Selectuur

Meest aannemelijke verdelingen

UDC 519.1.20

I. Inleiding

In het kader van zijn studies over informatietheorie op het gebied der telecommunicatie heeft S h a n n o n [1] enige jaren geleden de exponentiële en de normale verdeling als de *meest aannemelijke* gevonden: door de entropie te maximaliseren onder verschillende bijvoorwaarden. Van deze methode volgt hieronder een overzicht. Er zij op gewezen dat de gebruikte term „meest aannemelijke verdeling" geen verband heeft met de „aannemelijkste schatting" van parameters van een verdeling door middel van de zogenaamde maximum-likelihood-methode.

II. Methode

Zijn $p_1, \ldots, p_n$ de kansen in een discrete waarschijnlijkheidsverdeling ($p_i \geqq 0$, $\Sigma p_i = 1$), dan heet

$$H = -\Sigma p_i \log p_i$$

de entropie van de verdeling. Bij een continue verdeling met waarschijnlijkheidsdichtheid $p(x)$ ($p(x) \geqq 0$, $\int p(x)\,dx = 1$), is de entropie gedefinieerd als

$$H = -\int p(x) \log p(x)\,dx.$$

Volgens S h a n n o n is de entropie een maat voor de onbepaaldheid van de verdeling, d.w.z. voor het ontbreken van relatief grote waarden voor kans of kansdichtheid in de verdeling (zie de inleiding van [2]). Wil men de meest aannemelijke keus doen uit alle verdelingen die aan een of meer voorwaarden voldoen (b.v. gegeven gemiddelde of spreiding), maar waarvan men verder niets weet, dan zal men die met de grootste onbepaaldheid zoeken, dus die met de grootste entropie.

Het verband tussen entropie en aannemelijkheid van een verdeling kan in deze selectuur slechts gesuggereerd worden. Eerst door een analogie met het entropiebegrip uit de physica. Hier is de entropie een maat voor de waarschijnlijkheid van een toestand; volgens de tweede hoofdwet der thermodynamica streeft een afgesloten systeem naar de toestand met grootste entropie, die tevens de grootste waarschijnlijkheid heeft. Men zie de voordracht van Prof. v a n S o e s t op de Statistische Dag 1960 [3].

Nu is entropie en aannemelijkheid van een waarschijnlijkheidsverdeling iets anders dan entropie en waarschijnlijkheid van een physische toestand.

Dat er niettemin analogie kan zijn tot in de formules, leert een voorbeeld uit de kinetische gastheorie ([4], [5] of [6]), hier vereenvoudigd weergegeven.

Men heeft een verzameling van N objecten (moleculen b.v.), verdeeld over r hokjes. Aangenomen wordt dat de kans voor een object om in hokje i te zitten, $1/r$ is voor alle $i = 1, \ldots, r$. De waarschijnlijkheid van een toestand, bepaald door n_i objecten in hokje i ($i = 1, \ldots, r$; $\Sigma n_i = N$), is

$$W = \frac{N!}{n_1! n_2! \ldots n_r!} \cdot r^{-N}.$$

Voor de waarschijnlijkste toestand is dus W, of $\log W$, maximaal. Noemen we $f_i = n_i/N$ de fractie der objecten in hokje i en vervangen we voor grote waarden van n volgens Stirling $\log n!$ door $-n + (n + \frac{1}{2}) \log n + \frac{1}{2} \log 2\pi$, en dit door $-n + n \log n$, dan krijgen we

$$\log W \approx -N \Sigma f_i \log f_i.$$

Het gaat er dus om $-\Sigma f_i \log f_i$ maximaal te maken. Men vergelijke hiermee de uitdrukking $-\Sigma p_i \log p_i$ voor de entropie bij Shannon.

Eigenschappen van Shannons entropiebegrip vindt men b.v. behandeld in een betrekkelijk eenvoudig boekje van Yaglom en Yaglom [7] en in [2]. De zin der entropie kan als volgt plausibel worden gemaakt.

Willen we de onbepaaldheid van het resultaat van een experiment met n even waarschijnlijke uitkomsten door een functie $f(n)$ uitdrukken, dan moet $f(1) = 0$ zijn (dit betekent dat een *zekere* uitkomst onbepaaldheid 0 heeft). Als verder A en B twee onafhankelijke experimenten zijn met m resp. n even waarschijnlijke uitkomsten, dan heeft het experiment AB (combinatie van A met B) mn even waarschijnlijke uitkomsten; we willen nu dat de onbepaaldheid van AB gelijk is aan de som van de onbepaaldheden van A en B:

$$f(mn) = f(m) + f(n).$$

Hieraan en aan $f(1) = 0$ voldoet $f(n) = \log n$. Kiezen we dus als onbepaaldheid $\log n$, dan is $\log n = n\,(-1/n \log 1/n) = -\sum_{i=1}^{n} 1/n \log 1/n$; hierin is $1/n$ de kans van elk der uitkomsten van het experiment. Algemener: hebben de n uitkomsten ongelijke kansen $p_1, \ldots, p_n$, dan nemen we als onbepaaldheid $-\sum_{i=1}^{n} p_i \log p_i$. Dit is de entropie volgens Shannon. Men kan aantonen, dat de grootste onbepaaldheid verkregen wordt als alle $p_i = 1/n$ zijn, hetgeen ook intuïtief te verwachten was.

Voor twee onafhankelijke experimenten A en B geldt weer: onbepaaldheid van AB = onbepaaldheid van A + onbepaaldheid van B, want

$$-\Sigma\Sigma p_i q_j \log (p_i q_j) = -\Sigma p_i \log p_i - \Sigma q_j \log q_j.$$

Van een experiment met één (zekere) uitkomst is de onbepaaldheid

$$-1 \cdot \log 1 = 0.$$

Men kan vragen, of ook een andere definitie der entropie iets bruikbaars zou opleveren. In [2] wordt echter bewezen, dat als men uitgaat van bepaalde aan de entropie te stellen eisen, alleen de functie $-c\,\Sigma p_i \log p_i$ (c een constante) hieraan voldoet.

III. Enige eenvoudige verdelingen

In deze paragraaf worden enige eenvoudige verdelingen, bekend uit de statistiek, teruggevonden door (zoals reeds door Shannon gedaan is) de entropie maximaal te maken bij gegeven nevenvoorwaarden. Hierbij wordt een methode uit de variatierekening toegepast; men vindt deze b.v. in [8]. Het probleem is om in $\int f(x, y)\,dx$, waarin f een bekende functie van x en y, maar y een nog onbepaalde functie van x is, $y = y(x)$ zó te kiezen dat de integraal maximaal wordt; eventueel onder de nevenvoorwaarden

$$\int g_i(x, y)\,dx = a_i \quad (i = 1, \ldots, m),$$

waarbij steeds dezelfde $y = y(x)$ en hetzelfde integratie-interval gebruikt worden.

In ons geval is $f(x, y) = -y \log y$; er is minstens één bijvoorwaarde, nl. $\int y dx = 1$, want $y = p(x)$ is een kansdichtheid.

A. Eén bijvoorwaarde: kanssom = 1

Derhalve is het probleem: bepaal het maximum van

$$-\int p(x) \log p(x)\,dx, \text{ onder de bijvoorwaarde}$$
$$\int p(x)\,dx = 1$$

De multiplicatormethode van Lagrange geeft als noodzakelijke voorwaarde voor de oplossing van dit probleem, dat de partiële afgeleide naar $p(x)$ van de integrand van $\int \{-p(x) \log p(x) + \lambda p(x)\}\,dx$ gelijk 0 is. Dus:

$$-\log p(x) - 1 + \lambda = 0 \text{ of } p(x) = e^{\lambda - 1}.$$

De Lagrange-multiplicator λ wordt vervolgens bepaald uit de bijvoorwaarde $\int p(x)\,dx = 1$. Dit kan hier slechts, als tussen eindige grenzen wordt geïntegreerd, zodat *begrensdheid* van het interval een noodzakelijke voorwaarde is.

Het resultaat is de homogene (of rechthoekige) verdeling. Deze treedt dus

op als van een kansverdeling (uiteraard met kanssom = 1) niets anders is gegeven dan de grenzen van het interval; zulks ligt ook geheel in de lijn der verwachtingen.

B. Twee bijvoorwaarden: kanssom = 1 en gemiddelde = m

Nu moet het maximum bepaald worden van

$$-\int p(x) \log p(x)\, dx, \text{ onder de bijvoorwaarden}$$
$$\int p(x)\, dx = 1 \text{ en } \int xp(x)\, dx = m$$

De multiplicatormethode leidt nu tot de vergelijking:

$$-\log p(x) - 1 + \lambda + \mu x = 0 \text{ of } p(x) = e^{\lambda-1} \cdot e^{\mu x}.$$

De beide Lagrange-multiplicatoren volgen weer uit de bijvoorwaarden. De berekening gaat het gemakkelijkst als één der grenzen van het interval ∞ of $-\infty$ is en de andere eindig. Het resultaat is de exponentiële verdeling. Op het interval $(-\infty, \infty)$ bestaat geen oplossing.

Bij een begrensd interval kan zonder wezenlijke beperking $0 \leqq x \leqq 1$ genomen worden. Men vindt dan dat μ o.a. aan een transcendente vergelijking: $\mu = (1 + m\mu)(1 - e^{-\mu})$ moet voldoen. Een nadere beschouwing leert: $\mu = 0$ dan en slechts dan als $m = \frac{1}{2}$; dit geeft weer de homogene verdeling uit *A*. Voor $m \neq \frac{1}{2}$ is er ook precies één oplossing $\mu \neq 0$, die voldoet; als $0 < m < \frac{1}{2}$, dan $\mu < 0$; als $\frac{1}{2} < m < 1$, dan $\mu > 0$.

C. Twee bijvoorwaarden: kanssom = 1 en 2e moment t.o.v. c = σ^2

Nu luidt het probleem: maximaliseer $-\int p(x) \log p(x)\, dx$, onder de bijvoorwaarden $\int p(x)\, dx = 1$ en $\int (x-c)^2 p(x)\, dx = \sigma^2$.

Er moet nu voldaan zijn aan:

$$-\log p(x) - 1 + \lambda + \mu (x-c)^2 = 0 \text{ of } p(x) = e^{\lambda-1+\mu(x-c)^2}.$$

De Lagrange-multiplicatoren volgen uit de bijvoorwaarden. Voor een aan beide zijden onbegrensd interval vindt men de normale verdeling. Het gemiddelde blijkt gelijk aan *c* te zijn, m.a.w. de meest aannemelijke verdeling groepeert de waarnemingen om *c* als gemiddelde. Dit gemiddelde $m = c$ is echter *niet* a priori gegeven. Men kan aantonen, dat bij toevoeging van een derde bijvoorwaarde: gemiddelde $m = c$, de Lagrange-multiplicator die hierop betrekking heeft, nul moet zijn, m.a.w. de bijvoorwaarde is al vervuld.

De normale verdeling treedt dus op bij een aan beide zijden onbegrensd interval onder a priori gegeven variantie. Heeft het interval één of twee eindige

grenzen, dan treedt een verdeling met bovengenoemde $p(x)$ op, waarbij de parameter μ echter moet voldoen aan een transcendente vergelijking waarop we niet ingaan. Tenzij de ene grens $\pm \infty$ en de andere juist c is; dit geeft, afgezien van een factor 2, een helft van de normale verdeling.

Een bijzonder geval van C is

C'. Twee bijvoorwaarden: kanssom $= 1$ *en 2e moment t.o.v.* $0 = \sigma^2$.

Hier is nl. $c = 0$. Bij een aan weerszijden onbegrensd interval is de normale verdeling met gemiddelde 0 het meest aannemelijk. Beschouwt men alleen positieve of alleen negatieve waarden van x, dan resulteert (afgezien van een factor 2) de rechter- of de linkerhelft van de normale verdeling.

IV. Andere verdelingen en andere methoden

Met andere bijvoorwaarden dan in A, B, C van de vorige paragraaf kan men met dezelfde methode andere verdelingen trachten te vinden. In tal van gevallen stuit men dan op vergelijkingen voor parameters der verdeling, die niet of moeilijk expliciet oplosbaar zijn. Hiervan werden bij B en C al voorbeelden genoemd. Een ander voorbeeld is het stel bijvoorwaarden: interval van 0 tot ∞; $\int p(x)\,dx = 1$; $\int p(x) \log x\,dx = M$; $\int p(x)\,x\,dx = m$. Dit leidt tot $p(x) = e^{\lambda-1} x^{\mu} e^{\nu x}$, een gamma-verdeling.

Bovendien behoeven de bijvoorwaarden niet altijd in integraalvorm gegeven te zijn. B.v.: $\int p(x)\,dx = 1$ en $p(x) > 0$ voor alle reële x (zie ook IIIA). Nu is er zelfs geen $p(x)$ die de entropie maximaal maakt.

Het hangt dus van de vorm der bijvoorwaarden af, of men in de practijk iets aan de methode van Shannon heeft. Is alleen het gemiddelde of de variantie gegeven, dan krijgen we verdelingen als in III. Bij een groter aantal of minder eenvoudige voorwaarden zal de methode van Shannon wellicht dikwijls falen. Kan dit al bij theoretisch gefundeerde voorwaarden het geval zijn, bij een aan waarnemingsuitkomsten aan te passen verdeling zal men misschien niet eens een stel bijvoorwaarden voor het maximumprobleem kunnen formuleren. In al deze gevallen zijn andere methoden nodig om een geschikte verdeling te zoeken.

Litteratuur

[1] C. E. Shannon, Bell. Syst. Techn. Journ. **27**, 369 en 623, 1948.
[2] A. I. Khinchin, Mathematical foundations of information theory, 1957.
[3] J. L. van Soest, Statistica Neerlandica **14**, 249, 1960.
Zie ook A. J. Stam, Statistica Neerlandica **14**, 259, 1960.
[4] L. Boltzmann, Vorlesungen über Gastheorie, 3e druk, 1923.

[5] E. Bloch, Théorie cinétique des gaz, 1921.
[6] J. Zernike, Thermodynamica en Statistiek in de chemie, 1942.
[7] A. M. Yaglom et I. M. Yaglom, Probabilité et information, 1959.
[8] R. Courant, Diff. and integral calculus II, Chapter VII, 1948.

J. van Meurs
Mathematisch Centrum, Amsterdam

J. H. C. Lisman
Centraal Planbureau, 's-Gravenhage