DE STEEKPROEFMETHODE VAN A. VAN HEERDEN

door H. M. Wopkes

1. „Deze methode toch betekent een wezenlijke stap vooruit in de techniek van partiële waarneming in het kader van de accountantscontrole . . . Het ware te wensen dat collega Van Heerden door de besproken publicatie geprikkeld zou worden als vakman zijn methode in dit tijdschrift te bespreken.”

Aldus A. B. Frielink in een boekbespreking in het maandblad voor accountancy en bedrijfshuishoudkunde van december 1961. Daarbij ging het om het rapport S 274 van de Stichting Mathematisch Centrum Amsterdam over toepassingen van a-selecte steekproeven bij accountantscontroles, geschreven door J. Kriens en gepubliceerd in oktober 1960.

De Heer Frielink veronderstelt, dat dit geschrift „vanuit wiskundig standpunt bezien duidelijk en zonder feilen is”.

Overigens is de recensent van mening, dat „de onderhavige publikatie . . . in ernstige mate tekort” schiet „in aansluiting aan de problematiek van de accountantscontrole”.

De wens van de lezer moge ook zo snel vervuld worden ten aanzien van dingen, waarheen zijn begeerte uitgaat: in hetzelfde nummer van genoemd maandblad verscheen een studie van de heer A. van Heerden onder de titel „Steekproeven als middel van accountantscontrole”.

2. Na lezing van de hiervoor vermelde citaten zal men begrijpen, dat ik enigszins huiverig ben om de conclusies openbaar te maken, waartoe ik na grondige bestudering van de steekproefmethode van A. van Heerden ben gekomen.

De „guldenrangnummERMETHODE” leek mij aanvankelijk een prestatie van grote betekenis. Toch ben ik thans van opinie, dat deze werkwijze ons weinig verder brengt naar de oplossing van het probleem, waarvoor wij ons gesteld zien. Bovendien zie ik het gevaar, dat de manier, waarop de idee wordt gelanceerd, omringd met „frivolités mathématiques” ¹⁾ en statistisch-theoretische uiteenzettingen, velen zal doen terugschrikken voor een intensieve „investigation”. Echter: „het grootste kwaad is angst.” ²⁾ Daarom zou een eenvoudige handleiding, waarin de beginselen der differentiaal- en integraalrekening en van enige andere onderwerpen, verband houdend met ons vak, worden besproken, zeer welkom zijn. De draad van het artikel van A. van Heerden is echter zonder veel moeite te volgen. Daarbij kan men de „frivolités” terzijde laten.

Wel bedenke men, dat een toneelstuk eerst goed begrepen kan worden, wanneer de „terzijdes” worden verstaan.

3. De „guldenrangnummERMETHODE” komt hier op neer: „De wijze waarop een te controleren postenreeks zich in feite manifesteert”, wordt „geheel losgelaten en (in gedachten) getransformeerd in een reeks guldens, die (slechts toevallig) in posten van verschillende grootte zijn samengebracht. Aan elke gulden in deze reeks wordt nu een eigen rangnummer gegeven” (m.a.b. december 1961 blz. 459). En verder: „Statistisch is het probleem nu eenvoudig en teruggebracht tot de bekende

¹⁾ Dr. S. Kleerekoper, *Vergelijkend leerboek der bedrijfseconomie I* 1956 blz. 243.

²⁾ Epicurus, aangehaald door W. W. Sawyer in *Mathematician's delight*, Nederlands: *Wiskunde zonder omslag*, uitgave Het Spectrum.

vaas met witte en rode ballen" (art. blz. 460). De witte en rode ballen, hier genoemd, zijn identiek met de „goede en foutieve" guldens in de controlereeks. Als „goede guldens" worden aangemerkt de guldens, die onderdelen zijn van een post of posten, die door een document is of zijn gedekt, c.q. waarvan de rangnummers gedekt zijn. Met „foutieve guldens" is dit uiteraard niet het geval. De uitdrukking „goede en foutieve guldens" is van mij afkomstig; de heer Van Heerden spreekt van „foute" of „ongedekte" rangnummers (blz. 461 en 462). Het gaat er nu allereerst om vast te stellen, wat het eenvoudige probleem van de vaas met witte en rode ballen inhoudt.

4. Hoewel dit bekend mag worden verondersteld ga ik hierop nog even dieper in, omdat de uit te brengen kritiek is gebaseerd op een inzicht in het wezen van dit vraagstuk.

Bv.: wanneer men een vaas heeft met 100000 ballen, waarvan 40% wit is en 60% rood, dan is de eerste vraag, waarop geantwoord moet worden: hoeveel witte en hoeveel rode ballen zal men aantreffen bij een trekking van 10 stuks?

De verhouding in de steekproef zal gelijk zijn aan die in de populatie: 4 witte en 6 rode. Dit is echter een verwachting, gebaseerd op het gemiddelde, dat men zal vinden als men een voldoende grote serie van trekkingen van 10 stuks heeft verricht. Met behulp van de standaarddeviatie is te bepalen met een bepaalde mate van waarschijnlijkheid tussen welke grenzen b.v. het aantal te trekken witte ballen zal liggen.

Ook kan de kans of de waarschijnlijkheid worden vastgesteld om een zeker aantal witte of rode ballen of een combinatie te trekken met behulp van de ontwikkelde termen van het binomium van Newton; hier $(0,4 + 0,6)^{10}$.³⁾

Zo berekent men, dat het aantal witte ballen, dat getrokken zal worden in een steekproef van 10 bij een gemiddelde van 4, met een waarschijnlijkheid van 99,74% zal liggen tussen -0,65 en 8,65 (praktisch tussen 0 en 8 à 9). (Eerste probleem.)

De kans op het trekken van 10 witte ballen is slechts zeer gering, namelijk $0,4^{10}$ of 0,0001. (Tweede probleem.)

Er zijn hier blijkbaar twee verschillende problemen aan de orde. Het eerste omvat de begrenzing van het gemiddelde, waarmee men dus een eigenschap van de steekproef, c.q. van de massa (omkeervraagstuk), te weten wil komen. Het tweede vraagstuk gaat om het fixeren van de kans of waarschijnlijkheid van een bepaalde trekking.

Bij de „guldenrangnummermethode" hebben wij te maken met het laatste probleem.

5. Men zal eerst een hypothese moeten maken ten aanzien van de samenstelling van de massa (b.v. 40 % witte en 60 % rode ballen) om te komen tot een begrenzing van de trekking (de steekproef). Is de hypothese juist dan is de limitering correct binnen het aanvaarde waarschijnlijkheidsgebied. De uitspraak geldt voor de individuele trekking van 10 stuks.

³⁾ Er zijn in dit geval 11 mogelijkheden, waarvoor de kansen kunnen worden berekend, namelijk 10 witte ballen en 0 rode, 9 witte ballen en 1 rode, enz., tot en met 10 rode ballen en 0 witte. Elke volledige term geeft de kans aan. De exponenten bij 0,4 geven het aantal witte ballen aan, de exponenten bij 0,6 het aantal rode ballen.

De verhouding in de populatie (massa) is praktisch nooit bekend en men zal dus van een vermoeden moeten uitgaan. Dan spreekt men van een hypothese.

Het tweede vraagstuk van het berekenen van de kans op een bepaalde trekking ligt geheel anders. Ook hier zal men van een hypothese ten aanzien van de verhouding in de populatie moeten uitgaan. Maar de gevonden waarschijnlijkheid is niet te hanteren als uitkomst voor een enkele steekproef. Slechts bij een groot aantal trekkingen nadert de relatieve frequentie tot een grenswaarde. Wijlen Prof. Dr. O. Bakker heeft een proef genomen, waarbij het ging om het werpen van 10 ogen met 2 dobbelstenen. Daarbij bleek, dat eerst na ruim 2200 worpen het getal 0,07 te voorschijn kwam. De grenswaarde was 0,075, omdat het hier „valse” dobbelstenen betrof. Bij zuivere exemplaren zou het resultaat 0,083 zijn geweest. Het vanouds zo dierbare dobbelspel is één van de weinige voorbeelden, dat de uitkomst van tevoren berekenbaar is. ⁴⁾

6. Op bladzijde 463 en 464 van het december-nummer (1961) van het m.a.b. lees ik het volgende betreffende de „guldenrangnummERMETHODE”.

„Er moet echter wel op worden gewezen, dat de voordelen uitsluitend liggen in het technisch-statistische vlak. De uitspraak, die op grond van deze methode gedaan kan worden is identiek gebleven aan die, welke door Prof. de Wolff in zijn voordracht voor de Statistische Dag 1956 is geformuleerd nl.:

Indien in de steekproef geen enkele fout wordt aangetroffen, kan behoudens een waarschijnlijkheid E worden aangenomen, dat het toelaatbare totaalbedrag niet meer dan een fractie p onder het getoonde totaalbedrag van de onderzochte postenreeks ligt. De waarden voor E en p kunnen naar behoefte worden vastgesteld en bepalen de noodzakelijke omvang van de steekproef”.

(De gebruikte Griekse letters heb ik vervangen door de letters E en p).

Men kan dus concluderen, dat de uitspraak alleen iets zegt over de populatie, indien geen enkele fout is geconstateerd. In dit geval geldt met een waarschijnlijkheid E , dat de fout niet groter is dan pT , waarin T het totaal van de boekingen aangeeft.

7. Teneinde duidelijk naar voren te brengen wat deze waarschijnlijkheid E inhoudt is een berekening gemaakt voor de waarde van E bij een bepaalde p . E en p zijn uitgedrukt in procenten en gelden voor een steekproef van resp. drie, vier en vijf stuks. E is uiteraard 100 bij 0 fouten, terwijl E 0 is bij een foutenpercentage van 100. Zowel E als p bewegen zich van 0 tot 100.

Bij de tekening (grafiek I) is een staat gevoegd, vermeldende de bij de verschillende p 's behorende waarden van E , vermenigvuldigd met 1000 ($n=3$), met 10000 ($n=4$) en met 100000 ($n=5$). Het gering aantal elementen van de steekproef maakt het mogelijk het begrip E scherp te analyseren. Het groter worden van het aantal elementen van de steekproef doet E verminderen bij gelijke p .

Men kan uit de grafiek aflezen, dat men een kans op niet-ontdekking aanvaardt

⁴⁾ Men onderscheidt in dit verband twee gevallen:

- a. waarschijnlijkheid *à priori*: het aantal gunstige gevallen en het totaal aantal gevallen zijn beide met wiskundige juistheid te bepalen: de kans om met een zuivere dobbelsteen een zes te gooien;
- b. waarschijnlijkheid *à posteriori*: genoemde grootheden zijn *niet* beide met wiskundige zekerheid te bepalen: de sterftkans bij levensverzekering en de trefkans bij foutencontrole.

(dus 0 foute guldens) van 16% (horizontale as: E) bij een foutenpercentage van 30% voor een steekproef van vijf stuks.

Is het foutenpercentage (p) *kleiner* dan stijgt het risico (voor een foutenpercentage van 10% is het risico 59%), maar dit accepteert men welbewust. Het percentage van 30% fouten heeft schijnbaar het karakter van een maximum. (Zie echter noot 7.)

In het risico (E) zijn echter begrepen b.v. een risico van bijna 8%, dat het foutenpercentage 40% is, een risico van ruim 3% op een foutenpercentage van 50% en een risico van 1% op een foutenpercentage van 60%. Bij daling van E stijgt p bij gelijkblijvende grootte van de steekproef.

Niet-ontdekking betekent niet, dat het foutenpercentage niet meer dan 30% is. De toevoeging „met het risico van 16%” houdt in aanvaarding van *het gehele risicogebied links van $E = 16$* (het gearceerde gedeelte in de grafiek).

De niet ontdekte fout kan aanzienlijk groter zijn dan 30%.

Bij de beschouwing van E mag men zich niet beperken tot het risicomoment. Er is een *risicogebied* van 0 tot 16% voor een foutenpercentage van 100 tot 30%, waarbij het risico steeds kleiner wordt bij stijgende p .

Voor een grotere steekproef nadert de curve steeds meer de horizontale as: E toont een snelle vermindering bij een gelijke p . Dientengevolge is men eerder geneigd het risico te accepteren. Daarbij moet echter bedacht worden, dat limitering van het risico tot 1% bij een fout van 1% omvat het op zich nemen van een (theoretisch) oneindig aantal risico's (risico-momenten), lopend van 0 tot 1%, op fouten van 100 tot 1%.

8. Op blz. 469 (m.a.b. december 1961) wordt een staat gegeven, waarbij men zich heeft gebaseerd op „een bedrijfseconomisch verantwoorde controle-omvang” (blz.

467). Dit houdt in, dat „wij althans een poging moeten doen om te komen tot een tegen elkaar afwegen van nut en kosten”. (blz. 467)

Aanvankelijk was de vorm voor de bepaling van n (de grootte van de steekproef of de controle-omvang):
$$\frac{\log E}{\log (1 - p)} \quad (\text{blz. 460, formule C})$$

(uiteraard zijn hier zowel de Briggse als de natuurlijke logaritmen bruikbaar)
Hier is dus n afhankelijk van E en p .

Voor de bedrijfseconomisch verantwoorde controle-omvang vraagt men zich af bij welke steekproef-grootte de controle-kosten gelijk zijn aan het produkt van het risico van niet-ontdekking en de fout, die bij dit risico behoort. Dan wordt de

formule voor n :
$$\sqrt{\frac{T}{2,7183 \cdot k}} \quad (\text{blz. 468, formule E})$$

waarin T is het totaalbedrag van de postenreeks of het totaalbedrag der te onderzoeken massa en k de controlekosten per te onderzoeken eenheid.

In de toelichting bij tabel II (blz. 469) heeft de auteur het over „de controlekosten per post”. Hiermede zal waarschijnlijk bedoeld zijn „de controlekosten per te onderzoeken gulden”. Immers bij de „guldenrangnummermethode” wordt de idee van controleren van posten als zodanig geheel losgelaten. Het vaste bedrag 2,7183 is de waarde van het natuurlijke getal e . Blijkbaar is n nu geheel afhankelijk van T en k . In tabel II laat men p (en de daarvan afhankelijke E) wisselen bij een bepaalde steekproefgrootte.

9. Indien p de som der „foutieve guldens” voorstelt in een perunage van T (totaal der te controleren boekingen), dan wordt E , d.i. de kans op alleen „goede guldens” in de steekproef van 607 stuks (risico van niet-ontdekking van de fouten): $(1 - p)^{607}$. (a) ⁵⁾

Is p gelijk aan 1647 : 1000000 (f 1647 is het bedrag van de fout, waarbij het risico van niet-ontdekking 0,368 is, zie tabel II, tweede regel) of 1 : 607, dan wordt E :

$$\left(1 - \frac{1}{607}\right)^{607} \approx \frac{1}{e} = \frac{1}{2,7183} = 0,368. \text{ } ^6)$$

Dit betekent, dat men bij een fout van $1/607 \times 1000000 = f$ 1647,- een risico van niet-ontdekking loopt van 0,368 of van 36,8%.⁶⁾

Hier (fout resp. 1647 en 1621) wordt p bekend verondersteld.

In de andere gevallen wordt het risico (E) de bekende grootte, terwijl daaruit p met behulp van formule a wordt afgeleid.

10. De omvang van de steekproef is volgens de auteur verantwoord, indien

$$E \cdot F = n \cdot k$$

waarin E het risico van niet-ontdekking van de fout, F de fout, die bij dit risico behoort, en k de controlekosten per te onderzoeken gulden voorstellen.

⁵⁾ Indien p 0,4 is, dan is q (het perunage der „goede guldens”) 0,6 of $1 - p$. Dit steunt op de gelijkheid: $p + q = 1$. (zie paragraaf 4: $0,4 + 0,6 = 1$)

⁶⁾ Dit is dus de kans op 607 „goede guldens” in de steekproef van 607 stuks. De kans op alle overige combinaties (607 „foutieve guldens”, 606 „foutieve” + 1 „goede” gulden, enz.) is dan $1 - 0,368 = 0,632$.

In het gegeven voorbeeld (totaalbedrag f 1000000,—, controlekosten f 1,— per gulden) wordt de steekproefgrootte n berekend op 607 stuks. Dan wordt bovenstaande gelijkheid:

$$0,368 \cdot 1647 = 607.$$

De premisse is, dat het „nut” van het vinden van een fout bij de accountantscontrole maximaal gelijk is aan het bedrag van de fout. Als tweede uitgangspostelling geldt, dat het risico van niet-ontdekking maal de fout, die bij dit risico behoort, mag worden ingecalculceerd.

De vraag rijst: waarom wordt het risico van niet-ontdekking als een bepalende factor beschouwd voor het nut? Dit wordt niet gemotiveerd. Het wordt eenvoudig zó gesteld en wij zullen dit moeten aanvaarden op gezag van de auteur. Ik ben daartoe niet bereid.

Verder zijn de totale controlekosten afhankelijk van E en F en deze weer van T en k (zie hiervoor: formule E van de heer Van Heerden). E en p zijn niet beslissend meer. Uiteindelijk bepaalt het totaalbedrag der boekingen de steekproefgrootte (bij gelijkblijvende controlekosten per gulden).

11. Tot welke werkwijze leidt nu de beschouwde steekproefmethode? Aan de hand van tabel II zouden we kunnen zeggen: men is bereid een risico van 1 % te lopen ($E = 0,01$) niets te vinden, indien van het totaalbedrag een fractie van $p = 0,00757$ of ruim $\frac{3}{4}$ % (d.i. f 7570,—) niet gedekt zou zijn (bewoordingen van de heer Van Heerden blz. 460, 3de, 4de en 5de regel v.o.). Is het juist te stellen, dat men 1 % risico loopt geen enkele foutieve gulden op het spoor te komen en dat dan de totale fout f 7570,— is? (zie daarbij mijn paragraaf 7).

Men bedenke daarbij wel, dat men niets van de fouten in de populatie afweet. Men moet zijn conclusie baseren op de uitkomst van de steekproef en men doet dit bij de methode-Van Heerden met behulp van een kansberekening, waarvan bewezen is (zie mijn paragraaf 5), dat zij slechts bij zeer hoge uitzondering een juiste waarschijnlijkheid (E) voor een eenmalige steekproef oplevert.

Teneinde een bewijs te leveren voor het gevaar, dat aan de gepropageerde werkwijze verbonden is, is een grafiek vervaardigd aan de hand van tabel II (blz. 469) (grafiek II). Dezelfde gevallen zijn uitgewerkt, waarbij het gemiddelde ($\bar{x}$) der foutieve guldens in de steekproef van 607 stuks is berekend, de standaarddeviatie (Poisson) als tweedemachtswortel uit het gemiddelde en de grenzen (g_1 en g_2), waartussen het genoemde gemiddelde ligt met een waarschijnlijkheid van 95,44 %. Daarbij is dus aangenomen, dat de fout in de populatie bekend is (f 1150,—, f 1647,—, enz.). Dit gebeurt ook in Van Heerden's artikel. In mijn grafiek is bovendien een geval (3a) tussengevoegd.

12. Volgens de redactie van Van Heerden loopt men een risico van 50% niets te vinden, indien een bedrag van f 1150,— niet gedekt is (volgnr. 1). Uit grafiek II blijkt, dat bij het vinden van 0 fouten met een waarschijnlijkheid van ruim 95 % geconcludeerd kan worden, dat de fout in de massa zelfs wel f 6590,— kan zijn. Immers bij deze fout in de populatie zijn de grenzen in de steekproef bij het aangenomen waarschijnlijkheidsgebied (ruim 95 %) 0 en 8. Zie volgnummer 3a. Aanvaardt men een waarschijnlijkheid van 99,74 % (driemaal de standaarddeviatie = $3s$) dan kan de som der foutieve guldens tot f 15070,— oplopen en toch onopgemerkt blijven. Zie volgnummer 6.

Inderdaad vinden we bij v. H. een erkenning van deze mogelijkheid. Hij schrijft namelijk op blz. 461: „Vindt men bij controle van de aldus (door trekking van loterijbriefjes, die corresponderen met de rangnummers der guldens, H.M.W.) aangewezen posten geen fouten, dan staat dus (*behoudens een kans op het tegendeel van 1 %*) vast dat van dat totaalbedrag minder dan f 5000,— ($p \times$ totaalbedrag) ongedekt is.” (Cursivering van mij, H.M.W.)

13. Zowel op blz. 468 als op blz. 474 spreekt de heer Van Heerden van controlekosten per post”. Echter ook nog één keer van „controlekosten per te onderzoeken eenheid”. Dit laatste kan alléén juist zijn, daar de totale controlekosten bij hem gelijk moeten zijn aan $1647 \times 0,368 = f 607,—$, zijnde het aantal te onderzoeken n guldens van de steekproef (607) maal het tarief ($f 1,—$). (tabel II blz. 469)

Verder zouden de kosten bij detailcontrole gelijk zijn aan het totaal te controleren bedrag ad f 1000000,—. Hiermede zou een detailcontrole zonder meer veroordeeld zijn. Het voorbeeld is dus niet reëel te achten.

De statistische grens van de steekproefgrootte kan onder de economische grens liggen.

Zijn de controlekosten f 0,01 dan wordt n 6070. Het kan zijn, dat men een waarschijnlijkheid aanvaardt van 1 % op een onjuiste uitspraak ten aanzien van een fout van 1 %. Volgens tabel I (blz. 458 artikel v. Heerden) is in dit geval n 459. Een universele toepassing van formule E (blz. 468 artikel v. Heerden) moet dus zeker worden afgewezen. De „economische grens” kan soms in strijd zijn met het economisch principe.

De controle van een gulden betekent praktisch de controle van de gehele post, waarin het rangnummer voorkomt. Ook is het mogelijk, dat enige rangnummers in dezelfde post vallen. Een tarief per te onderzoeken eenheid, d.i. per te contro-

leren gulden voorzover deze voorkomt in de steekproef, kan niet voeren tot het juiste bedrag der totale controlekosten.

Het nut van de controle moet gezien worden los van het al of niet vinden van fouten. Daar dit algemeen bekend mag worden gesteld en door de heer Van Heerden niet zal worden tegengeworpen, behoeft hierop niet verder te worden ingegaan.

14. Tenslotte een samenvatting van de conclusies, die uit deze beschouwing voortvloeien.

Wat betekent de kans E? De aanvaarding van een risico-gebied, dat loopt van 0 tot de waarde, die men voor E heeft gekozen. De volgende mogelijkheden zijn begrepen in E:

1. de kans, dat de fout groter is dan $p \times T$ en dat men geen „foutieve gulden” op het spoor komt⁷⁾;
2. de kans, dat de fout groter is dan $p \times T$ en dat men inderdaad enige „foutieve gulden” ontdekt⁷⁾;
3. de kans, dat de fout kleiner is dan $p \times T$ en dat men eveneens „foutieve gulden” constateert;
4. de kans, dat de fout kleiner is dan $p \times T$ en dat de controle geen fouten oplevert; dit risico heeft men *a priori* op zich genomen.

De eerste drie gevallen veroorzaken de moeilijkheden. Bij 1 kan de fout zeer groot zijn, maar „it is all in the game”. Bij 2 en 3 rijst de vraag: wat moet men doen? Het is veelal bij uitzondering, dat men beide gevallen afzonderlijk kan determineren.

De feitelijke problematiek spitst zich in dit geval toe. De moeilijkheid wordt verplaatst, niet opgelost.

De „guldenrangnummermethode” leidt tot een controle van iedere post, waarin de rangnummergulden voorkomt (zie bijv. blz. 463 van het artikel van v. Heerden). De controlekosten zijn dus niet afhankelijk van de theoretische omvang van de steekproef, zoals die op „bedrijfseconomische gronden” is bepaald. Daarmede vervalt formule E van de heer Van Heerden.

Ook kan E (het risico van niet-ontdekking) geen rol spelen bij het bepalen van het bedrag van de nota van de accountant. Immers de „fout”, die beslissend is voor de grootte van E, is nimmer werkelijkheid maar uitvloeisel van een hypothese.

De verrichte arbeid bepaalt de controlekosten. Het nut van die arbeid is van vele imponderabilia afhankelijk. Wellicht overbodig moge ik verwijzen naar het artikel van Prof. A. M. van Rietschoten „De betekenis van de interne controle voor de accountantscontrole”, m.a.b. juni 1954, blz. 245.

⁷⁾ In wezen is p dus geen maximum! Zie hierbij mijn paragraaf 7.