STICHTING
MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM

STATISTISCHE AFDELING

Rapport S 260

Een methode van accountantscontrôle met behulp van steekproeven

door

J. Fabius

Oktober 1959

1. Inleiding.

Indien men met zekerheid wenst te weten of in een boekhouding fraude is gepleegd, zal men alle voorkomende posten moeten controleren. Zodra men immers volstaat met het controleren van een steekproef uit alle posten, loopt men het risico dat een eventuele fraude niet ontdekt wordt. Bij boekhoudingen, die op zodanige wijze zijn ingericht, dat fraude alleen gepleegd kan worden doordat te hoge bedragen geboekt worden, kan men echter steekproefsystemen ontwerpen waarbij het risico, dat een eventuele fraude van meer dan een van te voren vastgesteld bedrag f niet ontdekt wordt, hoogstens een van te voren gekozen waarde ε, bijv. 1% heeft.

In par. 2 wordt een door P. DE WOLFF (1959) ontworpen steekproefsysteem voor dit geval gegeven, waarna in par. 3 voor enkele concrete gevallen wordt nagegaan hoeveel posten men bij toepassing van dit systeem dient te controleren. De laatste paragraaf bevat de afleidingen.

2. Steekproefvoorschrift.

Zoals in par. 1 reeds werd uiteengezet, dient men het bedrag f en de kans ε van te voren te kiezen. Deze keuze kan men zien als een beleidskwestie. Men zou f de hoogste toelaatbare fraude kunnen noemen.

De procedure is nu als volgt:

Men kiest een geheel getal ν en vervolgens ν bedragen $x_1, \ldots, x_\nu$, zodanig dat $0 < x_1 < x_2 < \ldots < x_\nu$.

Het totale aantal posten zij nu N, en het aantal posten dat tussen x_{i-1} en x_i ligt zij N_i $(i = 1,2,\ldots,\nu)$. $(x_0=0)$. Het aantal posten groter dan x_ν noemen we $N_{\nu+1}$.

De posten zijn door middel van de getallen $x_1, \ldots, x_\nu$ dus naar opklimmende grootte in $(\nu+1)$ groepen ingedeeld. De $N_{\nu+1}$ posten die groter zijn dan x_ν dient men stuk voor stuk te controleren. Uit elk van de resterende groepen controleert men echter slechts een steekproef, men dien verstande, dat men uit de groep van posten die tussen x_{i-1} en x_i liggen een steekproef neemt van $\alpha_i N_i$ posten. De fractie α_i wordt hierbij gegeven door de formule

$$(2.1) \qquad \alpha_i = 1 - \varepsilon^{\frac{1}{\lambda_i}}, \text{ waarin } \lambda_i = \frac{f}{x_i} \quad .$$

Voor $\varepsilon = 0{,}01$ is α_i in figuur 1 grafisch uitgezet tegen λ_i, zodat men uit deze figuur voor $\lambda_i \leqslant 40$ de bijbehorende waarde van α_i kan aflezen[1]. Voor $\lambda_i > 40$ kan men gebruik maken van een benaderingsformule:

$$(2.2) \qquad \alpha_i = \frac{4{,}605}{\lambda_i} \quad \text{voor} \quad \varepsilon = 0{,}01 \text{ en } \lambda > 40.$$

Op deze wijze controleert men in totaal dus

$$\alpha_1 N_1 + \alpha_2 N_2 + \ldots + \alpha_k N_k + N_{k+1} \text{ posten.}$$

Vindt men hierbij een gefraudeerde post, dan dienen alle N posten gecontroleerd te worden. Vindt men echter geen enkele gefraudeerde post, dan mag men - behoudens een kans ε - concluderen, dat een eventuele niet-ontdekte fraude hoogstens f gulden bedraagt.

Uit het voorgaande blijkt, dat men - ook na de keuze van f en ε - nog een zekere mate van vrijheid heeft bij de constructie van het steekproefvoorschrift. Men mag immers de getallen $\nu, x_1, \ldots, x_\nu$ nog kiezen, waarna men de benodigde steekproefgrootten kan berekenen. Aangezien het aantal te controleren posten van deze keuze afhangt, zal er sprake zijn van een optimale keuze der getallen $\nu, x_1, \ldots, x_\nu$, optimaal in die zin, dat voor die keuze het aantal te controleren posten minimaal is.

Helaas is het niet mogelijk algemene regels te geven voor deze optimale keuze, zodat men voor elk geval apart door proberen er naar moet zoeken. Daar dit een zeer tijdrovend werk is, kan men echter in de meeste gevallen beter genoegen nemen met een min of meer arbitraire keuze, ook al zal deze uiteraard niet de optimale zijn.

Hier komt nog bij, dat de practische uitvoering van het nemen van de steekproeven veel ingewikkelder wordt naarmate ν

---

1) α_i kan uit de figuur in niet meer dan twee decimalen nauwkeurig worden afgelezen, hetgeen in de waarden $\alpha_i N_i$ tot fouten van enige procenten kan leiden. Indien een grotere nauwkeurigheid vereist is, moet derhalve formule (2.1) gebruikt worden.

groter gekozen wordt. Het is daarom aan te bevelen ν klein te kiezen, bijv. 1 of 2, ook al kan men door ν groter te nemen het aantal te controleren posten nog verder reduceren.

3. Enige voorbeelden.

a. Een debiteurensaldolijst bestaat uit 1662 posten, waarvan er 1052 kleiner zijn dan f 150.--, 204 tussen f 150.-- en f 250.--, en 172 tussen f 250.-- en f 500.-- liggen, terwijl de resterende 234 posten groter dan f 500.-- zijn. Het totaal is f 478.594,51. Men kiest f = 5000 en $\varepsilon = 0,01$. We kunnen nu voor verschillende keuzen van ν en $x_1, \ldots, x_\nu$ het aantal te controleren posten berekenen:

$\nu=1,\ x_1=250$

$N_1=1256 \qquad \lambda_1 = \frac{5000}{250} = 20 \qquad \alpha_1 = 0,206$

$N_2 = 406.$

Te controleren: 406 + 0,206.1256 = 665 posten.

$\nu=2,\ x_1=150,\ x_2=500$

$N_1=1052 \qquad \lambda_1 = \frac{5000}{150} = 33,33 \qquad \alpha_1 = 0,129$

$N_2 = 376 \qquad \lambda_2 = \frac{5000}{500} = 10 \qquad \alpha_2 = 0,369$

$N_3 = 234.$

Te controleren: 234 + 0,369.376 + 0,129.1052 = 508 posten.

$\nu=3,\ x_1=150,\ x_2=250,\ x_3=500$

$N_1=1052 \qquad \lambda_1=33,33 \qquad \alpha_1=0,129$

$N_2= 204 \qquad \lambda_2=20 \qquad \alpha_2=0,206$

$N_3= 172 \qquad \lambda_3=10 \qquad \alpha_3=0,369$

$N_4= 234.$

Te controleren:

234 + 0,369.172 + 0,206.204 + 0,129.1052 = 475 posten.

We zien dat de overgang van $\nu=1$ op $\nu=2$ nog een vrij aanzienlijke reductie van het aantal te controleren posten geeft, zodat het mogelijk is, dat $\nu=2$ te verkiezen is boven $\nu=1$, hoewel het nemen van de steekproeven voor $\nu=2$ bewerkelijker is dan voor $\nu=1$. De overgang op $\nu=3$ geeft echter geen noemenswaardige besparing meer.

b. Een lijst van inkoopfacturen, bestaande uit 3641 posten, bevat 2762 posten kleiner dan f 1.000.--, 333 posten tussen f 1.000.-- en f 1.500.--, 139 posten tussen f 1.500.-- en f 2.000.-- en 407 posten groter dan f 2.000.--.
Het totaal bedraagt ruim 3 miljoen gulden.
Men kiest f = 30.000 en $\varepsilon = 0{,}01$.

<u>$\nu=1$, $x_1=1500$</u>

$N_1=3095$ $\qquad \lambda_1= \frac{30.000}{1500} =20 \qquad \alpha_1=0{,}206$

$N_2= 546$.

Te controleren: $546 + 0{,}206.3095 = 1184$ posten.

<u>$\nu=2$, $x_1=1000$, $x_2=2000$</u>

$N_1=2762$ $\qquad \lambda_1= \frac{30.000}{1000} =30 \qquad \alpha_1=0{,}142$

$N_2= 472$ $\qquad \lambda_2= \frac{30.000}{2000} =15 \qquad \alpha_2=0{,}264$

$N_3= 407$.

Te controleren: $407 + 0{,}264.472 + 0{,}142.2762 = 924$ posten.

Ook hier geeft de overgang van $\nu=1$ op $\nu=2$ nog een kleine besparing. Het is echter zeer wel mogelijk dat men toch beter $\nu=1$ kan nemen vanwege het mindere werk.

c. Een lijst van 600 schadegevallen bestaat uit 454 posten kleiner dan f 50.--, 47 posten tussen f 50.-- en f 70.--, 26 posten tussen f 70.-- en f 100.-- en 73 posten groter dan f 100.--. Het totaal bedraagt ongeveer 5 ton. Men kiest f = 5000 en $\varepsilon = 0{,}01$.

$\nu=1,\ x_1=70$

$N_1=501 \qquad \lambda_1=\frac{5000}{70}=71{,}4 \qquad \alpha_1=\frac{4{,}605}{71{,}4}=0{,}064$

$N_2=99.$

Te controleren: $99 + 0{,}064.501 = 131$ posten.

$\nu=2,\ x_1=50,\ x_2=100$

$N_1=454 \qquad \lambda_1=\frac{5000}{50}=100 \qquad \alpha_1=\frac{4{,}605}{100}=0{,}046$

$N_2=73 \qquad \lambda_2=\frac{5000}{100}=50 \qquad \alpha_2=\frac{4{,}605}{50}=0{,}092$

$N_3=73.$

Te controleren: $73 + 0{,}092.73 + 0{,}046.454 = 101$ posten.

Het verschil tussen de aantallen te controleren posten bij $\nu=1$ en $\nu=2$ is ook hier niet bijzonder groot, zodat vermoedelijk $\nu=1$ de beste keuze is.

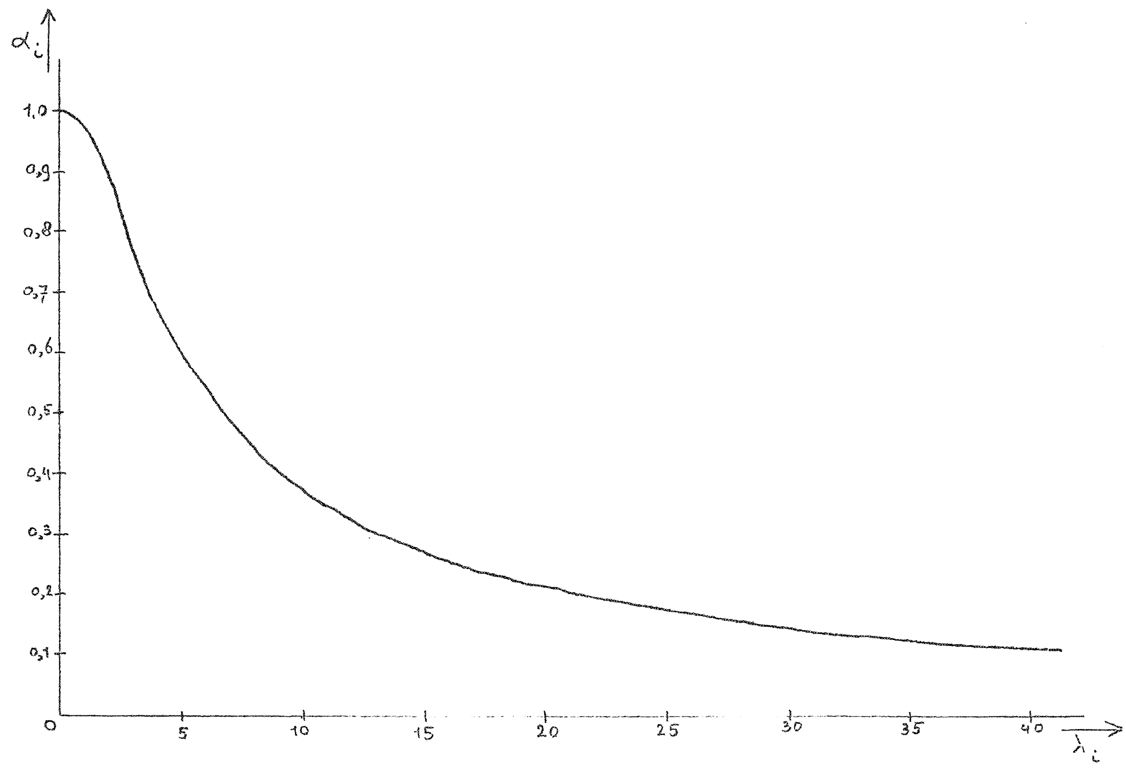

fig. 1: α_i als functie van λ_i voor $\varepsilon=0{,}01$

4. Afleiding van een steekproefsysteem bij accountantscontrole.

Gegeven zijn N posten, ingedeeld in $\nu+1$ klassen, zodanig, dat klasse i alle posten bevat die tussen x_{i-1} en x_i liggen, waarbij $0=x_0 < x_1 < x_2 \ldots < x_\nu < x_{\nu+1} = \infty$.

Aangenomen wordt, dat fraude alleen gepleegd kan zijn doordat één of meer te hoge bedragen zijn geboekt.

Men controleert de klasse $\nu+1$ volledig en uit elk der overige klassen een steekproef. Vindt men hierbij fraude, dan zal men alle posten controleren. Gevraagd wordt naar een betrouwbaarheidsgrens voor het gefraudeerde bedrag, indien men geen fraude vindt. Hierbij kunnen we ons beperken tot de klassen 1 t/m ν, daar de klasse $\nu+1$ geheel gecontroleerd wordt.

Zij de omvang van de steekproef die uit klasse i genomen wordt $n_i=\alpha_i N_i$ en laat de fractie gefraudeerde posten in diezelfde klasse p_i zijn. In dat geval is de kans dat geen fraude gevonden wordt:

$$P(p_1,\ldots,p_\nu) = \prod_{i=1}^{\nu} \frac{\binom{N_i-N_ip_i}{n_i}}{\binom{N_i}{n_i}} .$$

Aangezien steeds N_ip_i zeer klein zal zijn ten opzichte van N_i, kunnen we deze grootheid vervangen door

$$P(p_i,\ldots,p_\nu) = \prod_{i=1}^{\nu} (1-\alpha_i)^{N_ip_i} .$$

Als men geen fraude vindt, dan kan men met onbetrouwbaarheidsdrempel ε een betrouwbaarheidsgebied G opstellen voor het punt $(p_1,\ldots,p_\nu)$ in een ν-dimensionale ruimte:

$$G = \left\{ (p_1,\ldots,p_\nu) \,\middle|\, P(p_1,\ldots,p_\nu) \geqslant \varepsilon \right\} .$$

Voor alle punten van G geldt:

$$\sum_{1}^{\nu} \log\,(1-\alpha_i)^{N_ip_i} \geqslant \log\,\varepsilon$$

$$\therefore N_\nu p_\nu \leq \frac{1}{\log(1-\alpha_\nu)} \cdot \left\{\log \varepsilon - \sum_1^{\nu-1} N_i p_i \log(1-\alpha_i)\right\}.$$

Voorts is bij gegeven $p_1, p_2, \ldots p_\nu$ het gefraudeerde bedrag F hoogstens gelijk aan $\sum_{i=1}^{\nu} N_i p_i x_i$.

In alle punten van G is derhalve:

$$F \leq \sum_1^\nu N_i p_i x_i \leq \sum_1^{\nu-1} N_i p_i \left\{x_i - \frac{x_\nu}{\log(1-\alpha_\nu)} \log(1-\alpha_i)\right\} + \frac{x_\nu \log \varepsilon}{\log(1-\alpha_\nu)}.$$

Men kan het rechterlid van deze uitdrukking dus als betrouwbaarheidsgrens f voor F opgeven als men geen fraude heeft gevonden. Door de α_i zo te kiezen, dat

$$\frac{x_i}{\log(1-\alpha_i)} = \frac{x_\nu}{\log(1-\alpha_\nu)} \quad (i=1,2,\ldots,\nu),$$

krijgt men het gunstigste resultaat, d.w.z. een zodanig gebied G, dat voor elk punt buiten G ook $\sum_i^\nu N_i p_i x_i > f$ is, hetgeen meetkundig gemakkelijk is in te zien. Men heeft dan:

$$f = \frac{x_i}{\log(1-\alpha_i)} \cdot \log \varepsilon .$$

Omgekeerd kan men, als ν en $x_1, \ldots, x_\nu$ gegeven zijn bij f en de getallen α_i berekenen die hieraan voldoen:

$$\alpha_i = 1 - \varepsilon^{\frac{1}{\lambda_i}}, \text{ met } \lambda_i = \frac{f}{x_i}.$$

De zo verkregen getallen α_i stellen ons in staat de minimale steekproefgrootten te berekenen welke, indien geen fraude gevonden wordt, tot betrouwbaarheidsgrens f bij onbetrouwbaarheidsdrempel ε leiden.

Literatuur:

P. DE WOLFF (1959), Produktiviteitsverhoging bij accountantscontrole door toepassing van gelaagde steekproeven, Statistica Neerlandica 13 (1959) 215-232.