

HET PROGRAMMA PCALC VOOR HET UITVOEREN VAN KANSBEREKENINGEN

J.G. Bethlehem* en J.M. Buhrman**

1. Inleiding

PCALC is een statistische scherm-rekenmachine. Op het scherm van de micro-computer kunnen gewone en statistische berekeningen worden uitgevoerd. Daarbij kan gebruik worden gemaakt van variabelen, zodat het gebruik van kladpapier overbodig wordt. Voor wat betreft het gewone rekenwerk onderscheid PCALC zich niet van andere programmatuur op dat gebied. Het bijzondere van PCALC ligt in de statistische mogelijkheden: het programma kan niet alleen kansen uitrekenen bij een aantal klassieke verdelingen (binomiaal, Poisson, exponentieel, normaal, χ -kwadraat, t en F), maar ook krietieke waarden, verwachtingen en varianties.

PCALC is gemaakt door de Vakgroep Economie van de Rijksuniversiteit Groningen. Het kan worden besteld door overmaken van f 100,- op gironummer 1162221, t.n.v. S. Knypstra, onder vermelding van PCALC. Universiteiten betalen slechts f 10,- (administratiekosten). Het programma draait op MS-DOS microcomputers. Het is een zogenaamd 'memory-resident' programma. Dat betekent dat het (nadat het eenmaal is geladen) op elk willekeurig moment kan worden opgeroepen, of men nu bezig is in een tekstverwerker, spreadsheet, of welke programma dan ook. Alleen het gebruik binnen grafische programma's wordt afgeraden.

* Centraal Bureau voor de Statistiek, Hoofdafdeling Automatisering, Postbus 959, 2270 AZ Voorburg.

** Statisticalc, Van Dijkstraat 10, 1111 ND Diemen

2. Installatie

Op de geleverde diskette staan drie programma's:

- (1) PCALC.COM : Het eigenlijke rekenprogramma.
- (2) UITLEG.COM: Dit programma legt uit hoe u PCALC kunt gebruiken.
- (3) OPROEP.COM: Met dit programma stelt u de toetscombinatie in waarmee u PCALC oproept.

Daarnaast bevat de diskette nog een subdirectory VERSIE87. Hierin staan weer de bovengenoemde drie files, maar deze versies ondersteunen het gebruik van de 8087 mathematische co-processor.

Installeren van PCALC bestaat slechts uit het kopiëren van de drie files naar de juiste directory. Het laden en 'memory-resident' maken van PCALC gebeurt door de aanroep PCALC (voor een kleurenmonitor) of PCALC MONO (voor een monochrome monitor). PCALC kan daarna worden geactiveerd met toetscombinatie <ALT> P. Wordt de voorkeur gegeven aan een andere toetscombinatie dan kan deze worden ingesteld met het programma OPROEP.COM.

Na activering gebruikt PCALC de rechterhelft van het scherm als een soort rekenblad. Na laden met het commando PCALC MONO is op het scherm niet duidelijk de scheiding tussen rekenblad en rest van het scherm te zien. Wij gaven er daarom de voorkeur aan PCALC te laden alsof we een kleurenmonitor hadden. Bij gebruik van PCALC op een kleurenmonitor blijken niet alle kleuren even gelukkig te zijn gekozen: witte en gele tekst op een lichtblauwe achtergrond is slecht leesbaar. Het is niet mogelijk om de kleurinstelling te veranderen.

Het resident-karakter van het programma bleek niet altijd even goed te werken. Bij aanroep van PCALC uit de tekstverwerker PC-Write en vragen om uitleg met de toets F1, blijkt de uitleg soms, maar niet altijd, links i.p.v. rechts op het scherm te komen en daarna ook in de tekst te zijn opgenomen. Daarbij bleek de originele tekst dan te zijn overschreven. Omdat <ALT> P ook in de gebruikte tekstverwerker een gebruikte combinatie is, verschijnt het daarbij behorende klaver-symbool na elke aanroep van PCALC in de tekst. PCALC bleek niet te werken vanuit MULTIPLAN. Het werkte wel vanuit TURBO-PASCAL 3.0. PCALC werkte niet vanuit SIDEKICK, maar SIDEKICK wel vanuit PCALC. Kennelijk doet PCALC iets met DOS: soms blijkt na terugkeer

uit een programma de computer behoefte te hebben aan diskette met COMMAND.COM, terwijl dit zonder gebruik van PCALC niet het geval is.

Bij het programma zit geen documentatie op papier. De uitleg die door het programma wordt gegeven is in principe voldoende. Toch zullen er gebruikers zijn die wat meer willen weten over de gebruikte algoritmen en hun nauwkeurigheid. Hierover wordt geen informatie gegeven.

3. Werken met PCALC

PCALC is handig voor de schrijvers van statistiekboeken, de makers van statistieksommen en voor alle andere mensen die wel eens met kansen rekenen. Moeiteloos kunnen willekeurige kansen en kritieke waarden worden berekend uit de volgende verdelingen:

Binomiaal	Normaal	Exponentieel	Students t
Poisson	Homogeen	χ^2	F

De homogene en de exponentiële zijn geen huzarenstukjes; die kunnen met iedere brugklas-calculator worden berekend. Voor de volledigheid horen ze er natuurlijk bij. De andere verdelingen zijn interessanter, en met name de χ^2 -, t- en F-verdeling mogen als een waardevolle aanvulling op zowel tabellen als calculators worden beschouwd. De hypergeometrische ontbreekt, en die is voor sommige toepassingen toch wel handig. Het is echter de vraag of de relatief vele moeite voor deze verdeling opweegt tegen het relatief geringe gebruik ervan.

Berekenen van kansen gaat in het algemeen in twee stappen. Eerst wordt een stochast geïntroduceerd door aan een vrij gekozen letter een verdeling toe te kennen. Een voorbeeld: door de opdracht

$$x \sim P(2)$$

krijgt x een Poissonverdeling met verwachting 2. Wil men nu de kans $P(x \geq 3)$ weten, dan krijgt men het antwoord met

$$P(x \geq 3)$$

PCALC is getest bij het schrijven van een statistiekboek. Vanuit een tekstverwerker (PC-Write) werd PCALC aangeroepen om sommen uit te rekenen. Daar-

bij werd een vergelijking gemaakt met een programmeerbare calculator, de CASIO fx-4000P. De rekensnelheid van PCALC steekt gunstig af bij die van de fx-4000P. De meeste berekeningen duren niet langer dan 0,5 à 2 sec. Alleen voor zeer uitzonderlijke berekeningen is meer tijd nodig. De fx-4000P doet over lijvige berekeningen ruim viermaal zo lang. Ook de nauwkeurigheid lijkt ruim voldoende, minimaal 8 à 9 decimalen. Van botte normale benaderingen is zeker geen sprake. Het begint pas lang te duren bij het berekenen van bijvoorbeeld

$$P(x \leq 1500 \mid \text{Poisson}(1500^2/3))$$

de uitkomst 0,5000 vergt 25 sec. Met de functietoets F5 kan het aantal decimalen worden ingesteld. Helaas heeft dit echter geen invloed op al uitgevoerde berekeningen, zodat die dan weer opnieuw moeten worden uitgevoerd.

Als PCALC wordt aangeroepen vanuit een tekstverwerker, dan is dus PCALC actief en niet de tekstverwerker. Dit blijkt een nadeel op het moment dat het antwoord van een berekening moet worden overgenomen in de tekst. Om de tekstverwerker weer actief te maken moet PCALC worden verlaten, maar dan verdwijnt tevens het antwoord van het scherm. Er is dus eigenlijk toch een stukje papier nodig om het antwoord in de tekst te krijgen. Wat PCALC in deze situatie ontbeert is de mogelijkheid of om passief op het scherm aanwezig te blijven, of om de resultaten even naar file weg te schrijven, waarna deze informatie later in de tekstverwerker weer kan worden geïmporteerd.

Op een calculator kan verder worden gewerkt met de resultaten van een berekening, bijvoorbeeld 0,5 van de uitkomst aftrekken. Bij PCALC kan dit niet, maar dat is geen echt probleem. Aangezien PCALC de mogelijkheid biedt met variabelen te werken, is het probleem zonden moeite op te lossen. Stel dat de stochast x is geïntroduceerd. Door de aanroep:

$$q = P(x \leq 2500)$$

wordt de waarde van de betreffende kans aan de variabele q toegekend. Hierna kan de variabele q worden gebruikt tot hij een andere waarde krijgt, of de computer uitgaat.

Voorts heeft PCALC twee operatoren $E()$ en $V()$. Ze geven de verwachting en variantie van de stochast tussen de haakjes. Op deze wijze kan men gemakkelijk een kansberekening vergelijken met een normale benadering. Zo levert bijvoorbeeld

$$x \sim \text{bin}(100, 0.4)$$

$$P(x \leq 35)$$

de uitkomst 0.17947, terwijl

$$z \sim N(E(x), V(x))$$

$$P(z < 35.5)$$

de uitkomst 0.17916 geeft.

Tenslotte kan men kritieke waarden berekenen. Men geeft (bij een reeds gedefinieerde stochast x) eenvoudig de opdracht

$$P(x \leq k) = 0.025$$

en PCALC zoekt voor continue verdelingen de oplossing van de vergelijking en kent die waarde toe aan de variabele k . Voor discrete verdelingen worden de twee waarden van k gegeven waarmee de kans net onder resp. net boven de opgegeven 0,025 liggen.

Nadat een verdeling is gedefinieerd, wordt hiervan een keurige melding op het scherm gedaan. Het is alleen merkwaardig dat de vorm waarin PCALC deze melding geeft, niet als definitie wordt geaccepteerd. Is men de draad kwijt, dan kan men van alle bestaande variabelen (getallen en stochasten) een volledig overzicht krijgen.

Wil men het gehele computergeheugen voor andere doeleinden gebruiken, dan kan men PCALC op eenvoudige wijze verwijderen. Het is echter verstandig dit niet van uit een ander programma te doen, want dat wordt dan ook meteen verlaten met alle gevolgen van dien.

4. Nauwkeurigheid en snelheid

Om een idee te geven van de snelheid van de berekeningen is PCALC op een Olivetti M21 vergeleken met calculator CASIO fx-4000P (zie Buhrman, 1987). Voor een indruk van de nauwkeurigheid zijn de uitkomsten van PCALC vergeleken met Abramowitz en Stegun (1970) of Biometrika Tables, zie Pearson en Hartley (1970). In tabel 1 verwijst de letter A naar Abramowitz en Stegun, de letter B naar Biometrika Tables en de letter C naar de calculator CASIO fx-4000P.

Tabel 1. Een vergelijking van nauwkeurigheid en snelheid

Berekening	Uitkomst	Tijd (sec)	Bron
$P(\underline{x} \leq 24 \mid \text{bin}(74, .5))$	0,00168 61166 668 0,00168 61166 67	ca. 1,0 ca. 4,2	C
$P(\underline{x} \leq 100 \mid \text{bin}(500, .22))$	0,15236 42533 092 0,15237 42544	ca. 3,2 ca. 16,5	C
$P(\underline{y} \leq 10 \mid \text{Poisson}(15))$	0,11846 44115 3 0,11846 44115	ca. 0,3 ca. 1,2	C
$P(\underline{y} \leq 130 \mid \text{Poisson}(150))$	0,05318 28062 02 0,05318 28053 9	ca. 2,9 ca. 13	C
$P(\underline{y} \leq 1500 \mid \text{Poisson}(1500^2/3))$	0,50000 22054	ca. 25	
$P(\underline{y} \leq 3000 \mid \text{Poisson}(3000^2/3))$	0,50000 99352	ca. 50	
$P(\underline{t} > 2 \mid \nu=20)$	0,29632 768 0,02963	ca. 0,5	B
$P(\underline{t} > 2 \mid \nu=120)$	0,02387 926 0,02388	ca. 0,7	B
$P(\underline{t} > 1,96 \mid \nu=200)$	0,02569 24213 79065	ca. 0,8	
$P(\underline{t} > 1,96 \mid \nu=201)$	0,02499 78951 47888 (Standaardnormaal als $\nu > 200$) 0,02499 78951 48220 (standaardnormaal)		A
$P(\underline{\chi}^2 > 5 \mid \nu=2)$	0,08208 49986 0,08208 49986 ($= e^{-2,5}$)		C
$P(\underline{\chi}^2 > 10 \mid \nu=5)$	0,07523 525 0,07524		B

Tabel 1. Een vergelijking van nauwkerigheid en snelheid (vervolg)

Berekening	Uitkomst	Tijd (sec)	Bron
$P(\chi^2 > 60 \mid \nu=50)$	0,15724 203 0,15724		B
$P(F > 2 \mid \nu_1=4, \nu_2=8)$	0,1875 0,1875 (via $P(\underline{x} \leq -1 \mid \text{bin}(5, .5)))$)		C
$P(F > 1 \mid \nu_1=2, \nu_2=100)$	0,37152 78823 82 0,37152 78821 (via $P(\underline{x}=0 \mid \text{bin}(50, 1/51)))$)	ca. 0,7	C
$P(F > 2 \mid \nu_1=2, \nu_2=100)$	0,14071 26157 0,14071 26153 (via $P(\underline{x}=0 \mid \text{bin}(50, 1/26)))$)	ca. 0,7	C
$P(F > 2 \mid \nu_1=50, \nu_2=100)$	0,00168 61166 66 0,00168 61166 67 (via $P(\underline{x} \leq -24 \mid \text{bin}(74, .5)))$)	ca. 1,5	C

5. Gebruikersgemak

In het voorgaande zijn we met name ingegaan op de statistische kanten van PCALC. In zijn algemeenheid kent PCALC de volgende mogelijkheden:

- Alle rekenkundige bewerkingen: + - * / ^ en haakjes om de prioriteit vast te leggen.
- De gewone functies: wortel, kwadraat, sin, cos, tan, e-log (niet $^{10}\log$).
- De constanten e en pi.
- De operator ! (faculteit) en de operator 'boven' (voor 'n boven k') (maximaal 33!).
- Kansen en kritieke waarden uit de genoemde verdelingen.
- Verwachting en variantie van deze verdelingen.
- Editen in de tien voorafgaande commando's.
- Instellen van het aantal decimalen achter de komma tussen 2 en 11.
- Floating point alleen voor getallen groter dan 10^{15} .
- Werken met variabelen (namen mogen uit één letter bestaan); Hoofdletters en kleine letters duiden verschillende variabelen aan.

Zowel voor gewone variabelen als stochasten kunnen hoofletters als kleine letters worden gebruikt. Dit kan wel eens verwarrend werken en tot problemen leiden: stochasten kunnen niet worden gebruikt in gewone expressies en van gewone variabelen kunnen geen kansen worden uitgerekend. Het is misschien wel verstandig om daarvoor iets af te spreken, bijvoorbeeld hoofletters voor stochasten en kleine letters voor variabelen. Dit komt in ieder geval overeen met de internationale conventie.

Bij de definities van de stochastische grootheden is PCALC goed van begrip. Het maakt niet uit of men $v-p(5)$ definieert, dan wel $v-Poiss(2+3)$. De invoer van parameters mag via expressie plaatsvinden.

PCALC kan de meeste dingen die een gewone calculator kan. Het aantal functies is beperkter, maar het betreft functies die goed kunnen worden gemist ($^{10}\log$ en hyperbolische). De goniometrische functies werken alleen in radialen. Inverse goniometrische functies ontbreken. Bij gebruik van zeer grote of zeer kleine getallen zijn er enige verschillen in het voordeel van de gewone calculators. Bovendien worden zeer kleine getallen (kleiner dan $0,5 \cdot 10^{-11}$) op 0 afgerond.

PCALC heeft geen statistiek-modus; het berekenen van de variantie van een rij getallen of een frequentietabel wordt daardoor een vervelende aangelegenheid. Daar staat tegenover dat PCALC de 10 laatste regels onthoudt, waarin de mogelijkheid tot editen bestaat. Men kan dus in de regel waarmee een gemiddelde is uitgerekend in een later stadium de kwadraten toevoegen.

Hinderlijk is het feit dat getallen niet met een decimale punt mogen beginnen, dat is inmiddels zo'n gewoonte geworden! Ook op het punt van haakjes is PCALC iets pietluttiger dan andere calculators, waarbij de sluithaakjes aan het eind van een berekening kunnen worden weggelaten. Dat laatste druist weliswaar in tegen ons wiskundig purisme, maar het is wel makkelijk.

Ook SIDEKICK bevat een calculator. Onze versie daarvan is erg pover, en kan alleen eenvoudige algebra met vier cijfers vast achter de komma. Het enige punt waarop SIDEKICK het wint, is de mogelijkheid om een getal uit de calculator naar de tekstverwerker over te brengen

6. Wensen

Alhoewel PCALC een leuk programma is en in de huidige versie best bruikbaar, blijven er toch nog wel wat wensen over:

- Een handleiding waarin de algoritmen staan beschreven, met een aanduiding van de herkomst ervan, of een literatuurverwijzing daarnaar.
- Een stand waarin het PCALC-scherm blijft staan, maar waarin het programma van waaruit PCALC werd aangeroepen weer actief wordt.
- De mogelijkheid getallen te laten beginnen met een decimale punt.
- Een rekenblad dat in hoogte en plaats variabel is. Ook de kleuren moeten kunnen worden ingesteld.
- De mogelijkheid functies te definieren, b.v. $\text{var}(n,p) = n*p*(1-p)$, waarna var de waarde heeft die volgt uit de dan geldende waarden van n en p.
- Opnemen van de hypergeometrische verdeling.
- De mogelijkheid de uitkomsten naar file te schrijven.

Literatuur

- Abramowitz, M., & I.A. Stegun (1970), Handbook of Mathematical Functions, Dover Publications, New York.
- Buhrman, J.M. (1987), Toepassingen Casio fx-4000P/7000G, Statisticalc, Diemen.
- Pearson, E.S., & H.O. Hartley (1970), Biometrika Tables for Statisticians, Volume I, Cambridge University Press.