uiteengezette principe de waarden van ξ_i en η_i gevonden worden bij bekend veronderstelde α, β, σ_1 en σ_2 en vervolgens die waarden van deze vier grootheden bepaald worden, die F tot maximum maken. Om de maximum waarde van een factor F_i te vinden, moet het rechterlid k^2 van de formule van ellips E gevonden worden, die door het punt (x_i,y_i) gaat en daar een lijn parallel aan de lijn ℓ met formule $\eta = \alpha + \beta\xi$ raakt, of, wat op hetzelfde neerkomt, het rechter lid van de ellips E' met middelpunt in (x_i,y_i), die de lijn $\eta = \alpha + \beta\xi$ raakt.

Hiertoe worde de oorsprong verplaatst naar (x_i,y_i). De transformatievergelijkingen luiden $\xi' = \xi - x_i$, $\eta' = \eta - y_i$. De vergelijking van ℓ wordt nu:

$$\eta' + y_i = \alpha + \beta\xi' + \beta x_i$$

of wel

$$\eta' = (-y_i + \alpha + \beta x_i) + \beta\xi',$$

die van de ellips E'

$$\frac{\xi'^2}{k^2\sigma_1^{\ 2}} + \frac{\eta'^2}{k^2\,\sigma_2^{\ 2}} = 1$$

Uit de bovenbewezen hulpstelling volgt, dat deze rechte aan de ellips zal raken indien:

$$(-y_i + \alpha + \beta x_i)^2 = k^2(\sigma_2^{\ 2} + \beta^2\,\sigma_1^{\ 2}).$$

Derhalve geldt:

ne. $$k^2 = \frac{(-y_i + \alpha + \beta x_i)^2}{\sigma_2^{\ 2} + \beta^2\,\sigma_1^{\ 2}}$$

De waarde van F wordt dus met behulp van deze maximum-likelihood redenering als volgt:

$$F = \frac{1}{\sigma_1^{\ n}\sigma_2^{\ n}\pi^n} \quad \text{exp.} - \tfrac{1}{2}\,\frac{\Sigma(-y_i + \alpha + \beta x_i)^2}{\sigma_2^{\ 2} + \beta^2\,\sigma_1^{\ 2}}$$

De normale maximum-likelihood methode zegt nu, dat de beste schattingen $\hat{a}$, $\hat{b}$, $\hat{s}_1$ en $\hat{s}_2$ voor α, β, σ_1, en σ_2 gevonden worden, door die waarden voor deze grootheden te kiezen, die F maximum en dus - log F minimum maken. Het minimum van

$$-\log F = \frac{1}{2}\,\frac{\Sigma(y_i - \alpha - \beta x_i)^2}{\sigma_2^2 + \beta^2\sigma_1^2} + n \log \sigma_1 + n \log \sigma_2 + n \log \pi$$

wordt gevonden door de partiële afgeleiden van - log F naar α, β, σ_1 en σ_2 gelijk aan 0 te stellen. Dit levert:

$$-\left[\frac{\partial \log F}{\partial \alpha}\right]_{\alpha = \hat{a}} = -\frac{\Sigma(y_i - \hat{a} - \hat{b}x_i)}{\hat{s}_2^2 + \hat{b}^2\hat{s}_1^2} = 0 \qquad \text{III}$$

$$-\left[\frac{\partial \log F}{\partial \beta}\right]_{\beta = \hat{b}} = -\frac{\Sigma x_i(y_i - \hat{a} - \hat{b}x_i)}{\hat{s}_2^2 + \hat{b}^2\hat{s}_1^2} - \frac{\hat{b}\hat{s}_1^2\Sigma(y_i - \hat{a} - \hat{b}x_i)^2}{(\hat{s}_2^2 + \hat{b}^2\hat{s}_1^2)^2} = 0 \qquad \text{IV}$$

$$-\left[\frac{\partial \log F}{\partial \sigma_1}\right]_{\sigma_1 = \hat{s}_1} = -\frac{\hat{b}^2\hat{s}_1\Sigma(y_i - \hat{a} - \hat{b}x_i)^2}{(\hat{s}_2^2 + \hat{b}^2\hat{s}_1^2)^2} + \frac{n}{\hat{s}_1} = 0 \qquad \text{V}$$

$$-\left[\frac{\partial \log F}{\partial \sigma_2}\right]_{\sigma_2 = \hat{s}_2} = -\frac{\hat{s}_2\Sigma(y_i - \hat{a} - \hat{b}x_i)^2}{(\hat{s}_2^2 + \hat{b}^2\hat{s}_1^2)^2} + \frac{n}{\hat{s}_2} = 0 \qquad \text{VI}$$

Door optelling van V en VI na vermenigvuldiging met $\hat{s}_1$ resp. $\hat{s}_2$ wordt gevonden:

$$\frac{-(\hat{s}_2^2 + \hat{b}^2\hat{s}_1^2)\Sigma(y_i - \hat{a} - \hat{b}x_i)^2}{(\hat{s}_2^2 + \hat{b}^2\hat{s}_1^2)^2} + 2n = \frac{-\Sigma(y_i - \hat{a} - \hat{b}x_i)^2}{(\hat{s}_2^2 + \hat{b}^2\hat{s}_1^2)} + 2n = 0 \qquad \text{VII}$$

of wel

$$\frac{\Sigma(y_i - \hat{a} - \hat{b}x_i)^2}{2n} = \hat{s}_2^2 + \hat{b}^2\hat{s}_1^2 \qquad \text{VIII}$$

Uitkomst VII toegepast op vergelijking V levert:

$$\frac{-\hat{b}^2\hat{s}_1}{\hat{s}_2^2 + \hat{b}^2\hat{s}_1^2} \cdot 2n + \frac{n}{\hat{s}_1} = 0$$

$$\frac{2\hat{b}^2\hat{s}_1^2}{\hat{s}_2^2 + \hat{b}^2\hat{s}_1^2} = 1 \qquad \text{IX}$$

$$\hat{s}_2^2 = \hat{b}^2\hat{s}_1^2 \qquad \text{X}$$

Uitkomst VII toegepast op vergelijking IV levert:

$$\frac{-\Sigma x_i\ (y_i - \hat{a} - \hat{b}x_i)}{\hat{s}_2^{\ 2} + \hat{b}^2\hat{s}_1^{\ 2}} - \frac{\hat{b}\hat{s}_1^{\ 2}}{\hat{s}_2^{\ 2} + \hat{b}^2\hat{s}_1^{\ 2}} \cdot 2\,n = 0 \qquad \text{XI}$$

In combinatie met IX volgt hieruit:

$$\frac{-\Sigma x_i\ (y_i - \hat{a} - \hat{b}x_i)}{\hat{s}_2^{\ 2} + \hat{b}^2\hat{s}_1^{\ 2}} - \frac{n}{\hat{b}} = 0 \qquad \text{XII}$$

$$-\Sigma x_i\ (y_i - \hat{a} - \hat{b}x_i) - \frac{n}{\hat{b}}\,(\hat{s}_2^{\ 2} + \hat{b}^2\hat{s}_1^{\ 2}) = 0$$

Substitutie van $(\hat{s}_2^{\ 2} + \hat{b}^2\hat{s}_1^{\ 2})$ uit VIII geeft:

deze volgt direct uit IV door X in te vullen.

$$-\Sigma x_i\ (y_i - \hat{a} - \hat{b}x_i) - \frac{\Sigma(y_i - \hat{a} - \hat{b}x_i)^2}{2\,\hat{b}} = 0$$

$$\Sigma\, 2\hat{b}x_i\ (y_i - \hat{a} - \hat{b}x_i) + \Sigma(y_i - \hat{a} - \hat{b}x_i)^2 = 0$$

$$\Sigma(y_i - \hat{a} - \hat{b}x_i)\ \left\{(y_i - \hat{a} - \hat{b}x_i) + 2\hat{b}x_i\right\} = 0$$

$$\Sigma(y_i - \hat{a})^2 - \Sigma\,\hat{b}^2x_i^{\ 2} = 0$$

$$\Sigma y_i^{\ 2} - 2\hat{a}\,\Sigma y_i + n\hat{a}^2 = \hat{b}^2\Sigma x_i^{\ 2} \qquad \text{XIII}$$

Uit III volgt:

$$\Sigma y_i - n\hat{a} = \hat{b}\,\Sigma x_i \qquad \text{XIV}$$

Het kwadraat van XIV, gedeeld door n, afgetrokken van XIII, geeft:

$$\Sigma y_i^{\ 2} - \frac{(\Sigma y_i)^2}{n} = \hat{b}^2\left\{\Sigma x_i^{\ 2} - \frac{(\Sigma x_i)^2}{n}\right\}$$

$$\Sigma(y_i - \bar{y})^2 = \hat{b}^2\,\Sigma\,(x_i - \bar{x})^2$$

$$\hat{b}^2 = \frac{\Sigma(y_i - \bar{y})^2}{\Sigma(x_i - \bar{x})^2} \qquad \text{XV}$$

Uit XIV volgt:

$$\hat{a} = \bar{y} - \hat{b}\bar{x} \qquad \text{XVI}$$

Tenslotte geeft XII in combinatie met X:

$$\frac{n}{\hat{b}}\left(\hat{s}_2^{\,2} + \hat{b}^2 s_1^{\,2}\right) = 2\, n\hat{b}\hat{s}_1^{\,2} = -\Sigma x_i(y_i - \hat{a} - \hat{b}x_i)$$

$$\hat{s}_1^{\,2} = -\frac{\Sigma x_i y_i - \hat{a}\Sigma x_i - \hat{b}\Sigma x_i^2}{2n\hat{b}} = \frac{\hat{b}\Sigma(x_i - \bar{x})^2 - \Sigma(x_i - \bar{x})\,(y_i - \bar{y})}{2n\hat{b}} \qquad \text{XVII.}$$

Het teken $\hat{b}$ moet dusdanig gekozen worden, dat $\hat{s}_1^{\,2}$ positief wordt. fout. Indien de covariance van x en y positief is, kan dit alleen bij positieve $\hat{b}$, indien de covariance negatief is alleen bij negatieve $\hat{b}$.

Tenslotte volgt uit X en XV

$$\hat{s}_2^{\,2} = \hat{b}^2\hat{s}_1^{\,2} = \frac{\hat{b}^2\Sigma(x_i - \bar{x})^2 - \hat{b}\Sigma(x_i - \bar{x})\,(y_i - \bar{y})}{2n}$$

$$= \frac{\Sigma(y_i - \bar{y})^2 - \hat{b}\Sigma(x_i - \bar{x})\,(y_i - \bar{y})}{2n} \qquad \text{XVIII.}$$

Formules XVI - XVIII geven de oplossing van het probleem.

Opgemerkt moge nog worden, dat WALD bij zijn methode van oplossen tot de betrekking (in onze notatie uitgedrukt)

$$\sqrt{\frac{\Sigma(y_i - \bar{y})^2}{n}} = \beta\sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n}}$$

komt (loc.cit. formule 9), die klaarblijkelijk samenhangt met onze formule XV.

Met bij voorbaat onze dank voor Uw hulp, verblijf ik

hoogachtend
De Afdeling Bewerking
Waarnemingsuitkomsten
het Hoofd

(Th.J.D.Erlee).

27 April 1949.

Aan de Heer Th.J.D. Erlee,
Koningskade 12,
D e n H a a g .

Zeer Geachte Heer Erlee,

De beantwoording van Uw brief van 30 (15) Maart j.l. heeft enige vertraging ondervonden, doordat er - zoals ik de Heer Drion enige weken geleden reeds kort mededeelde - enig bezwaar tegen te maken is, en het enigszins moeilijk bleek, duidelijk aan te geven, waarop dit gegrond is.

In het betoog wordt uitgegaan van een 2n-dimensionale wh-verdeling voor de monsters, voor welker dichtheid ik f_{2n} in plaats van F zal schrijven. Op grond van de onderstellingen die gemaakt worden[1]) is

$$\text{II} \qquad f_{2n} = \frac{1}{(2\pi\sigma_1\sigma_2)^n} \exp\left\{-\frac{1}{2}\,\frac{\sum(x_i-\xi_i)^2}{\sigma_1^2} - \frac{1}{2}\,\frac{\sum(y_i-\alpha-\beta\xi_i)^2}{\sigma_2^2}\right\}$$

Voor de ξ_i wordt nu een aannemelijkste schatting bepaald. De hiervoor gebruikte geometrische methode geeft hetzelfde resultaat als dat waartoe de methode van Fisher, toegepast op de ξ_i alleen, zou leiden, t.w.

$$\hat{\xi}_i = \frac{\sigma_2^2 x_i + \beta\sigma_1^2(y_i-\alpha)}{\sigma_2^2+\beta^2\sigma_1^2}$$

Daaruit volgt met $\hat{\eta}_i = \alpha+\beta\hat{\xi}_i$:

$$\frac{x_i-\hat{\xi}_i}{-\beta\sigma_1^2} = \frac{y_i-\hat{\eta}_i}{\sigma_2^2} = \frac{y_i-\alpha-\beta x_i}{\sigma_2^2+\beta^2\sigma_1^2}$$

Substitueert men de $\hat{\xi}_i$ voor ξ_i in II, dan ontstaat Uw formule op pag. 4 onderaan[1]), t,w.

$$\text{II}^* \qquad f_{2n} = \frac{1}{(2\pi\sigma_1\sigma_2)^n} \exp\left\{-\frac{1}{2}\,\frac{\sum(y_i-\alpha-\beta x_i)^2}{\sigma_2^2+\beta^2\sigma_1^2}\right\}$$

) De factor 2π was per abuis door π vervangen.

27 April 1949.

Deze grootheid echter is geen 2n-dimensionale verdelingsdichtheid meer. Integreert men nl. over $x_1, \cdots, x_n, y_1, \cdots, y_n$ dan wordt de integraal $+\infty$. De gevonden waarden voor $\hat{\xi}_i$ vormen dus niet - zoals het geval schijnt te zijn - de aannemelijkste schatting voor de ξ_i, maar in het geheel geen schatting op grond van de gemaakte onderstellingen.

Beschouwen we ter toelichting het eenvoudiger geval van een enkele te schatten parameter θ en zij $f(x/\theta)$ de verdelingsdichtheid van iedere waarneming. Vindt men door oplossing van de vgln van Fisher een waarde $\hat{\theta}$, dan zal men deze alleen dan als aannemelijkste schatting kunnen beschouwen, indien zij gelegen is binnen het waardegebied dat θ doorlopen kan om een verdeling van het onderstelde type te geven. Dit is echter niet steeds het geval. Beschouwen we ter illustratie het geval van 2 onbekende parameters μ en σ ener normale verdeling, en zij de opgave gesteld, deze uit één waarneming x te schatten. De vgln worden

$$\frac{x-\mu}{\sigma^2} = 0 \quad \text{en} \quad \frac{(x-\mu)^2}{\sigma^3} = \frac{1}{\sigma}$$

Men zou geneigd zijn, met σ^2 resp. σ^3 te vermenigvuldigen en te concluderen tot $\mu = x$, $\sigma = 0$. D.w.z. men zou concluderen, dat de gevonden waarde tevens de ware ware, en dat de spreiding nul ware. De conclusie echter is fout, omdat de verdeling voor $\sigma = 0$ niet meer normaal (maar éénwaardig) is, zodat men met de gemaakte onderstellingen in strijd komt. (Bovendien voldoen $\mu = x$, $\sigma = 0$ niet noodzakelijk aan de gevonden vgln, daar de linkerleden onbepaald worden. Bovendien maken deze waarden $f(x/\theta) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}$

27 April 1949.

neemt men eerst $\sigma=0$ en dan $\mu=x$, dan wordt $f=0$;

niet tot een maximum; neemt men eerst $\mu=x$ en dan $\sigma=0$, dan wordt $f=+\infty$; laat men μ en σ tegelijk tot hun grenswaarden naderen, dan hangt het af van de weg in het (μ, σ)-vlak waarlangs deze bereikt worden, of men een limiet voor f vindt, en, zo ja, welke). De conclusie behoorde dus te luiden: onder alle mogelijke schattingen (μ, σ) is geen de aannemelijkste.

In het door U aangegeven geval is de situatie analoog, zij het gecompliceerder. Ondersteld was, dat de 2 n waarnemingen onderling onafhankelijk waren. Men schat nu de parameters ξ_i en vindt een verdeling, waarbij slechts n waarnemingscombinaties onafhankelijk zijn, doordat er n (niet slchts stochastische maar zelfs exacte) afhankelijkheden t.w.

$$\frac{x_i-\hat{\xi}_i}{-\beta\sigma_1^2} = \frac{y_i-\hat{\eta}_i}{\sigma_2^2}$$

tussen bestaan. Geconcludeerd moet dan ook worden, dat er geen aannemelijkste schatting voor de ξ_i bestaat.

Inderdaad gaat, zoals de heer Hemelrijk opmerkte, de waarde II* voor $\beta=0$ en $\sigma_1\to 0$ met $\sigma_2\neq 0$, $\sigma_2\neq\infty$, $y_i\neq \alpha$ naar $+\infty$, zodat zij dan dus zeker geen maximum heeft.

Ter toelichting van deze wellicht ongewone situatie diene nog het volgende, waarbij we evengoed β als gegeven kunnen beschouwen, daar de moeilijkheid alleen in de schatting der ξ_i gelegen is.

Ondersteld is, dat bij iedere waarneming x en y stochastisch variabel zijn. Ditzelfde geldt dan ook voor de coördinaten x', y' t.o.v. een coördinatenstelsel, waarvan de x'-as langs de ijklijn valt en de y'-as en t.o.v. de ellipsen aan toegevoegd is (Of σ_2^2/σ_1^2 en daarmede de richting der y'-as bekend is, doet hier niet ter zake). Bovendien zijn x' en y' op grond van de ondersteldte normaliteit onderling onafhankelijk.

27 April 1949.

Bij de schatting van α treden nu geen bijzonderheden op. Anders echter is het bij de schatting der ξ_i'. Om dit in te zien, kunnen we de y_i' en hun verdeling geheel buiten beschouwing laten.

Men vraagt dan dus uit n onderling onafhankelijke normaal verdeelde waarnemingen x_i', alle met dezelfde spreiding σ' ($= \sqrt{\beta^2\sigma_1^2 + \sigma_2^2}$) en elk met een onbekend gemiddelde ξ_i', deze gemiddelden en spreiding te schatten. Dit is precies het geval, dat boven voor n = 1 als voorbeeld behandeld werd. Men vindt $\xi_i' = x_i'$ en $\underline{\underline{\sigma' = 0}}$, d.w.z. de ondersteIde normaliteit is als men deze parameterwaarden accepteert niet meer aanwezig.

Ook hier is dus geen aannemelijkste schatting aanwezig. Tracht men de moeilijkheid te ontgaan door de aanvankelijke onderstelling der normaliteit te laten vallen, dan vindt men als aannemelijkste schattingen eveneens $\xi_i' = x_i$, $\sigma' = 0$ d.w.z. als men voldoende vrijheid voor de keuze der verdelingen der toelaat, dan worden de aannemelijkste - op grond van telkens één waarneming - degene, waarbij telkens de waargenomen waarde de enig mogelijke is.

Wel kan men dan nog van de bepaling ener aannemelijke schatting afzien en zich met een gebied van voldoend aannemelijke schattingen, dus een betrouwbaarheidsgebied voor de ξ_i tevreden stellen, een wijze van behandelen, die ook overigens zekere voordelen biedt.

Overigens zal het U wellicht interesseren, dat inmiddeld de Heer J. Hemelrijk naar aaleiding van een ander, maar verwant probleem op zeer eenvoudige wijze een betrouwbaarheidsgebied voor de parameters van de ijklijn heeft aangegeven, met behulp ener methode die met die van de zgn. "order-statistics" verwant is. Zodra dit volledig is opgeschreven zullen we er U gaarne mede in kennis stellen.

In de hoop U hiermede de gewenste voorlichting gegeven te hebben, verblijf ik,

Hoogachtend,

w.g. D. van Dantzig.