

De Studentized bootstrap methode in de accountantscontrole

Ir. G.H. Van, Prof. Dr. R.J. Grübel, Drs. G.B. Broeze

1 Inleiding

In de accountantscontrole worden statistische methodes gebruikt om een betrouwbaarheidsboven-grens (vaak ook 'maximale fout' genoemd) te construeren voor de fout (in guldens) in een financiële verantwoording. In 1961 heeft Van Heerden zijn baanbrekende artikel over de guldenssteekproef gepubliceerd (Van Heerden 1961). Het idee hiervan is om bij uitspraken over de omvang van een monetaire fout niet de posten, maar de guldens als steekproefeenheid te zien. Het percentage foute guldens is in dit type steekproef een zuivere schatter voor het foutpercentage in de massa. Een verfijning op dit idee is door Stringer (1963) gepubliceerd. Door het gebruik van zogenaamde taintings werd een gulden niet als (helemaal) goed of (helemaal) fout aangemerkt, maar als voor een deel (de tainting) fout. De erbij behorende schatter is - voor zover van steekproeven gebruik gemaakt wordt - alom verspreid in de accountantscontrole en staat bekend als de Stringerbound.

Een nadeel van de Stringerbound is, dat zijn mathematische eigenschappen grotendeels onbekend zijn. Hij blijkt daarnaast, meestal in hoge mate, conservatief te zijn (zie o.m. Bickel 1992).²

Ir. G.H. Van studeerde Technische wiskunde aan de TU-Delft. Sinds 1994 is hij werkzaam als automatiseringsdeskundige bij TAS Top Talent. Prof. Dr. R.J. Grübel was tot 1993 universitair hoofddocent bij de TU-Delft. Thans is hij als professor werkzaam bij het Institut für Mathematische Stochastik, Universität Hannover. Drs. G.B. Broeze studeerde wiskunde aan de Universiteit van Amsterdam, waaraan hij 15 jaar als universitair docent was verbonden (faculteit Sociale Wetenschappen). Sinds 1986 is hij als statisticus werkzaam bij de Algemene Rekenkamer.

Ook andere 'standaard' statistische technieken, onder meer die welke op de normale verdeling zijn gebaseerd, leveren geen bevredigende resultaten (zie bijv. Tamura et al. 1989). Dit hangt samen met het feit dat de meeste posten in een verantwoording in het algemeen correct zijn geboekt, waardoor een steekproef doorgaans slechts weinig fout geboekte posten bevat. Daarmee bevat de steekproef dus ook weinig informatie over de verdeling van de fouten. Omdat daarnaast de verdeling van de fouten vaak in hoge mate afwijkt van de veronderstelde normale verdeling, is simpelweg vergroten van de steekproefomvang niet toereikend (zie ook Bauer 1988). Dat is bovendien vaak te duur.

In dit artikel komt de vraag aan de orde of de zogenaamde 'bootstrap methodes' een uitweg uit deze problematiek leveren.

2 Bootstrap methodes

Vanaf ongeveer 1960 wordt met behulp van de computer statistisch onderzoek gedaan naar methodes die beter gebruik maken van de in de steekproef aanwezige informatie. De bootstrap methodes staan hierbij zeer in de belangstelling.

Een sleutelbegrip in bootstrap methodes is de zogenaamde "empirische verdeling". De empirische verdeling van een variabele is diens (cumulatieve) frequentieverdeling in de steekproef. Bootstrap methodes hebben als gemeenschappelijke noemer dat zij zich rechtstreeks baseren op deze empirische verdeling. Hierin onderscheiden zij zich van veel statistische methodes die zich baseren op de veronderstelde verdeling van de variabele in de populatie. Men neemt dan aan dat deze tot een bepaalde familie behoort (bijvoorbeeld tot de bekende normale verdeling(en)). Die aanname maakt het mogelijk om met het gebruik

van (schattingen van) slechts enkele parameters (bij normale verdeling gemiddelde en standaarddeviatie) uitspraken te doen. Als de aannames betreffende de verdeling kloppen zijn deze methodes sterk, als ze niet kloppen is er een probleem. Om dat op te lossen zijn zogenaamde 'nonparametrische of verdelingsvrije' methodes ontwikkeld.

Bootstrap methodes nu opereren niet slechts verdelingsvrij, maar maken gebruik van alle informatie die in de empirische verdeling aanwezig is, dus ook van de grootte van de in de steekproef gevonden waarden voor de variabele.

Bij een schatting van het gemiddelde van een variabele in een te onderzoeken populatie, kan dit onder meer door het herhaald trekken van steekproeven uit de al beschikbare steekproef. Door, bijvoorbeeld 500 keer, zo'n 'resample' te trekken kan men 500 keer het bij het resample behorende gemiddelde berekenen. Voor een 95% betrouwbaarheidsbovengrens neemt men nu het 95-e kwantiel van de 500 gemiddelden.

In het onderzoek waarvan dit artikel verslag doet wordt een verfijning van de methode gebruikt, de 'Studentized bootstrap'. In een appendix wordt deze methode nader toegelicht.

3 Het onderzoek

Dit artikel is gebaseerd op een (simulatie-) onderzoek, dat door Van als afstudeeropdracht voor de TU-Delft bij de Algemene Rekenkamer te Den Haag werd uitgevoerd. Het onderzoek had betrekking op de toepasbaarheid en kwaliteit van diverse schattingsmethodes (evaluatiemethodes van steekproefresultaten) voor de accountantscontrole. In dit artikel wordt voornamelijk ingegaan op onderzoeksresultaten betrekking hebbende op de Studentized bootstrap methode, in vergelijking met de Stringerbound. Voor een volledig overzicht wordt verwezen naar Van (1993).

In het onderzoek is gekeken naar het gedrag van schattingsmethodes, waarbij:

- onderboekingen in een verantwoording buiten beschouwing worden gelaten: het onderzoek is gericht op verantwoordingen die geen onderboekingen bevatten,
- er wordt uitgegaan van 'Probability Proportional to Size' (pps)-steekproeven; deze zijn verkregen door het trekken van monetaire eenheden (zonder teruglegging).

Deze wijze van trekken levert een klassieke guldensteekproef op.

De schattingsmethodes zijn onderzocht met behulp van met de computer gegenereerde populaties van posten. Van zo'n populatie is het werkelijke foutpercentage (of -bedrag) bekend. Uit deze gegenereerde populatie werd (pps) een steekproef getrokken. In de getrokken steekproef werd de (95%) betrouwbaarheidsbovengrens (maximale fout) berekend met behulp van (onder meer) de Stringerbound en de Studentized bootstrap methode. Omdat de werkelijke fout bekend is, kan worden vastgesteld of de berekende maximale fout ook - als gewent - groter is dan de werkelijke fout. Als gevolg van steekproeffluctuaties hoeft dit niet noodzakelijk zo te zijn, óók bij correcte methodes. Er is immers 'genoegen genomen' met een onbetrouwbaarheid van 5%.

Door het herhaald trekken van steekproeven (in het onderzoek zijn 1000 steekproeven per populatie getrokken) kan worden vastgesteld wat (bij benadering) de betrouwbaarheid van de methodes is. Een bij een gekozen betrouwbaarheid van 95% correcte methode zal in rond 950 van de 1000 keer een maximale fout opleveren die groter is dan de werkelijke fout, met andere woorden een overdekkingspercentage hebben in de buurt van 95%.

Het spreekt vanzelf dat de bruikbaarheid van het (simulatie-)onderzoek staat of valt met de realiteitswaarde van de gegenereerde populaties, waar het de verdeling van de fouten over de posten betreft. Om in dit opzicht realistische populaties te kunnen genereren, zijn de verdelingseigenschappen van 5, na anonimisering, door de accountantsdienst van het Ministerie van (toen nog) Onderwijs en Wetenschappen ter beschikking gestelde datasets onderzocht in een data analyse. De eigenschappen zijn in een aantal van de gesimuleerde populaties ingegeven.

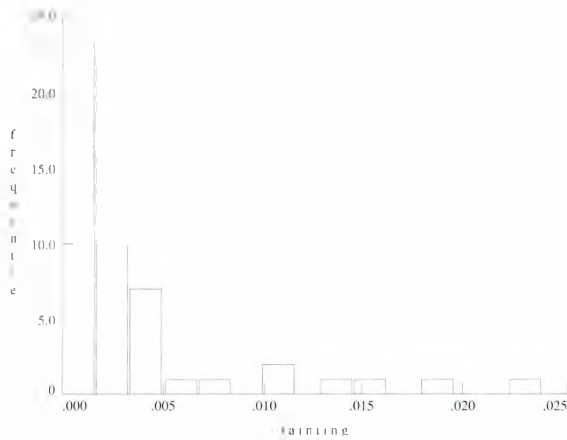
De in dit artikel opgenomen onderzoeksresultaten geven aan dat de Studentized bootstrap methode in deze relevante situaties bij gelijke nominale betrouwbaarheid nauwkeuriger bovengrenzen levert dan bijvoorbeeld de Stringerbound methode.

4 Data analyse

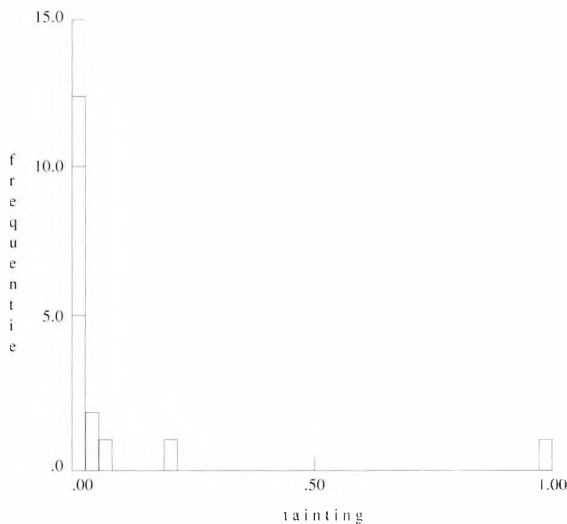
In de analyse van een verantwoording worden de volgende karakteristieke punten onderscheiden:

- het aandeel fout geboekte ('getainte') posten,
- de verdeling van de boekwaarden,
- de verdeling van de taintings van de fout geboekte posten.

In de figuren 1 en 2 zijn histogrammen van de taintings van de fout geboekte posten van twee



Figuur 1: Histogram van de taintings van de getainte posten bij een van de onderzochte datasets



Figuur 2: Histogram van de taintings van de getainte posten bij een van de onderzochte datasets

van de 5 onderzochte datasets afgebeeld.

De figuren 1 en 2 zijn representatief voor alle 5 onderzochte datasets. De histogrammen duiden op een zeer scheve verdeling (in niet-cumulatieve vorm ook *kansdichtheid* genoemd) van de taintings van de getainte posten. Het grootste deel van de kansmassa is gelegen bij de kleine waarden van de taintings. Zoals uit figuur 2 blijkt kan zich de situatie voordoen dat een steekproef een of meerdere 100%-overboekingen bevat, zodat de kansdichtheid dan een massapunt bevat bij de waarde 1 voor de taintings.

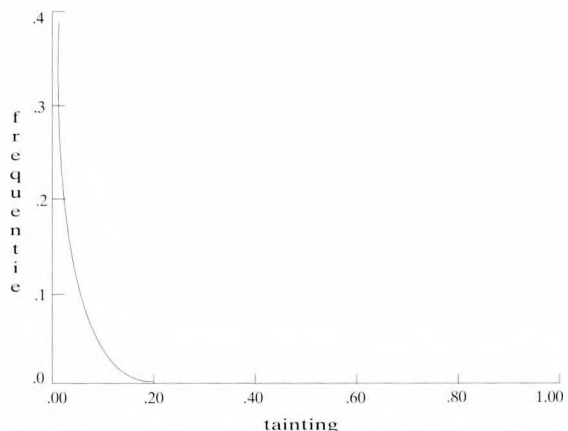
5 Opzet van de simulatiestudie

De in paragraaf 3 in hoofdlijnen weergegeven opzet van de simulatiestudie kan nu als volgt

nader worden ingevuld. Met behulp van een computer wordt een populatie aangemaakt (met een bekende fout!) die bestaat uit (boekwaarden van) posten en bijbehorende taintings. Hierbij wordt gevarieerd op de grootte van de populatie, het percentage getainte posten en de verdeling van de taintings groter dan nul. Voor deze verdeling is op grond van de data-analyse (par. 4) gekozen voor een Beta(α, β)-verdeling (uitgezonderd de situatie van 100%-overboekingen), waarbij over de parameters α en β kan worden gevarieerd (zie voor de eigenschappen van de Beta-verdeling Johnson & Kotz (1970a)). Voor de verdeling van de boekwaarden is, eveneens op grond van data-analyse, gekozen voor een vaste exponentiële verdeling (zie eveneens Johnson & Kotz (1970b)).

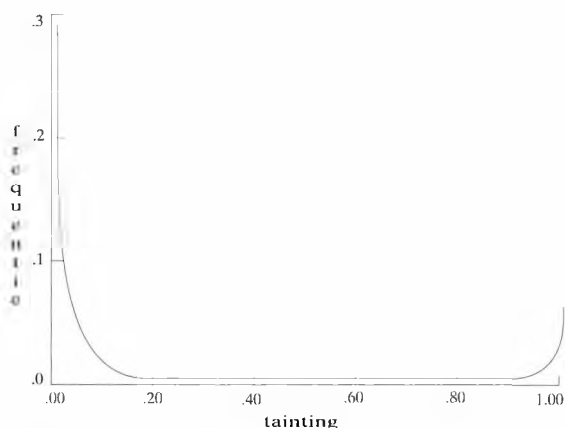
Uit een aldus gegenereerde populatie wordt herhaalde malen een pps-steekproef getrokken. Er kan hierbij gevarieerd worden over de grootte van een steekproef en het aantal trekkingen van (telkens verschillende!) steekproeven; standaard is echter gekozen voor 1000 trekkingen. Bij het onderzoek is voornamelijk gebruik gemaakt van de populatiegroottes 5000 en 10.000, waarbij de getrokken steekproeven in grootte variëren van 50 tot 500 posten. Het ingestelde percentage getainte posten in een populatie varieert van 5% tot 20%. Bij deze instellingen varieert het percentage fout geboekte geldeenheden (guldens) van 0.05% tot 10%.

In het verrichte onderzoek worden vijf typen taintverdelingen van getainte posten onderscheiden, waarvan met name de typen I en II voor accountancy doeleinden interessant zijn. (Data met een taintverdeling van getainte posten van bijv. het type I worden in dit artikel verder standaard aangeduid met data van het type I). De typen I (zie ook figuur 1) en II worden uitgebeeld in de figuren 3 en 4 in deze paragraaf.



Figuur 3: Type I; gespiegelde J-vorm

Onder het type I vallen de taintverdelingen van getainte posten die als dichtheidsfunctie een gespiegelde J-vorm bezitten; variërend van een zeer korte staart (steile gespiegelde J-vorm) tot een lange staart (waarbij de kansmassa gelegen is over het hele bereik van de taintings). Het type I representeert de situatie dat er (gerelateerd aan de boekwaarde) meer kleine fouten zijn gemaakt dan grote fouten. Een praktijkvoorbeeld van de gespiegelde J-vorm is al gegeven in figuur 1. Een taintverdeling van het type I wordt verkregen met behulp van een Beta(α , β)-verdeling met $0 < \alpha < 1$ en $\beta > 1$.



Figuur 4: Type II; linker U-vorm

Onder het type II vallen de taintverdelingen van getainte posten die als dichtheidsfunctie een U-vorm bezitten 'waarbij de linkerpoot van de U groter is dan de rechterpoot', ofwel, waarbij de meeste kansmassa links van het midden ligt. Dit geeft, ruwweg gezegd, aan dat de meeste fouten (gerelateerd aan de boekwaarde) of klein of groot zijn, waarbij er echter in totaal meer kleine dan grote fouten zijn. Een taintverdeling van het type II wordt verkregen met behulp van een Beta(α , β)-verdeling met $0 < \alpha < \beta < 1$.

In de simulatiestudie zijn voorts de volgende keuzen gemaakt:

- voor de populaties/verantwoordingen die zijn aangemaakt volgens de typen I en II is de tainting van een post onafhankelijk gekozen van de bijbehorende boekwaarde; deze onafhankelijkheid werd geïndiceerd door de data-analyse;
- steekproeven met minder dan 2 getainte posten worden uitgesloten van trekking, wegens de noodzakelijke berekening van de steekproefstandaarddeviatie. Dit levert een enigszins vertekend beeld op van de werkelijke kwaliteit

van de onderzochte methodes, echter, bij steekproeven met een voldoende aantal getainte posten is dit verwaarloosbaar. Het minimale verwachte aantal getainte posten in een steekproef is gesteld op 5 (bijvoorbeeld gerealiseerd met het trekken van een steekproef ter grootte van minimaal 100 posten uit een populatie met 5% getainte posten).

6 Resultaten

De conclusies van het onderzoek zijn gebaseerd op simulatie-gegevens, verkregen in enkele honderden uren computer(reken)tijd. In deze paragraaf wordt een representatief deel van de resultaten getoond in de tabellen 1 en 2.

Verdeling van getainte posten	: Beta (0.80;50.00)
Populatiegrootte	: 10000
Steekproefgrootte	: 50 (en 250)
Foutpercentage in posten	: 15.00%
Foutpercentage in guldens	: 0,24%
Aantal steekproeftrekkingen	: 1000

gewenst	Studentized bootstrap	Stringerboud
0.800	0.805 (0.791)	1,000 (1,000)
0.900	0.881 (0.892)	1,000 (1,000)
0.950	0.935 (0.943)	1,000 (1,000)
0.990	0.976 (0.988)	1,000 (1,000)

Tabel 1: Onderzoeksresultaten bij data van het type I

De tabel kan als volgt gelezen worden. In de linkerkolom van de tabel staan de nominale betrouwbaarheden, waarbij bijvoorbeeld 0,800 aangeeft dat in het ideale geval bij 80% van de steekproeven een interval wordt bepaald waarin de werkelijke fout van de populatie is gelegen. De overige kolommen geven de overdekkingspercentages van de onderzochte methodes, gebaseerd op 1000 herhaalde steekproeftrekkingen. Hierbij geeft bijvoorbeeld 0,805 aan dat bij 805 van de 1000 steekproeftrekkingen (bij een steekproefgrootte van 50 posten) een correct interval werd bepaald; bij een steekproefgrootte van 250 posten was dit het geval bij 791 van de 1000 trekkingen. Merk opnieuw op dat zelfs bij een ideale methode niet verwacht mag worden dat de nominale betrouwbaarheid precies wordt behaald in 1000 trekkingen; bij 1000 maal opwerpen van een zuivere munt mag evenmin verwacht worden dat er precies 500 keer kop boven komt.

Tabel 1 is typerend voor de vele gevonden resultaten bij data van het type I. Het blijkt dat

met behulp van de Studentized bootstrap methode de nominale betrouwbaarheden dicht worden benaderd. De Stringer (upper)bound methode is zeer conservatief: de overdekkingspercentages die met deze methode worden verkregen zijn doorgaans gelijk aan 1.

Verdeling van getainte posten :	Beta (0,20;0,40)	
Populatiegrootte :	10000	
Steekproefgrootte :	50 (en 250)	
Foutpercentage in posten :	10,00%	
Foutpercentage in guldens :	3,20%	
Aantal steekproeftrekkingen :	1000	
gewenst	Studentized bootstrap	Stringerbound
0.800	0.885 (0,825)	0.995 (0,895)
0.900	0.944 (0,930)	1.000 (0,967)
0.950	0.977 (0,979)	1.000 (0,987)
0.990	0.993 (0,999)	1.000 (0,998)

Tabel 2: Onderzoeksresultaten bij data van het type II

Tabel 2 geeft inzicht in de prestaties van de onderzochte methodes bij data van het type II. Evenals bij het type I geldt dat de Stringer (upper)bound methode zeer conservatief is, zij het bij type II minder dan bij type I. De Studentized bootstrap methode is doorgaans conservatief, maar steeds minder dan de Stringerbound methode. Bovendien tendert de Studentized bootstrap naar minder conservatisme bij toenemende steekproefgrootte.

In het simulatie-onderzoek zijn ook populaties met een rechter U-vorm (meer grote dan kleine taintings) en een J-vorm (vooral grote taintings, maar geen puntmassa in 1) onderzocht. Hierbij is met een te gering aantal herhalingen gewerkt om stellige uitspraken te kunnen doen. De steeds aanwezige tendens was echter dat Studentized bootstrap evenals de Stringerbound in deze gevallen conservatieve bovengrenzen oplevert, zij het minder extreem. Als de verdeling van de taintings een puntmassa in 1 heeft, geeft de Studentized bootstrap te lage overdekkingspercentages en kan dus niet zonder meer gebruikt worden.

7 Conclusie

De introductie van de Studentized bootstrap methode in de accountantscontrole lijkt, gezien de gevonden resultaten, een welkome aanvulling op de reeds bestaande en toegepaste methodes. Vanwege de diversiteit van verdelingen voor

getainte posten en de onvermijdelijke handicap van een gering aandeel getainte posten in een accountancy steekproef, lijken zowel de Stringerbound methode als de Studentized bootstrap methode belangrijke evaluatietechnieken in de accountantscontrole.

De grote kracht van de Stringerbound is dat deze methode geen variantieschattingen vereist en daarom toepasbaar is bij steekproeven die geen of slechts weinig getainte posten bevatten. De Studentized bootstrap methode daarentegen kan slechts toegepast worden als een steekproef minimaal twee getainte posten bevat. Bovendien geldt dat de Studentized bootstrap methode bij steekproeven met weinig (minder dan 5) getainte posten nog onvoldoende is onderzocht en daarom niet toegepast kan (mag) worden.

Het is in de praktijk van de accountantscontrole niet uit te sluiten dat 'slechte' methodes onder specifieke omstandigheden betere prestaties leveren dan 'goede' methodes. Omdat de precieze omstandigheden in werkelijkheid echter nooit (voldoende) bekend zijn, dient men desondanks een algemeen geldende richtlijn te volgen.

Op grond van de verkregen onderzoeksresultaten kan worden geconcludeerd dat - onder de in het artikel gestelde restricties en bij data van het type I of II - de Studentized bootstrap methode de meest nauwkeurige betrouwbaarheidsbovengrens levert voor de fout in een verantwoording, tenzij sprake is van een steekproef met minder dan 5 getainte posten. In het laatste geval wordt de Stringerbound methode geadviseerd.

Aantal getainte posten in de steekproef	
< 5	≥ 5
Stringerbound methode	Studentized bootstrap methode

Tabel 3: Aanbevolen methodes (onder de in het artikel gestelde restricties en bij data van het type I of II)

8 Bruikbaarheid

De bruikbaarheidstabel van de vorige paragraaf impliceert een behoorlijke beperking van onze bevindingen. Immers in heel veel situaties zal men steekproeven vinden met minder dan 5 fouten. Al dan niet in combinatie met deze beper-

king zal men vaak ook geen populatie van het type I (gespiegelde J-verdeling van de taintings) mogen veronderstellen (bijvoorbeeld ingeval een steekproef een tainting van omvang I bevat, zoals in figuur 2).

Er zijn echter situaties waarop de bevindingen van toepassing zijn (zie paragraaf 5). Bij controles in de overheidssector zal men vooral populaties met relatief veel fouten tegenkomen. Dit komt onder meer doordat in foutdefinities bij controle van overheidsjaarrekeningen, formele aspecten zwaarder wegen dan bij controle van bedrijfsjaarrekeningen. Ook bij interne controle of controle van voorraden zal men situaties met relatief veel fouten tegen kunnen komen. Dit betekent dat gebruik van onze bevindingen in een niet te verwaarlozen aantal gevallen tot de mogelijkheden zal behoren.

LITERATUUR

- Bauer, K.P., (1988), *Der Einsatz stichprobengestützter Prüfungsmethoden unter besonderer Berücksichtigung mathematisch-statistischer Verfahren*. Presentatie Niederlandisch-Deutscher Workshop, Treuarbeit AG Frankfurt.
- Bickel, P.J., (1992), Inference and Auditing: The Stringer-bound, *Int. Statist. Rev.* 60, 179-209.
- Efron, B., (1979), 'Bootstrap methods: Another look at the jackknife.' *Ann. Statist.* 7, 1-26.
- Efron, B., (1982), *The Jackknife, the Bootstrap and Other Resampling Plans*. SIAM, Philadelphia.
- Fox, J. and J.S. Long, (1990), (second printing), *Modern methods of Data Analysis*, Sage Publications, Newbury Park.
- Hall, P., (1992), *The bootstrap and Edgeworth expansion*. Springer, New York.
- Heerden, A. van, (1961), Steekproeven als middel van accountantscontrole. *Maandblad voor Accountancy en Bedrijfs-huishoudkunde* 35, 453-475.
- Johnson & Kotz, (1970a), *Continuous Univariate Distributions-2*, Wiley & Sons, New York.
- Johnson & Kotz, (1970b), *Continuous Univariate Distributions-1*, Wiley & Sons, New York.
- Leslie, D.A., A.D. Teitlebaum and R.J. Anderson, (1980), *Dollar Unit Sampling, a practical guide for auditors*. Pitman Publishing Limited, London.
- Pap, G. and M.C.A. van Zuijlen, (1995), On the asymptotic behaviour of the Stringerbound' to appear in: *Statistica Neerlandica* Vol. 49
- Stringer, K.W., (1963), *Practical aspects of statistical sampling in auditing*. Proceedings of the American Statistical Association.
- Tamura et al.: Panel on Nonstandard Mixtures of Distributions, (1989), *Statistical Models and Analysis in Auditing.* *Statistical Science* 4, 2-33.
- Van, G.H., (1993), *Bootstrap methodes in auditing*, TU Delft / Algemene Rekenkamer Den Haag.

NOTEN

- 1 De auteurs danken M.A.M. de Bruijn RA, werkzaam bij de Accountantsdienst van het Ministerie van Onderwijs en Wetenschappen, voor zijn adviezen bij de totstandkoming van dit artikel.
- 2 In een door Pap en Van Zuijlen uitgevoerde studie blijkt het conservatisme voor betrouwbaarheden kleiner dan 50% niet altijd op te gaan (zie Pap en Van Zuijlen (1995)).
- 3 De naam 'bootstrap' is bedacht door Brad Efron, die in 1979 de mogelijkheden van de bootstrap methode in een publikatie beschreef (Efron, 1979). In 1982 verscheen Efron (1982), een beschrijving van o.a. de bootstrap methode in boekvorm.