

## EXPLORING OPERATIONS RESEARCH AND STATISTICS IN THE MICRO LAB: ORSTAT

A.J. Hundepool \*)

### Samenvatting

ORSTAT is een pakket programma's op het gebied van de operationele research en de statistiek. De programma's zijn duidelijk ontwikkeld met als doelgroep het onderwijs. De programma's zijn gebruikersvriendelijk, al lijken ze voor grote professionele toepassingen toch wat beperkt. Globaal bestrijken de programma's de volgende onderdelen:

- Lineaire programmering (minimaliseren van een lineaire functie onder lineaire restricties; ook geheeltallige programmering)
- Dynamische programmering (kortste paden)
- Waarschijnlijkheidsrekening en Statistiek (wet van de grote aantallen, simulaties en centrale limiet stelling)

### 1. Technische gegevens en installatie.

Naam	ORSTAT(exploring Operations Research and STATistics in the micro lab)
Release	1990
Besturingssysteem	MS-DOS 2.10 of hoger
Minimale configuratie	512-K RAM, een floppy-disc drive, CGA, EGA of VGA kaart voor de grafische modules
Prijs	Fl 80.-
Begeleidend boek	Exploring Operations Research and Statistics in the micro lab door E. Kalvelagen en H.C. Tijms; Prentice-Hall Intern. Editions

Op zich kunnen de programma's vanaf de diskette gedraaid worden, maar veelal zal men er toch de voorkeur aan geven de programma's op een harde schijf te installeren, alvorens men ze gebruikt. Het installeren verloopt eenvoudig, omdat de programmatuur simpel met **COPY A:\*. \*** op de harde schijf gekopieerd kan worden. De omvang van de software (die op twee 360 Kb diskettes wordt geleverd) bedraagt ca. 600 Kb. Een kleine slordigheid is de aanwezigheid op de eerste diskette van een bestand AUTOEXEC.BAT, dat waarschijnlijk nog rondslinger uit een vorig release, want het commando ORSTAT is niet (meer) beschikbaar; als autoexec.bat lijkt hij niet erg bruikbaar.

---

\*) Centraal Bureau voor de Statistiek, Hoofdafdeling Automatisering, Postbus 959, 2270 AZ Voorburg, tel. 070-3375038. De in deze bespreking weergegeven opvattingen zijn die van de auteur en komen niet noodzakelijk overeen met het beleid van het Centraal Bureau voor de Statistiek.

De software is verdeeld over twee programma's, ORSTAT1 en ORSTAT2. De reden hiervoor is waarschijnlijk dat de auteurs het degenen die vanaf diskettes werken niet te lastig willen maken. In elk geval wordt dan voorkomen, dat het systeem om het wisselen van diskettes gaat vragen. Een klein schilletje om beide programma's voor diegenen die op een hard disk werken, was toch eenvoudig toe te voegen geweest. Dit zal toch de meerderheid van de gebruikers zijn. Voor beide onderdelen is er een installatie-programma (INSTALL1 en INSTALL2), waarmee men kan aangeven of men een monochrome, dan wel een kleuren-monitor gebruikt. Men moet dit dan dus 2 keer aangeven. Een slordigheidje is dat de cursor na afloop van ORSTAT soms geheel verdwenen is, dan wel een andere kleur heeft gekregen.

## 2. Het gebruik

In eerste instantie is voor het testen gebruik gemaakt van een Tulip PC-2. De programmatuur draait er wel op, maar de (Tulip eigen) gecombineerde Hercules-CGA monitor van de Tulip leverde problemen op. Het installatie-programma biedt de keuze uit een monochrome of een kleuren monitor. In beide gevallen leverde het scherm een moeilijk leesbaar geheel op. Gespikkelde letters op een gespikkelde ondergrond. Daarom is voor het verdere testen maar uitgeweken naar een (krachtiger) Olivetti M290 en een Philips 386SX in een Novell netwerk.

Bij het opstarten van ORSTAT1 krijgt men de keuze uit de volgende onderdelen:

- Linear programming
- Dynamic programming
- Law of Large Numbers
- Central Limit Theorem
- How to gamble if you must
- Kelly Betting
- Statistical Tables and Graphs

terwijl ORSTAT2 de volgende onderdelen bevat

- Integer programming
- Traveling Salesman problem
- Q-Q plot

Het hele systeem is menu-gestuurd en werkt dan ook meestal goed. Soms mis je het echter dat de <Escape> niet werkt om een menu te verlaten. Bij een keuze-menu met opties genummerd van 1 tot 8 zou je verwachten, dat het aanslaan van een getal als een hot-key de menu-optie zou activeren. Het resultaat is nu echter, dat de cursor-balk slechts verspringt naar de menu-keuze, die je dan nog moet activeren met <Enter>.

Als help-faciliteit is soms een inloop-scherm beschikbaar bij de hoofdonderdelen van het pakket. Deze schermpjes zijn soms zo vol dat het je de moed ontnemt ze te lezen. Een context-gevoelige help-faciliteit (via F1) zou soms wel erg handig zijn.

In de rest van dit verslag zullen we de verschillende onderdelen van het ORSTAT pakket nader beschouwen.

### 2.1. Linear programming

Lineaire programmering vormt een van de hoofdmoten van het ORSTAT systeem. In het begeleidende boek neemt het 70 van de 160 tekst-pagina's in beslag. Het lineaire programmeringsprobleem is het minimaliseren van een lineaire functie onder een aantal lineaire restricties. In formule vorm is dit het volgende probleem:



$$\text{Max } \sum_{j=1}^n c_j x_j \quad \left| \quad \begin{array}{l} \sum_{j=1}^n a_{ij} x_j \leq b_i \\ \text{en } x_j \geq 0 \end{array} \right. \quad \text{voor } i = 1, \dots, m, \quad j = 1, \dots, n$$

Een te minimaliseren functie is eenvoudig om te zetten in bovenstaand probleem door de  $c_j$  van teken te verwisselen. Voor de niet-negativiteits-eis van de  $x_j$  is ook een eenvoudige transformatie voorhanden. Vervang  $x$  door  $x^+ - x^-$ , waarbij  $x^+$  en  $x^- \geq 0$  zijn. Een grote klasse van problemen is in deze vorm te gieten. Voor dit probleem is een efficiënte oplossingsmethode beschikbaar, de z.g. simplex-methode.

Door de efficiëntie van de simplex-methode, zeker vergeleken met methoden voor niet-lineaire problemen, is deze methode zeer populair geworden. Vaak worden er voor niet-lineaire problemen om efficiënte-redenen vervangen door een lineaire benadering.

Een eenvoudig te gebruiken LP-programma is dan ook zeer welkom. In Orstat wordt het probleem eenvoudig gespecificeerd door een tekstje in te typen als in dit voorbeeldje:

```
max 4.8 xa + 12 xb + 6 xc + 7.2 xd
0.7 xa + 0.75 xb + 0.55 xc + 0.34 xd ≤ 400
0.55 xa + 0.82 xb + 0.8 xc + 0.55 xd ≤ 480
0.24 xa + 0.32 xb + 0.45 xc + 0.27 xd ≤ 220
1.9 xa + 2.5 xb + 1.8 xc + 2.0 xd ≤ 1500
xa ≥ 100
```

Eerst wordt de doelfunctie opgegeven en dan de restricties. Wordt het probleem groter en past een restrictie niet meer op een regel, dan moeten alle restricties afgesloten worden door een ;. Met een beetje inspanning had hier een elegantere oplossing gevonden kunnen worden. Nu heeft het toevoegen van één lange restrictie tot gevolg dat je overal ;'s moet toevoegen en niet moet vergeten bij de opties de ; als scheider aan te geven i.p.v. de nieuwe regel. De probleemdefinitie kan buiten Orstat met een eigen tekstverwerker voorbereid worden, maar kan ook tijdens een Orstat sessie met de ingebouwde eenvoudige editor ingevoerd worden.

Als de probleemdefinitie ingevoerd is, wordt eerst de syntax gecontroleerd en daarna wordt het probleem opgelost. Het resultaat is een listing met de resultaten. Hier kan men dan doorheen scrollen. Het doet een beetje ouderwets aan, maar de informatie is wel duidelijk. Het is ook mogelijk de resultaten naar een file weg te schrijven, maar dan kan het resultaat tijdens de sessie niet meer worden gezien.

Het geheel maakt echter een solide indruk en is zonder veel voorkennis goed te gebruiken.

## 2.2. Dynamic programming

Onder dit hoofdstuk wordt een systeem aangeboden om kortste paden te vinden in een netwerk. Er zijn twee vormen van kortste paden mogelijk, de zg. 'shortest path' en 'safest path'. Bij de laatste gaat het dan om een soort overlevingskansen. Het geheel wordt aan de gebruiker gepresenteerd in een rechthoekig netwerk. Men begint altijd links onder en gaat naar rechts boven. Dit is echter niet echt een beperking.

Bij dit onderdeel had ik het gevoel dat het user-interface niet echt lekker aanvoelde. Een onbedoelde menu-keuze kan niet ongedaan gemaakt worden met een <Escape>. Men zal dan het spelletje spelen tot het einde.

Om de kosten van de verschillende paden te bepalen, kan men kiezen uit random kosten, constante kosten of zelf te bepalen kosten. Het zelf invoeren van de kosten gaat nogal omslachtig, maar dit is toch de optie die je meestal wilt. Constante kosten kies je tenslotte alleen maar als een start-situatie.

Als de kosten eenmaal bepaald zijn, kan het systeem een optimale oplossing bepalen of kan de gebruiker zelf een oplossing gaan zoeken.

In dit onderdeel mis ik ernstig de mogelijkheid resultaten te bewaren. Een eenmaal met veel moeite ingevoerd probleem, kan niet even weggeschreven worden. Hierdoor is het ook niet mogelijk met meerdere situaties te spelen, want men moet steeds weer de andere situatie invoeren.

### 2.3 Law of large numbers

Onder dit item kan men spelen met een dobbelsteen met gelijke danwel ongelijke kansen. Bij de ongelijke kansen moet men dan zelf de kansen aangeven. Een beperking is dat de kansen tot 1 moeten optellen. Bij statistici is dit gebruikelijk, maar velen geven b.v. kansen op in procenten of willen een verdeling aangeven b.v. de 1 heeft een  $2 \times$  zo grote kans als de rest en zou dan de waarden 2,1,1,1,1,1 willen opgeven. Helaas moet je dan  $2/7$  en  $1/7$  in decimale breuken opgeven. Zelf even schalen was toch een kleinigheid geweest. Ook nu weer viel het gemis van de <Escape> op.

De resultaten van de dobbelsteen worden grafisch weergegeven als een vollopend staaf-diagram. Daarna wordt nog gevraagd of men statistics wil zien. Dan wordt er een  $\chi^2$  uit-gerekend. Op het scherm wordt daar geen nadere toelichting bij gegeven, hoewel dat toch niet zou misstaan.

Derhalve maar eens het boek opgeslagen. Na enige uitleg over de  $\chi^2$  staat dan onderaan bladzijde 103 de volgende zin: 'It will be clear that in this particular example you should be suspicious when the  $\chi^2$  statistic has an exceptionally small value'. Het voorbeeldje is het resultaat van  $1200 \times$  gooien met de verdeling 195, 202, 199, 196, 205, 203. Dit lijkt op een misser.

Mijn ervaringen met deze dobbelsteen leverde  $\chi^2$ -en op van 3.64, 4.84, 4.09. Ik krijg dan meer het gevoel, dat de random-generator niet aan de te stellen eisen voldoet.

### 2.4 Central Limit Theorem

Hier wordt de bekende centrale limiet stelling gevisualiseerd aan de hand van de verdeling van de som van het aantal ogen van een dobbelsteen. Op het scherm verschijnt als resultaat de bekende klok-kromme. Net als bij het vorige onderwerp kan hier gekozen worden uit een dobbelsteen met gelijke kansen en zelf te specificeren kansen. De basis van de berekeningen is niet een simulatie, maar een analytische berekening van de verdeling. Verder kunnen de resultaten nog in een Q-Q-plot bekeken worden. De Q-Q-plot komt als laatste onderwerp nog aan bod.



## 2.5 How to gamble if you must

Hier komen we terecht in het VU-casino. Het spel 'red and black' is een gokspel waarbij men met een bepaalde kans kan winnen en dan de inzet dubbel uitbetaald krijgt. De gebruiker heeft een start-kapitaal en moet een doel-kapitaal zien te bereiken. Als parameters worden opgegeven het aantal keren dat gespeeld wordt en de winstkans. Het is de bedoeling, dat de gebruiker per spel een bedrag inzet. Met de opgegeven kans wordt dan de inzet gewonnen dan wel verloren. In een eenvoudig grafiekje is de ontwikkeling van het spel te volgen.

## 2.6. Kelly Betting

J.F. Kelly heeft zich in 1956 verdiept in de optimale strategie bij dit soort gokspelletjes. Helaas gaat dit systeem slechts op als de winst-verwachting groter is dan de verlies-verwachting. Het komt erop neer dat, als de winstkans  $p$  is, en de verlieskans  $q=1-p$ , dat de optimale strategie is een fractie  $p-q$  van het vermogen in te zetten. Er zijn echter weinig casino's te vinden die de toepassing van Kelly's strategie mogelijk maken.

Na de specificatie van het start-kapitaal en de winst-kans moet de gebruiker een eigen inzet-fractie opgeven. Het programma vergelijkt dan die kans met de strategie van Kelly en laat dit in een eenvoudig grafiekje zien. Soms ontspoot de grafiek echter en verschijnen er twee Kelly strategieën, toen ik de simpele situatie opgaf om met de kans 0.5 de inzet dubbel terug te krijgen.

## 2.7 Statistical Tables and Graphs

Hier kan de gebruiker van een aantal bekende verdelingsfuncties grafieken te zien krijgen en percentielen en overschrijdingskansen laten uitrekenen. Dit onderdeel vereist wel dat men bekend is met de eigenschappen van de verschillende verdelingen. Er is hier geen on-line help aanwezig en in het boek wordt het in een paar bladzijden afgedaan. De volgende verdelingen zijn beschikbaar .

- Normaal
- Chi-kwadraat
- Student's T
- Binomiaal
- Poisson
- Hypergeometrisch
- Fisher's F

## 2.8.Integer programming

De probleemstelling bij integer programming is vrijwel dezelfde als bij lineaire programmering, met dien verstande dat op de variabelen extra restricties gelegd kunnen worden. Variabelen kunnen geheel (integer) en binair (0/1) zijn. Dit wordt eenvoudig opgegeven door de regel '\$binary varx1' of '\$integer varx1' aan de probleem-definitie toe te voegen. Voor de oplossing van het probleem wordt nu een 'branch and bound' methode gebruikt om het optimum te vinden. Verder wekt het net zo als lineaire programmering.

## 2.9. Travelling salesman problem

Het bekende handelsreizigers-probleem wordt hier op een aardige manier visueel gepresenteerd. Op een twee-dimensionaal tableau worden de te bezoeken lokaties interactief aangegeven met de cursor-toetsen. Als test kan ook een random patroon geselecteerd worden. Het programma biedt dan verschillende oplossings-strategieën:

- random tour
- user tour
- nearest neighbour
- nearest insertion
- farthest insertion
- greatest angle
- branch and bound (exact)

Tijdens het afwerken van een bepaalde strategie wordt dan in beeld gebracht hoe de oplossing tot stand komt. Opmerkelijk is dat de 'farthest insertion' het nog zo slecht niet doet. Hierbij wordt steeds de verste lokatie opgenomen in het reeds bestaande pad en op de meest ideale positie tussengevoegd. Ik vond dit een leuk onderdeel van het pakket.

Een paar kanttekeningen zijn, dat een eenmaal ingevoerde verdeling van lokaties niet bewaard kan worden. Morgen nog eens spelen met eenzelfde probleem verplicht je tot het opnieuw invoeren ervan. Ook miste ik sterk de mogelijkheid (b.v. via <Escape> onderweg een oplossing te verlaten. Bij een niet langer gewenste strategie met een fors aantal steden is dit hinderlijk. Alleen <Ctrl Break> werkt dan nog.

## 2.10. Q-Q plot

De Q-Q plot biedt de mogelijkheid de waarschijnlijkheidsverdeling van een te onderzoeken gegevens verzameling visueel te vergelijken met bepaalde verdelingen. Als de verdeling van de te onderzoeken gegevensverzameling overeenkomt met een van de 6 in het programma bekende verdelingen resulteert dit in een plot van de gegevens volgens een rechte lijn. Dit is een eenvoudige visueel inzichtelijke methode om een indruk te krijgen van de verdeling in de te onderzoeken gegevens.

Het Q-Q plot onderdeel biedt de mogelijkheid de gegevensverzameling te simuleren, als een mogelijkheid inzicht in de methode te krijgen. De te onderzoeken verdelingen zijn:

- Uniform
- Normaal
- Exponentieel
- Lognormaal
- WeiBull

Helaas is het alleen mogelijk de gegevens interactief in te voeren. Bewaren tot een volgende keer, dan wel inlezen van een set gegevens is er niet bij. Dit beperkt de gebruiker toch wel erg sterk. Ook heb ik het als irritant ervaren dat niet overal de Escape werkt om terug te keren naar een hoger menu. Een ondoordachte keus leidt er nu toe dat je maar wat moet knoeien om terug te komen. Jammer dat dit overigens leuke onderdeel ontsierd wordt door deze beperkingen.

### 3. Conclusie

De serie programma's van ORSTAT zijn duidelijk bedoeld voor de lessituatie. De onderdelen over lineaire en integer programmering spraken mij het meeste aan. Zeker voor het oplossen van relatief bescheiden problemen kan iemand met weinig ervaring en een klein beetje kennis van de lineaire programmering hier vlot mee uit de voeten. Ook het handelsreizigers probleem wordt visueel aardig in beeld gebracht. De andere onderdelen beschouw ik meer als leuk meegenomen, dan als een reden ORSTAT aan te schaffen.

De algemene kanttekeningen bij dit pakket zijn dat er niet overal (graag via F1) een help-faciliteit aanwezig is en dat de Escape soms node wordt gemist om een onbedoelde keuze ongedaan te maken.



