NOTA 1048 april 1978

HET BENADEREN VAN DE NORMALE VERDELING

J.R. Maassen

I N H O U D

INLEIDING

Het berekenen van kansen van een normale verdeling kan niet rechtstreeks. De oorzaak hiervan is het feit dat de integraal die de verdeling beschrijft niet oplosbaar is. De numerieke waarde van kansen van een normale verdeling is te verkrijgen uit numerieke integratie of uit benaderingsformules. Hier werd gewerkt met benaderingsfuncties zoals die ontwikkeld zijn door HASTINGS (1955).

BENADERING VOLGENS HASTINGS

Hastings geeft een benaderingsformule voor de in de fysica bekend zijnde zogenaamde error-functie welke met de normale verdeling verwant is.

$$\Phi\ (x) = \frac{2}{\sqrt{\pi}} \int_{o}^{x} e^{-t^2}\ dt \qquad (1)$$

waarbij x = o of positief reëel en eindig

e = basis van het natuurlijk logarithmenstelsel

De benaderingsfunctie is als volgt opgebouwd:

Indien

$$\eta = \frac{1}{1 + pX} \qquad (2)$$

waarin p = constante (zie fig. 1, 2 en 3)

$$\Phi(X) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$$

$$0 \leq X < \infty$$

$$\eta = \frac{1}{1 + pX}$$

$$\Phi^*(X) = 1 - (a_1\eta + a_2\eta^2 + a_3\eta^3)\ \Phi'(X)$$

$p = .47047$ $\quad a_1 = .3084284$

$a_2 = -.0849713$

$a_3 = .6627698$

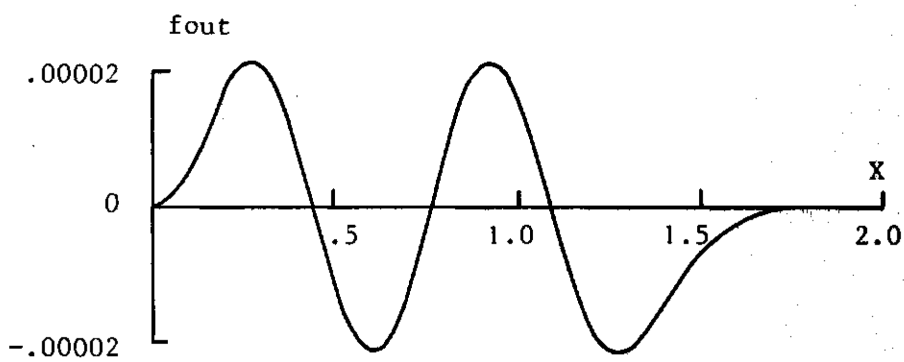

Fig. 1. Benadering van de error-functie met een 3e-graads functie van η (volgens HASTINGS)

$$\Phi(X) = \frac{2}{\sqrt{\pi}} \int_0^X e^{-t^2} dt$$

$$0 \leq X < \infty$$

$$\eta = \frac{1}{1 + pX}$$

$$\Phi^*(X) = 1 - (a_1\eta + a_2\eta^2 + a_3\eta^3 + a_4\eta^4)\Phi'(X)$$

$p = .381965$

$a_1 = .12771538$
$a_2 = .54107939$
$a_3 = -.53859539$
$a_4 = .75602755$

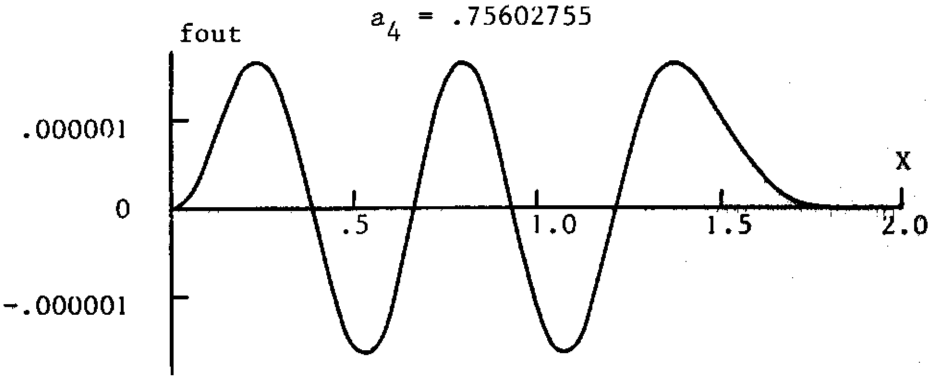

Fig. 2. Benadering van de error-functie met een 4e-graads functie van η (volgens HASTINGS)

$$\Phi(X) = \frac{2}{\sqrt{\pi}} \int_0^X e^{-t^2} dt$$

$$0 \leq X < \infty$$

$$\eta = \frac{1}{1 + pX}$$

$$\Phi^*(X) = 1 - (a_1\eta + a_2\eta^2 + a_3\eta^3 + a_4\eta^4 + a_5\eta^5)\ \Phi'(X)$$

$p = .3275911$ $\quad a_1 = .225836846$

$a_2 = -.252128668$

$a_3 = 1.259695130$

$a_4 = -1.287822453$

$a_5 = .940646070$

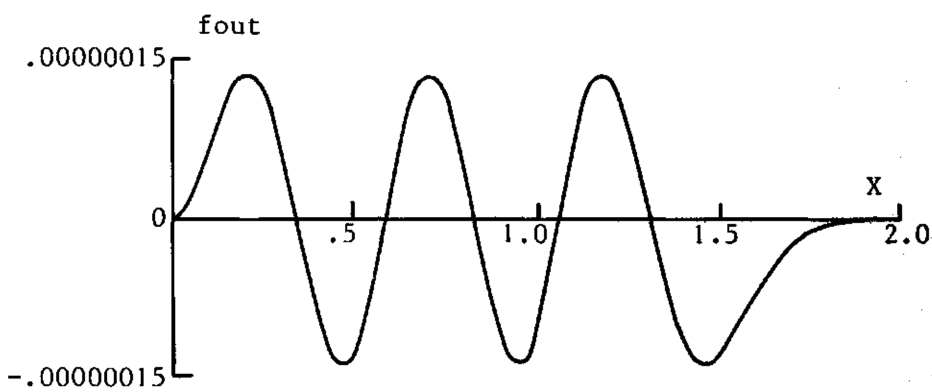

Fig. 3. Benadering van de error-functie met een 5e-graads **functie** van η (volgens HASTINGS)

dan wordt als benadering voor de error-functie gegeven

$$\Phi^*(x) = 1 - c(\eta).\Phi'(x) \qquad (3)$$

waarin $\Phi'(x)$ de afgeleide is van $\Phi(x)$ naar x

en c = een functie van η

In (3) is dan $\Phi'(x) = \frac{2}{\sqrt{\pi}} e^{-x^2}$

Afhankelijk van de gewenste nauwkeurigheid, wordt door Hastings een 3e-, 4e-of 5e-graads functie van η gegeven. In fig. 1, 2 en 3 is deze functie nader uitgewerkt. Tevens is bij elke figuur de fout weergegeven die bij de betreffende benadering wordt gemaakt. Opvallend is dat bij waarden van x die groter zijn dan 1,50 de gemaakte fout zeer snel naar 0 loopt.

HET VERBAND MET DE NORMALE VERDELING

De error-functie Φ houdt als volgt verband met de cumulatieve normale verdeling F:

$$\Phi(x) = 2\,F(x\sqrt{2}) - 1 \qquad (4)$$

waarin

F (...) = verdelingsfunctie van de normale verdeling met verwachting 0 en variantie 1

dus

$$F(x\sqrt{2}) = \tfrac{1}{2}\{1 + \Phi(x)\} \qquad (5)$$

Met de benaderingen voor Φ door Φ^* geeft dit

$$F^*(x\sqrt{2}) = \tfrac{1}{2}\{1 + \Phi^*(x)\}$$

en, in verband met (3),

$$= \tfrac{1}{2}\{2 - c(\eta).\Phi'(x)\}$$

Stel nu dat $\underline{u}$ normaal verdeeld is met parameters μ en σ, dan is de transformatie naar de standaard-normale verdeling de functie F(z) waarin

$$z = \frac{u-\mu}{\sigma} \qquad (6)$$

Uit (4) volgt $\qquad x = \frac{z}{\sqrt{2}}$

zodat $\qquad x = \frac{u-\mu}{\sqrt{2}.\sigma}$

Verder is gegeven dat $\qquad \eta = \frac{1}{1+pX}$

dus de benadering wordt

$$F^*(\frac{u-\mu}{\sigma}) = \tfrac{1}{2}\{2-c(\eta).\Phi'(\frac{u-\mu}{\sigma\sqrt{2}})\}$$

DE INVERSE FUNCTIE

Uitgaande van de normale verdeling is het mogelijk om van een overschrijdingskans de bijbehorende drempelwaarde x(q) terug te berekenen.

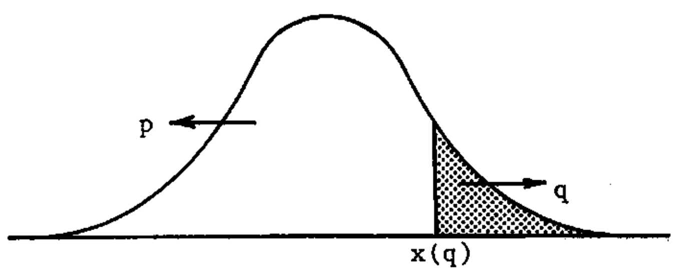

Fig. 4. p = kans op een waarde $\leqslant$ x(q)
q = $\qquad$ > x(q)

Als we de kans dat een variabele een waarde kleiner of gelijk x aanneemt gelijk stellen aan p, dan is de kans dat dit niet gebeurt gelijk aan 1-p, hetgeen we q noemen (gearceerd in fig. 4)

Nu geldt

$$q = \frac{1}{\sqrt{2\pi}} \int_{x(q)}^{\infty} e^{-\frac{1}{2}t^2} \tag{7}$$

Door HASTINGS wordt gesteld dat als

$$\eta = \sqrt{\ln(\frac{1}{q^2})} \tag{8}$$

een benadering voor x(q) gegeven kan worden door

$$X^*(q) = \eta - d(\eta) \tag{9}$$

Hierin is d een rationele 3e graads functie van η.
In fig. 5 is de term d gegeven en is de fout weergegeven die bij de benadering wordt gemaakt.

TOEPASBAARHEID

Gebleken is dat de toepasbaarheid van de functies van HASTINGS groot is. Bovendien zijn de functies gemakkelijk en goedkoop te programmeren. Dit was aanleiding ten behoeve van een neerslag-onderzoek het programma ERROR te ontwikkelen. Voor het bepalen van kansen in een normale verdeling bleek de formule van Hastings een zeer goede benadering te zijn. Zelfs zo nauwkeurig dat het niet mogelijk bleek op de conventionele manier, dit is via numerieke integratie volgens de symplex-methode, een even nauwkeurig resultaat te verkrijgen, binnen dezelfde grenzen, zowel wat betreft computertijd als computerkosten. Bij stapgrootte van 10^{-5} bleek de CYBER-computer van IWIS-TNO al meer dan 1 uur rekentijd nodig te hebben om van een relatief klein interval de cumulatieve kans te berekenen. Zelfs dan kwam er nog een vrij grote onnauwkeurigheid in de resultaten voor. De benadering van Hastings is daarentegen altijd tot minstens 6 decimalen nauwkeurig, wordt direkt berekend en gaat binnen een fractie van de tijd benodigd voor het integreren.

$$q = \frac{1}{\sqrt{2\pi}} \int_{x(q)}^{\infty} e^{-\frac{1}{2}t^2} dt$$

$$0 < q \leq .5$$

$$\eta = \sqrt{\ln \frac{1}{q^2}}$$

$$X^*(q) = \eta - \frac{a_0 + a_1\eta + a_2\eta^2}{1 + b_1\eta + b_2\eta^2 + b_3\eta^3}$$

$a_0 = 2.515517$ $\quad b_1 = 1.432788$
$a_1 = .802853$ $\quad b_2 = .189269$
$a_2 = .010328$ $\quad b_3 = .001308$

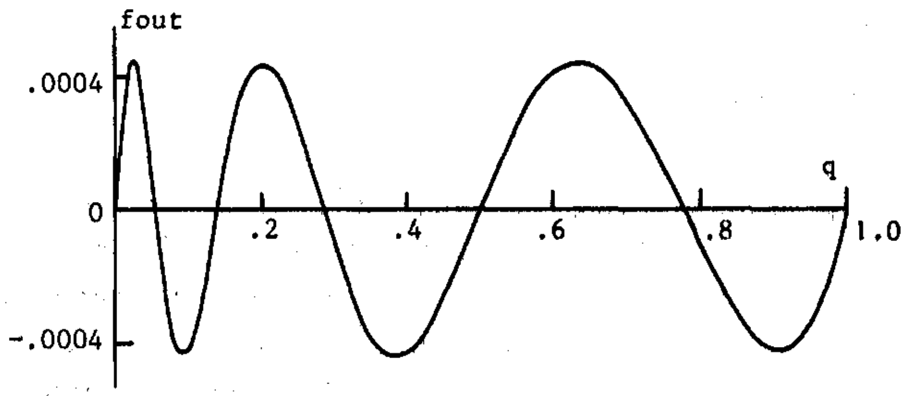

Fig. 5. Benadering van de drempelwaarde met een 3e-graads functie van η (volgens HASTINGS)

PROGRAMMA ERROR

Het programma ERROR is zodanig ontwikkeld dat men er kansen en drempelwaarden mee kan berekenen voor verdelingen die al dan niet afgeknot zijn. Voorwaarde is dat de verdelingen wel normaal zijn. Tevens is het mogelijk de verwachtingswaarde van afgeknotte verdelingen te berekenen. Voor het berekenen van kansen is de benadering uit fig. 3 gebruikt, omdat deze het nauwkeurigst is. Het berekenen van de verwachtingswaarde onder afknotting, zal aan de hand van fig. 6 duidelijk worden.

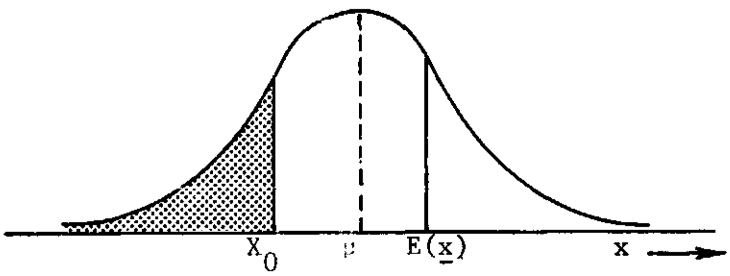

Fig. 6. Voorbeeld van een afgeknotte verdeling

Stel we hebben een normale verdeling met verwachtingswaarde μ. Als nu bekend is dat waarden die kleiner zijn dan x_0 niet kunnen voorkomen, dan kunnen we de verdeling beschouwen als zijnde afgeknot in het punt x_0. Het is dan ook duidelijk, dat als het gearceerde gedeelte in fig. 6 niet 'meedoet', het gemiddelde, dus de verwachtingswaarde $E(\underline{x})$, naar rechts opschuift. Deze nieuwe waarde $E(\underline{x})$ wordt in programma ERROR ook berekend.

Afhankelijk van wat men in de eerste gegeven-kaart opgeeft, wordt een aantal berekeningen uitgevoerd. In deze eerste gegeven-kaart worden globale gegevens aan de computer verstrekt, die in volgende kaarten gespecificeerd moeten worden. Wordt b.v. in de eerste gegeven-kaart opgegeven dat de verdeling niet standaard-normaal is, dan moeten in de tweede gegeven-kaart de parameters van de verdeling worden gespecificeerd. Wordt opgegeven dat de verdeling is afgeknot, dan moet in een volgende kaart opgegeven worden in welk punt de verdeling is afgeknot. Evenzo moet in de eerste gegeven-kaart opgegeven worden welke bewerking men uitgevoerd wil hebben: het berekenen van onderschrijdingskansen, het berekenen van de verwachtingswaarde onder afknotting, of het uit een op te geven overschrijdingskans terugberekenen van de drempelwaarde.

Wil men een combinatie van bewerkingen uit laten voeren, dan is dit ook mogelijk. Aan de hand van voorbeelden in de bijlagen wordt dit verduidelijkt.

LITERATUUR

HASTINGS, C., 1955. Approximations for digital computers, Princeton University Press, Princeton, New Yersey, pp. 201

Ter verduidelijking van de positie van de gegevens op ponskaart staan de kaartkolomnummers vermeld op de input-voorbeelden.

voorbeeld nr. 1 van in/output programma ERROR

In de programma-besturingskaart wordt opgegeven dat alle bewerkingen gewenst zijn.

In de 2e kaart worden de parameters van de verdeling opgegeven

In de 3e kaart wordt opgegeven waar de verdeling is afgeknot

In de 4e kaart wordt een aantal waarden opgegeven waarvoor de onderschrijdingskans moet worden berekend

In de 5e kaart wordt de overschrijdingskans opgegeven, waarvoor de bijbehorende drempelwaarde moet worden berekend.

```
DE WERKELIJKE INPUT STAAT IN HET OMLIJNDE BLOK

KAARTKOLOM
1234567890        20        30        40        50        60        70        80
         .         .         .         .         .         .         .         .
--------------------------------------------------------------------------------
AAAAA
    0    1
    0
 -3        -2        -1         0         1         2         3
%EOF
0.5
%EOF
--------------------------------------------------------------------------------
```

PARAMETERS VAN DE VERDELING ZIJN: MU = 0.00 EN SIGMA = 1.0000

KANSTABEL VAN N(0.00, 1.0000) VOOR DE VOLGENDE OPGEGEVEN DREMPELWAARDEN

DE VERDELING IS AFGEKNOT BIJ X= 0.000

X =	-3.000	-2.000	-1.000	0.000	1.000	2.000	3.000
PHI =	0.00000000	0.00000000	0.00000000	0.00000000	.68268947	.95449987	.99730007

DE VERWACHTINGSWAARDE ONDER AFKNOTTING IN HET PUNT 0.000 BEDRAAGT .797885

BIJ DE OVERSCHRIJDINGSKANS .50000000 VAN N(0.00, 1.0000) HOORT DE DREMPELWAARDE .674
EOI ENCOUNTERED AFTER COPY OF FILE
0, RECORD 1

voorbeeld nr. 2 van in/output programma ERROR

In de programma-besturingskaart wordt opgegeven dat de verdeling standaard-normaal en niet afgeknot is. Gevraagd wordt kansen en de drempelwaarde, behorend bij een op te geven overschrijdingskans, te berekenen.

In de 2e kaart wordt een aantal waarden opgegeven waarvoor de onderschrijdingskans moet worden berekend

In de 3e kaart wordt de overschrijdingskans opgegeven, waarvoor de bijbehorende drempelwaarde moet worden berekend.

```
DE WERKELIJKE INPUT STAAT IN HET OMLIJNDE BLOK

KAARTKOLOM
1234567890          20         30         40         50         60         70         80
---------.---------.---------.---------.---------.---------.---------.---------.
**A*A
 -3         -2         -1          0          1          2          3
%EOF
0.5
%EOF
--------------------------------------------------------------------------------
```

```
PARAMETERS VAN DE VERDELING ZIJN:  MU =  0.00   EN SIGMA = 1.0000


KANSTABEL VAN N(  0.00, 1.0000) VOOR DE VOLGENDE OPGEGEVEN DREMPELWAARDEN

DE VERDELING IS NIET AFGEKNOT


X   =      -3.000      -2.000      -1.000       0.000       1.000       2.000       3.000
PHI =   .00134997   .02275006   .15865526   .50000000   .84134474   .97724994   .99865003

ARGUMENT TOO SMALL
ERROR NUMBER   115  DETECTED BY EXP


BIJ DE OVERSCHRIJDINGSKANS .50000000 VAN N(  0.00, 1.0000) HOORT DE DREMPELWAARDE   -.000
 EOI ENCOUNTERED AFTER COPY OF FILE
          0, RECORD  1
```

voorbeeld nr. 3 van in/output programma ERROR

In de programma-besturingskaart wordt opgegeven dat van de standaard-normale verdeling de verwachtingswaarde onder afknotting moet worden berekend.

In de 2e kaart wordt opgegeven waar de verdeling is afgeknot.

```
DE WERKELIJKE INPUT STAAT IN HET OMLIJNDE BLOK

KAARTKOLOM
1234567890         20        30        40        50        60        70        80
         .         .         .         .         .         .         .         .
--------------------------------------------------------------------------------
***A*
  -1.
%EOF
--------------------------------------------------------------------------------
```

```
PARAMETERS VAN DE VERDELING ZIJN:  MU =  0.00   EN SIGMA = 1.0000


ARGUMENT TOO SMALL
ERROR NUMBER    115  DETECTED BY EXP

ARGUMENT TOO SMALL
ERROR NUMBER    115  DETECTED BY EXP


DE VERWACHTINGSWAARDE ONDER AFKNOTTING IN HET PUNT -500.000 BEDRAAGT   0.000000
 EOI ENCOUNTERED AFTER COPY OF FILE
          0, RECORD  1
```

voorbeeld nr. 4 van in/output programma ERROR

In de programma-besturingskaart wordt opgegeven dat alle bewerkingen gewenst zijn

In de 2e kaart worden de parameters van de verdeling opgegeven

In de 3e kaart wordt opgegeven waar de verdeling is afgeknot

In de 4e kaart wordt een aantal waarden opgegeven waarvoor de onderschrijdingskans moet worden berekend

In de 5e kaart wordt de overschrijdingskans opgegeven, waarvoor de bijbehorende drempelwaarde moet worden berekend.

bijlage 4 vervolg

```
DE WERKELIJKE INPUT STAAT IN HET OMLIJNDE BLOK

KAARTKOLOM
1234567890        20        30        40        50        60        70        80
         .         .         .         .         .         .         .         .
--------------------------------------------------------------------------------
AAAAA
    2    5
  -3.
 -3        -2        -1         0         1         2         3
%EOF
0.5
%EOF
--------------------------------------------------------------------------------
```

```
PARAMETERS VAN DE VERDELING ZIJN:  MU =  2.00   EN SIGMA = 5.0000


KANSTABEL VAN N(  2.00, 5.0000) VOOR DE VOLGENDE OPGEGEVEN DREMPELWAARDEN

DE VERDELING IS AFGEKNOT BIJ X=          -3.000


X   =      -3.000      -2.000      -1.000       0.000       1.000       2.000       3.000
PHI =  0.00000000   .06323220   .13739648   .22098319   .31150733   .40571328   .49991924


DE VERWACHTINGSWAARDE ONDER AFKNOTTING IN HET PUNT   -3.000 BEDRAAGT   3.438000


BIJ DE OVERSCHRIJDINGSKANS .50000000 VAN N(  2.00, 5.0000) HOORT DE DREMPELWAARDE    2.999
 EOI ENCOUNTERED AFTER COPY OF FILE
         0, RECORD  1
```