

STICHTING
MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM

S 47 (M11)

Hoek-toets voor onafhankelijkheid

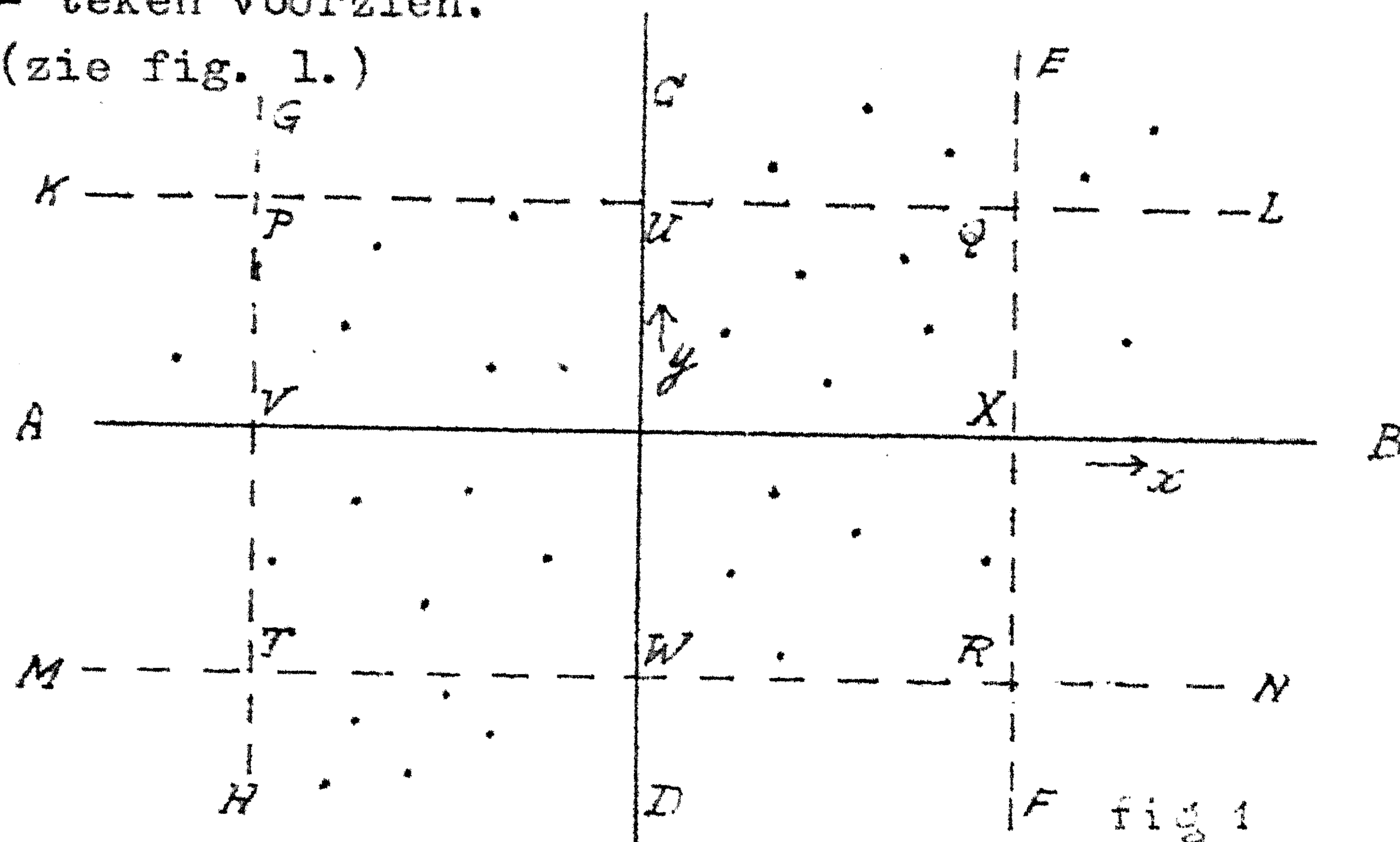


Hoek - toets voor onafhankelijkheid ¹⁾

Deze methode dient om de onafhankelijkheid te toetsen van 2 continuë stochastische grootheden. De toets is geldig zonder verdere onderstellingen over de vorm van de waarschijnlijkheidsverdelingen. Zij berust op n paren waarnemingen $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$. Deze getallenparen worden als punten in een vlak getekend.

We trekken nu 2 lijnen AB (horizontaal) en CD (verticaal), zo, dat onder en boven AB evenveel punten liggen, en evenzo evenveel punten links als rechts van CD . Van de vier kwadranten, die ze ontstaan, worden de kwadranten rechts boven en links onder van een + teken en de beide andere van een - teken voorzien.

(zie fig. 1.)



We tellen nu de punten af, beginnende aan de rechterkant van het diagram en naar links gaande, volgens afnemende x , tot we, om een volgend punt te vinden, de horizontale lijn AB moeten passeren. Tussen dit volgende punt en het laatst getelde punt trekken we de verticale lijn EF . Van links af tellend, volgens opklimmende x vinden we op analoge wijze de lijn GH .

Van bovenaf beginnend, tellen we de punten volgens afnemende y tot we de lijn CD moeten passeren. Daar trekken we de rechte KL . Aan de onderzijde bepalen we op analoge wijze de rechte MN .

¹⁾ Dit memorandum is slechts bedoeld ter oriëntatie en streeft niet naar volledigheid of volledige exactheid.

Er zijn nu een aantal punten geïsoleerd buiten de op deze wijze verkregen vierhoek $PQRT$ (zie fig. 1)
We bepalen nu de zgn. kwadranten som \underline{S} van deze aantallen geïsoleerde punten. Onder deze kwadrantensom verstaan we de som van het aantal punten boven KL , onder MM , links van GH en rechts van EF in het eerste en derde kwadrant, verminderd met het aantal punten buiten deze lijnen in het tweede en vierde kwadrant. De punten in de hoeken EQL , MTH , FRN en GPN worden dus dubbel geteld.

Vb. in fig. 1. wordt :

$$\begin{array}{r} \text{Hoek } CUL \quad + 5 \\ \text{,, } EXB \quad + 3 \\ \text{,, } GVA \quad - 1 \\ \text{,, } MWD \quad + 5 \\ \hline \underline{S} = + 12 \end{array}$$

De verdelingsfunctie van \underline{S} is bekend en vrijwel onafhankelijk van het aantal waarnemingen. Is aan de onafhankelijkheid voldaan, dan kan men betrekkelijk kleine waarden van $|\underline{S}|$ verwachten. Is er afhankelijkheid, dan is er meer kans op grote waarden van $|\underline{S}|$. Daarom neemt men kritieke zônes van de vorm $|\underline{S}| \geq |\underline{S}_0|$. Voor een onbetrouwbaarheidsdrempel $\alpha = 0,05$ is $\underline{S}'_0 = 11$, voor $\alpha = 0,01$ is $\underline{S}'_0 = 19$.
Voor het geval dat één of meer lijnen door één of meer waargenomen punten gaan wordt een kleine verandering in de werkwijze aangebracht. Men zie hiervoor de litteratuur.

Litteratuur

P.S. Olmstead, John W. Tukey

A Corner Test for Association
Ann. of Math. Stat. 1947,
Vol. 18, p. 495-513

A.M. Mood,

Introduction to the Theory of Statistics, 1950
p. 410.