Mathematics. — M. J. VAN UVEN: *Scheeve correlatie tusschen twee veranderlijken.* (Communicated by Prof. A. A. NIJLAND).

(Communicated at the meeting of April 27, 1929).

In het artikel „Over het bewerken van Scheeve Correlatie" (Versl. K. A. v. W. **34**, pp. 787 en 965; **35**, p. 129) hebben we getracht uit een willekeurige frequentieverdeeling van twee veranderlijken, x en x', een normale verdeeling te construeeren van twee (unimodulaire) veranderlijken t en t', die zoo eenvoudig mogelijk van x en x' afhangen. De aldaar ontwikkelde methode kan zeer beknopt worden samengevat als volgt:

De frequentieverdeeling is gegeven door:

Voor Y_{kl} exemplaren geldt:

$$x_{k-1} < x < x_k, \; x'_{l-1} < x' < x'_l \; (k=1, \dots n; l=1, \dots n'; \sum_{k=1}^{n} \sum_{l=1}^{n'} Y_{kl} = N)$$

Als hulpfuncties voeren we in:

$z(x)$, waarbij $z_k = z(x_k)$ bepaald wordt door

$$\Theta(z_k) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{z_k} e^{-v^2} dv = s_k = \frac{\sum_{i=1}^{k} \sum_{j=1}^{n'} Y_{ij}}{N},$$

$z'(x')$, waarbij $z'_l = z'(x'_l)$ bepaald wordt door

$$\Theta(z'_l) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{z'_l} e^{-v^2} dv = s'_l = \frac{\sum_{i=1}^{n} \sum_{j=1}^{l} Y_{ij}}{N},$$

$\zeta(x, x')$, waarbij $\zeta_{k-1/2, l} = \zeta(x_{k-1/2}, x'_l)$ bepaald wordt door

$$\Theta(\zeta_{k-1/2, l}) = \sigma_{k-1/2, l} = \frac{\sum_{j=1}^{l} Y_{kj}}{\sum_{j=1}^{n'} Y_{kj}},$$

$\zeta'(x, x')$, waarbij $\zeta'_{k, l-1/2} = \zeta'(x_k, x'_{l-1/2})$ bepaald wordt door

$$\Theta(\zeta'_{k, l-1/2}) = \sigma'_{k, l-1/2} = \frac{\sum_{i=1}^{k} Y_{il}}{\sum_{i=1}^{n} Y_{il}}.$$

De differentiaal-kansformule luidt nu:

$$dW = \frac{\sqrt{1-\gamma^2}}{\pi} e^{-(t^2-2\gamma t t'-t'^2)}\, dt\, dt' \rightleftarrows e^{-(z^2+\zeta^2)} \frac{\partial \zeta}{\partial z'}\, dz\, dz' = \frac{1}{\pi} e^{-(z'^2+\zeta'^2)} \frac{\partial \zeta'}{\partial z}\, dz\, dz'.$$

Door

$$\gamma = \cos\omega,\ z^2+\zeta^2 = r^2,\ z'^2+\zeta'^2 = r'^2,\ zz'-\zeta\zeta' = A,\ z\zeta'+z'\zeta = B$$

te stellen krijgen we

$$\gamma = \cos\omega = \frac{A}{rr'},\ \sqrt{1-\gamma^2} = \sin\omega = \frac{B}{rr'}.$$

Voeren we nu twee rechthoekige (tegengesteld georiënteerde) coördinaatstelsels (z, ζ), (z', ζ') in, zóó, dat $\angle (z, z') = \omega$, dan wordt het stel z, ζ afgebeeld door een punt P en het stel z', ζ' door een punt P', zóó, dat OP' langs OP valt (fig. 1).

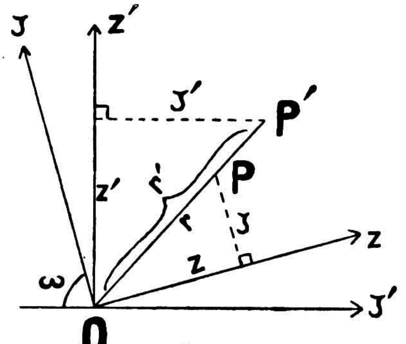

fig. 1

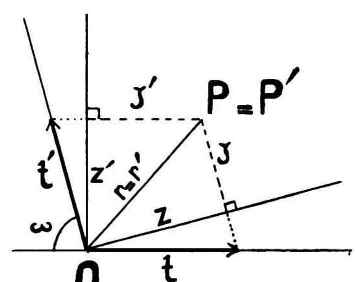

fig. 2

Wanneer nu in elk stel x_k, x'_l gevonden wordt:

I. $r' = r$, zoodat P' met P samenvalt (fig. 2),

II. $\operatorname{tg}\omega = \dfrac{B}{A} = \text{constant}$, zoodat de hoek ω bij elk stel x, x' even groot uitvalt, dan blijkt er lineaire correlatie te bestaan tusschen de (unimodulaire) functies

$$t = \frac{z\,(x)}{\sin\omega} \quad \text{en} \quad t' = \frac{z'\,(x')}{\sin\omega},$$

terwijl de correlatiecoëfficient is: $\gamma = \cos\omega$ (Versl. K. A. v. W. **34**, p. 978). Is er *wel* voldaan aan $r' = r$, terwijl daarentegen ω veranderlijk blijkt te zijn, dan kunnen we een constanten hoek $\bar{\omega}$ (bij voorkeur een gemiddelde waarde van de veranderlijke ω) invoeren en het stelsel (z', ζ') vervangen door een stelsel $(\bar{z}', \bar{\zeta}')$ zóódanig, dat $\angle (z, \bar{z}') = \bar{\omega}$ (fig. 3).

Daarmee hebben we dan lineaire correlatie tot stand gebracht tusschen de (unimodulaire) functies

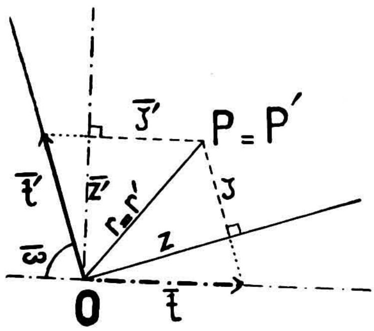

fig. 3

$$\bar{t} = \frac{z\,(x)}{\sin \bar{\omega}} \quad \text{en} \quad \bar{t}' = \bar{\zeta} + \cot \bar{\omega} \,.\, z,$$

met correlatiecoëfficient $\bar{\gamma} = \cos \bar{\omega}$.

Daar $\zeta = \frac{z' - \cos \omega \,.\, z}{\sin \omega}$, hebben we

$$\bar{t}' = (\cot \bar{\omega} - \cot \omega)\, z + \frac{z'}{\sin \omega} =$$

$$= \frac{\sin\,(\omega - \bar{\omega})}{\sin \omega \,.\, \sin \bar{\omega}} \,.\, z + \frac{z'}{\sin \omega},$$

conform met verg. (42) l.c. p. 981.

Terwijl $\bar{t}$ nu een zuivere functie is van x, is $\bar{t}'$ een gemengde functie van x en x' (zie ook de uitspraak op p. 982 l.c.).

Is evenmin voldaan aan $r' = r$, dan kunnen we P en P' tot één enkel punt P^* herleiden. Men past dus als het ware een vereffeningsmethode toe, die tusschen de twee verschillende punten een compromis sluit. In het aangehaalde artikel hebben we voor dit „vereffeningspunt" P^* gekozen: het op OPP' gelegen punt P^*, waarvoor $OP^* = R = \sqrt{rr'}$ (verg. Versl. K. A. v. W. **35**, p. 129). Daardoor werden zoowel z als z' aangetast; en het resultaat was, dat de unimodulaire veranderlijken t en t' beide gemengde functies werden van x en x'. Vandaar de uitspraak op p. 130 l.c. Wel is waar werd (zie de noot op p. 130 l.c.) de theoretische mogelijkheid erkend één der veranderlijken t en t' tot een zuivere functie van x (of x') te maken, maar het werd betwijfeld, of deze theoretische oplossing practische beteekenis zou hebben.

Het is ons intusschen gebleken, dat men ook in het algemeenste geval $(r' \neq r)$ zeer eenvoudig een zoodanige oplossing (bijv.: $t\,(x)$, $t'\,(x, x')$) kan construeeren.

Opdat t een zuivere functie zij van x, hebben we slechts:

1^0. een constanten hoek $\bar{\omega}$ in te voeren,

2^0. de functie $z\,(x)$ ongewijzigd te laten.

Van de ontelbare manieren, waarop we een dergelijke oplossing kunnen verkrijgen, noemen we er twee, die o. i. bijzonder de aandacht verdienen.

a. We behouden z en z'; de punten P en P' worden dan verplaatst naar een punt P^*, waarvoor $z^* = z$, $z'^* = z'$ (fig. 4). Daarbij veranderen natuurlijk ζ en ζ'. ζ gaat over in

$$\zeta^* = \frac{z' - \cos \omega \,.\, z}{\sin \omega} = - \cot \omega \,.\, z + \frac{z'}{\sin \omega}.$$

De hoek ω wordt ook hier ontleend aan

$$\operatorname{tg}\omega = \frac{B}{A} = \frac{z\zeta' + z'\zeta}{zz' - \zeta\zeta'};$$

ζ en ζ' werken dus nog mede tot het bepalen van ω.

Voeren we nu nog den constanten hoek $\bar{\omega}$ in (zie boven), dan komen

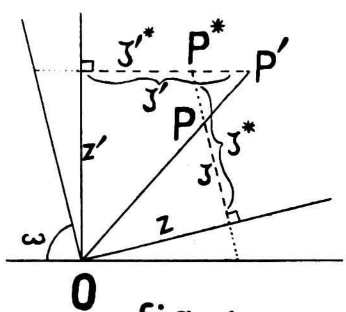

fig. 4

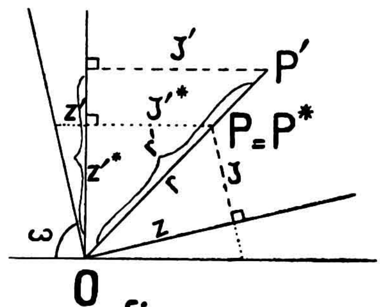

fig. 5

we tot de unimodulaire functies:

$$\bar{t} = \frac{z}{\sin\bar{\omega}},\ \bar{t}'^{\star} = \zeta^{\star} + \cot\bar{\omega}\,.\,z = (\cot\bar{\omega} - \cot\omega)\,.\,z + \frac{z'}{\sin\omega} =$$

$$= \frac{\sin(\omega - \bar{\omega})}{\sin\omega\,.\,\sin\bar{\omega}}\,.\,z + \frac{z'}{\sin\omega},$$

dus tot dezelfde functies $(\bar{t}, \bar{t}')$ als in het geval $r' = r$.

b. We behouden z en ζ, d. w. z. we geven P' geheel prijs (of, anders gezegd: P' wordt vereffend tot P) (fig. 5). De coördinaten z' en ζ' van P' dienen nu alleen ter bepaling van den hoek ω. De functies z' en ζ' gaan nu resp. over in

$$z'^{\star} = \frac{r}{r'}\,.\,z', \quad \zeta'^{\star} = \frac{r}{r'}\,.\,\zeta'.$$

$z'^{\star}$ wordt zoodoende een gemengde functie van x en x'.

Invoering van den constanten hoek $\bar{\omega}$ brengt ons dan tot de unimodulaire functies

$$\bar{t} = \frac{z}{\sin\bar{\omega}}, \quad \bar{t}'^{\star} = \zeta + \cot\bar{\omega}\,.\,z.$$

Hier geldt

$$\zeta = \frac{z'^{\star} - \cos\omega\,.\,z}{\sin\omega} = -\cot\omega\,.\,z + \frac{r}{r'}\frac{z'}{\sin\omega},$$

dus

$$\bar{t}'^{\star} = (\cot\bar{\omega} - \cot\omega)\,.\,z + \frac{r}{r'}\,.\,\frac{z'}{\sin\omega} = \frac{\sin(\omega - \bar{\omega})}{\sin\omega\,.\,\sin\bar{\omega}}\,.\,z + \frac{r}{r'}\,.\,\frac{z'}{\sin\omega}.$$

De eerste methode (*a*) heeft het voordeel, dat men werkt met de veranderlijken z en z', die veel zuiverder bepaald zijn dan de grootheden ζ en ζ'. Immers, om $\zeta_{k,l}$ en $\zeta'_{k,l}$ te bepalen, heeft men moeten interpoleeren tusschen de waarden $\zeta_{k-{}^1\!/_2,l}$, resp. $\zeta'_{k,l-{}^1\!/_2}$. Bovendien zijn deze laatste op zichzelf onzekerder dan de veranderlijken z en z', omdat ze uit kleinere frequentiesommen zijn berekend.

Het, overigens niet zeer belangrijke, nadeel van deze methode is, dat men bij het terugrekenen van de frequenties, hetzij uit (z, ζ^*), hetzij uit (z', ζ'^*), in geen van beide gevallen de gegeven frequenties precies terugkrijgt. Men kan echter de aldus berekende frequenties opvatten als vereffende empirische frequenties.

De tweede methode (*b*) heeft het voordeel, dat men bij het terugrekenen van de frequenties uit (z, ζ) de gegeven frequenties precies terugkrijgt. Daartegenover staat het nadeel, dat men, door zich uitsluitend naar het ééne verdeelingsprincipe (z, ζ) te richten, de frequentieverdeeling hoogst eenzijdig behandelt, en dat men met name de betrekkelijk nauwkeurige grootheden z' bijna geheel veronachtzaamt.

Intusschen moet de praktijk in elk geval uitmaken, welke der beide vereffeningsmethoden de voorkeur verdient.

SKEW CORRELATION BETWEEN TWO VARIABLES.

Summary.

The aim of this paper is: to complete the results obtained in our former paper "On Treating Skew Correlation", [Proceed. of the K. Ak. v. Wet. Amsterdam, Vol. **38**, p. 797 and p. 919, Vol. **39**, p. 580 (referred to by S. C. *a, b, c*)], and to amend the conclusion enunciated S. C. *c*, p. 581, also making the footnote (p. 581) superfluous.

Instead of adjusting the points $P(z, \zeta)$ and $P'(z', \zeta')$ to the point P^*, situated on OPP', so that $OP^{*2} = OP \times OP'$ ($R^2 = r\,r'$, cf. (47), S. C. *c*, p. 581), we can better adjust P and P' in such a way *that one of the two unimodular variables t and t' becomes a function of only one of the primary variables x and x'.*

In order to construct $t(x)$, $t'(x, x')$, it is necessary:

1°. to introduce a constant angle $\overline{\omega}$,

2°. to leave the function $z(x)$ unaltered.

Among the innumerable methods, by which we can construct $t(x)$, $t'(x, x')$, we recommend especially the following two:

a. We keep z and z'. Then the points P and P' are shifted to a point P^* for which $z^* = z$, $z'^* = z'$ (fig. 4); thus ζ is changed into

$$\zeta^* = \frac{z' - \cos\omega \,.\, z}{\sin\omega} = -\cot\omega \,.\, z + \frac{z'}{\sin\omega}.$$

The angle ω is still taken from $\operatorname{tg}\omega = \dfrac{B}{A} = \dfrac{z\zeta' + z'\zeta}{zz' - \zeta\zeta'}$ (cf. (32), S.C. *b*, p. 928). Introducing the constant angle $\overline{\omega}$ (mean of the variable ω), *we*

obtain the same unimodular functions that we arrived at in the case $P' = P$ $(r' = r)$ (cf. (42), S C. *b*, p. 935).

b. We keep z and ζ, or: we adjust P' to P (fig. 5). Here the unadjusted values z', ζ' serve only to determine ω (in cooperation with z, ζ). z' is afterwards adjusted to $z'^* = \frac{r}{r'} \cdot z'$, becoming in this way a mixed function of x and x'. Introducing the constant $\bar{\omega}$, we obtain at last:

$$\bar{t} = \frac{z}{\sin \bar{\omega}} \quad , \quad \bar{t}'^* = \frac{\sin (\omega - \bar{\omega})}{\sin \omega \, . \, \sin \bar{\omega}} \, . \, z + \frac{r}{r'} \, . \, \frac{z'}{\sin \omega} \, .$$

As to (*a*): Advantage: we operate with z and z', which are determined more accurately than ζ and ζ' (these latter being computed from small frequencies, and affected by the incertainties of interpolation).

Drawback: By computung back the frequencies from (z, ζ^*) [or from (z', ζ'^*)] we do not exactly get back the given frequencies.

As to (*b*): Advantage: By computing back the frequencies from (z, ζ) we obtain precisely the given frequencies.

Drawback: The frequency distribution is treated very partially, the comparatively accurate quantities z' being almost wholly neglected.