

## COMPUTACIÓN DE FUNCIONES DISTRIBUCIÓN

LUIS M. BOGGIA y OSCAR M. SORARRAIN \*

Los métodos clásicos de integración numérica se reducen a la elección "conveniente" de ciertas abscisas  $x_i$  y de ciertos correspondientes coeficientes de peso  $c_i$  tales que aproximadamente se cumpla que:

$$\int_a^b f(x) dx \cong \sum_0^n c_i f(x_i) \quad (1)$$

Las  $f(x_i)$  pueden ser valores exactos (cuando  $f(x)$  es conocida explícitamente) o si no aproximaciones tomadas de curvas que se ajustan a un gran número de datos experimentales.

El hecho de que la elección de los  $x_i$  pueda llamarse o no conveniente radica, entre otras cosas, en el sentido o alcance que se dé al signo  $\cong$  de la (1) y sobre todo del conocimiento o grado de confianza que se tenga de los valores  $f(x_i)$ . En aquellos casos en que la función se conozca en forma explícita, debe tenerse cuidado con la posible existencia de singularidades en el intervalo de integración, que puedan invalidar el cómputo.

En los casos en que  $f(x)$  sea conocida como polinomio que aproximadamente se ajusta a una serie de datos experimentales, suele usarse como criterio para tal aproximación (que da el alcance del símbolo  $\cong$ ) el de LEGENDRE o de los mínimos cuadrados. Este criterio supone que el conocimiento de  $f(x)$  se tiene con una precisión relativa que es a su vez una función peso  $p(x)$ , y que la mejor aproximación la da aquel polinomio, de grado  $n$ ,  $P_n(x)$ , tal que se cumpla que

$$\int_a^b p(x) [f(x) - P_n(x)]^2 dx = \text{MINIMO} \quad (2)$$

\* Dr. Luis M. Boggia, Profesor e Investigador de la Facultad de Ciencias Físicomatemáticas y Profesor de la Facultad de Ciencias Económicas de la Universidad Nacional de La Plata.

Dr. Oscar M. Sorrain, Profesor e Investigador de la Facultad de Ciencias Físicomatemáticas de la Universidad Nacional de La Plata. (La Dirección.)

una manera de hallar tal polinomio es suponer que deba ser de la forma:

$$P_n(x) = \sum_0^n a_r \Phi_r(x) \quad (3)$$

en la que cada  $\Phi_r(x)$  es a su vez un polinomio de grado  $r$ .

Si ahora por simplicidad indicamos con  $A(x)$  al integrando de la fórmula (2) la condición de mínimo se traduce en que deben satisfacerse las siguientes igualdades

$$\frac{\partial A(x)}{\partial a_r} = 0 \quad r = 0, 1, \dots, n \quad (4)$$

Este sistema de  $n+1$  ecuaciones, llamadas ecuaciones normales, permiten calcular los parámetros  $a_r$  en función de los polinomios  $\Phi_r(x)$  elegidos y de la función de peso  $p(x)$ .

En el caso en que el intervalo de integración es semiinfinito, por ejemplo:

$$\int_a^\infty \varphi(x) dx \quad (1')$$

debe primero trasladarse el origen mediante un cambio de coordenadas para que la integral tome la forma

$$\int_0^\infty \varphi(x) dx$$

luego para asegurar la convergencia de la integral (en tal intervalo) resulta razonable elegir como función peso  $a$

$$p(x) = e^{-ax}$$

de manera que la integral se transforme en

$$\int_0^\infty \varphi(x) dx = \int_0^\infty e^{-ax} f(x) dx \quad (5)$$

y de acuerdo al criterio de LEGENDRE deberá cumplirse las (2), es decir

$$\int_0^\infty e^{-ax} [f(x) - P_n(x)]^2 dx = \text{MINIMO} \quad (2')$$

En este caso se conocen (ver HILDEBRAND: *Introduction to numerical analysis*) tanto las  $\Phi_r(x)$  como los  $a_r$  de la ecuación (3).

El exponente  $a$  que aparece en la (5) puede elegirse igual a 1, siempre

que ello no haga divergente a la integral (2'), caso contrario deberá hacerse tan grande como sea necesario.

Precisamente en el caso en que  $\alpha = 1$ , las  $\Phi_r(x)$  son directamente los polinomios de LAGUERRE de orden  $r$

$$\Phi_r(x) = e^x \frac{d^r}{dx^r} (x^r e^{-x}) = L_r(x) \quad (6)$$

Dando sucesivos valores 0,1,2,... a  $r$ , se hallan los polinomios

$$\begin{aligned} L_0(x) &= 1 \\ L_1(x) &= 1 - x \\ L_2(x) &= 2 - 4x + x^2 \\ L_3(x) &= 6 - 18x + 9x^2 - x^3 \\ L_4(x) &= 24 - 96x + 72x^2 - 16x^3 + x^4 \\ L_5(x) &= 120 - 600x + 600x^2 - 200x^3 + 25x^4 - x^5 \end{aligned} \quad (7)$$

pudiendo los sucesivos obtenerse mediante la útil fórmula de recurrencia

$$L_{n+1}(x) = (1 + 2n - x) L_n(x) - n^2 L_{n-1}(x) \quad (8)$$

Puede también probarse que en el intervalo  $0, \infty$  estos polinomios poseen raíces distintas reales, y positivas  $x^k$ .

En el caso general en que  $\alpha$  es distinto de 1, resulta ser:

$$\Phi_r(x) = e^{\alpha x} \frac{d^r}{dx^r} (x^r e^{-\alpha x}) \quad (9)$$

y por lo tanto:

$$\Phi_r(x) = L_r(\alpha x) \quad (10)$$

Recordemos ahora la fórmula interpolación de HERMITE. Esta fórmula se usa cada vez que de una función se conozcan  $n+1$  ordenadas  $y_i$  y sus correspondientes  $n+1$  derivadas  $y'_i$ . Para obtener un polinomio que se ajuste a esas  $2n+2$  condiciones habrá que poner

$$P_n(x) = \sum_0^n h_i(x) y_i + \sum_0^n h_i^*(x) y_i' \quad (11)$$

y hallar los  $h_i(x)$  y los  $h_i^*(x)$  de manera que

$$P_n(x_i) = y_i \quad (12)$$

y que

$$P_n'(x_i) = y_i' \quad (13)$$

En este caso  $P_n(x)$  pasaría por los  $n+1$  puntos y tendría en ellos las  $n+1$  derivadas dadas y habríamos llegado a una aproximación de orden  $2n+1$ .

La obtención de los coeficientes del desarrollo de HERMITE no es muy difícil si se introducen los conocidos polinomios de LAGRANGE de orden  $i$

$$l_i(x) = \frac{K=i}{\prod_{\substack{K=0 \\ K \neq i}}^n} \frac{x - x_K}{x_i - x_K} \quad (14)$$

que son polinomios de grado  $n$  que obviamente se hacen igual a cero para todos los  $x_K$  excepto para  $K=i$  para el cual se hace igual a uno.

Usando la notación de KRONECKER se podría poner:

$$l_i(x_j) = \delta_{ij}$$

Para evitar excluir el factor correspondiente al caso  $K=i$  suele usarse a veces la expresión equivalente a la (14)

$$l_i(x) = \frac{\prod_0^n (x - x_K)}{(x - x_i) \prod_0^n (x_i - x_K)}$$

en la que  $\prod$  indica la derivada del producto.

Si ahora derivamos la (11) e imponemos a  $P_n(x)$  y  $p'(x)$  las condiciones (12) y (13) se obtendría

$$\begin{aligned} h_i(x) &= [1 - 2l_i'(x_i)(x - x_i)] [l_i(x)]^2 \\ h_i^*(x) &= (x - x_i) [l_i(x)]^2 \end{aligned} \quad (16)$$

Es decir que estos coeficientes permitirían hallar el polinomio  $P_n(x)$ , (con aproximación de orden  $2n+1$ ) cualesquiera sean las  $n+1$  abscisas  $x_K$

que se hubieran elegido. Si usáramos dicha expresión para aproximar la función  $f(x)$  del integrando de la (5) se obtendría

$$\int_0^{\infty} e^{-x} f(x) dx = \int_0^{\infty} \sum_0^r e^{-x} h_i(x) y_i dx + \int_0^{\infty} \sum_0^n e^{-x} h_i^*(x) y_i' dx + E_n \quad (17)$$

En la que con  $E_n$  se indica el error correspondiente a esta aproximación y cuya cota es:

$$E_n = \frac{[(n+1)!]^2}{(2n+2)!} f^{(2n+2)}(\xi) \quad 0 \leq \xi \leq \infty \quad (18)$$

Volviendo ahora a la (17) y si suponemos  $\alpha = 1$ , puede ponerse

$$\begin{aligned} \int_0^{\infty} e^{-x} f(x) dx &= \sum_0^n \int_0^{\infty} e^{-x} h_i(x) y_i dx + \sum_0^n \int_0^{\infty} e^{-x} h_i^*(x) y_i' dx + E_n \\ &= \sum_0^n c_i y_i + \sum_0^n c_i^* y_i' + E_n \end{aligned} \quad (19)$$

Nada se ha dicho hasta ahora sobre la elección de las abscisas  $x_k$  que son arbitrarias. Pero si se eligen los  $x_k$  de manera que sean las  $n+1$  raíces reales, positivas y distintas del polinomio de LAGUERRE de orden  $n+1$ , puede probarse (ver HILDEBRAND, libro citado) que resultan

$$c_i = \frac{[(n+1)!]^2}{[L_{n+2}(x_i)]^2} x_i \quad (20)$$

$$c_i^* = 0$$

y por lo tanto la (19) se reduciría a

$$\int_0^{\infty} e^{-x} f(x) dx = \sum_0^n c_i y_i + E_n \quad (19')$$

que es nuestra fórmula inicial (1).

El método en sí no es más que una cuadratura de GAUSS, aunque con el auxilio de los polinomios de LAGUERRE para obtener una aproximación de orden mayor, de allí el nombre de método de GAUSS-LAGUERRE con que se lo conoce en la literatura.

Integrales del tipo (1') se presentan corrientemente en estadística cuando se requiere hacer uso de las llamadas función distribución o función probabilidad acumulada.

Se define como tal a la función:

$$\Phi(x) = \int_{-\infty}^x \delta(y) dy$$

en la que  $\delta(y)$  es "alguna función frecuencia y como tal debe satisfacer la integral

$$\int_{-\infty}^{\infty} \delta(y) dy = 1$$

En el caso particular en que

$$\delta y = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}$$

se dice que la función es NORMAL o gaussiana y en tal caso

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-y^2/2} dy \quad (21)$$

es la llamada función distribución normal.

Evidentemente  $\Phi(x)$  es una función positiva, monótonamente creciente entre 0 y 1, y el integrando tiene un comportamiento en el infinito que permite aplicarle el método de LAGUERRE-GAUSS.

Para ello primero hagamos un cambio de variable tal que los límites de integración sean 0 e infinito

$$y = x - Y$$

Entonces

$$\int_{-\infty}^x f(y) dy = - \int_{+\infty}^0 f(x - Y) dY = \int_0^{\infty} f(x - Y) dY \quad (22)$$

y por lo tanto, sustituyendo y llevando a la forma (5)

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-y^2/2} dy = \frac{1}{\sqrt{2\pi}} \int_0^{\infty} e^{-Y} e^{[Y - \frac{(x-Y)^2}{2}]} dY \quad (23)$$

similar a la (19').

La integración se limita a tomar de tablas, los valores de las raíces de los polinomios de LAGUERRE, calcular los correspondientes  $c_i$  con la fórmula (20) y efectuar la sumatoria (19').

Cuando no se disponga de tablas adecuadas habrá que resolver previamente, las ecuaciones de LAGUERRE que se obtienen igualando a cero los polinomios  $L_n(x)$  que vimos en (7).

En nuestro caso disponemos de una modesta tabla con sólo cuatro y cinco puntos tomadas del texto de HILDEBRAND y la utilizaremos para hallar mediante computadora, una tabla de la función distribución gaussiana, para  $x$  variando de 0 a 1,5 a intervalos de 0.01.

Con el programa que se adjunta la computadora realizó en menos de tres minutos las ciento cincuenta integrales aproximadas.

Para el caso en que sólo se usaron cuatro puntos la aproximación obtenida difiere de tablas conocidas a lo sumo en un error del orden del 4 %. Tomando un punto más se lograron aproximaciones con un error máximo relativo del 3 %. No se pudieron hacer mejores aproximaciones por no disponerse en nuestro medio de tablas de polinomios de LAGUERRE y de sus raíces. Se está desarrollando un programa que llenará ese vacío y que se publicará próximamente.

El método puede ser aplicado a cualquier función similar con sólo cambiar unas pocas instrucciones del programa adjunto.

**APÉNDICE**  
**PROGRAMA USADO**

DIMENSIÓN C(10), Y(10), FI(10)

```
20 PRINT 100
   PRINT 101
   I = 0

5  I = I + 1
   READ 130, C(I), Y(I)
   IF(C(I))10, 30, 10

10 PRINT 130, C(I), Y(I)
   GO TO 5

30 N = 1 -- 1
   DENOM = SQRTF (2.*3.1415926)
   PRINT 103
   X = 0.

1  DO 25 J = 1,10
   SUM = 0.
   DO 15 I = 1,N

15 SUM = SUM + C(I)*EXPF(Y(I) - (Y(I) - X)**2*0.5)
   FI(J) = SUM/DENOM

25 X = X + 0.01
   EQUIS = X - 0.1
   PRINT 106, EQUIS, (FI(J), J = 1, 10)
   IF(X - 1.5)1, 1, 2

2  PRIN 104
   PRINT 105
   PAUSE
   GO TO 20

100 FORMAT(17X5HDATOS,/)
101 FORMAT(3X12HCOEFICIENTES,6X8HABSCISAS,/)
```



```
103 FORMAT(4×1HX,5×4HO.00,4×4HO.01,4×4HO.02,4
      ×4HO.03,4×4HO.04,4×4HO.05,4×4HO.06,4×4HO.07,4×4HO.08,4×4HO.09,/)
104 FORMAT(///,46 HESTA EN PAUSA, SI TIENE MÁS DATOS APRIETE
      START)
105 FORMAT (1 × 47H SINO APRIETE RESET, INSERT,4900796, RELÉASE,
      START,////)
106 FORMAT(F6.2,10(2×F6.4))
130 FORMAT(LE15.10)
      END
```

**DATOS**

COEFICIENTES	ABSCISAS
.60315400	.32254800
.35741900	1.74576100
.03888790	4.53662000
.00053929	9.39507100

**FUNCION DISTRIBUCION NORMAL**

X	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.00	.4934	.4975	.5017	.5058	.5100	.5141	.5183	.5225	.5267	.5308
.10	.5350	.5392	.5434	.5476	.5519	.5561	.5603	.5645	.5687	.5730
.20	.5772	.5814	.5857	.5899	.5942	.5984	.6026	.6069	.6111	.6154
.30	.6196	.6238	.6281	.6323	.6365	.6407	.6450	.6492	.6534	.6576
.40	.6618	.6660	.6702	.6744	.6786	.6827	.6869	.6911	.6952	.6993
.50	.7035	.7076	.7117	.7158	.7199	.7239	.7280	.7320	.7361	.7401
.60	.7441	.7481	.7520	.7560	.7599	.7638	.7677	.7716	.7755	.7793
.70	.7832	.7870	.7907	.7945	.7982	.8020	.8056	.8093	.8130	.8166
.80	.8202	.8237	.8273	.8308	.8342	.8377	.8411	.8445	.8479	.8512
.90	.8545	.8577	.8610	.8642	.8673	.8705	.8736	.8766	.8796	.8826
1.00	.8856	.8885	.8913	.8941	.8969	.8997	.9024	.9050	.9077	.9102
1.10	.9128	.9153	.9177	.9201	.9224	.9248	.9270	.9292	.9314	.9335
1.20	.9356	.9376	.9396	.9415	.9433	.9452	.9469	.9487	.9503	.9519
1.30	.9535	.9550	.9564	.9578	.9592	.9605	.9617	.9629	.9640	.9651
1.40	.9661	.9671	.9680	.9688	.9696	.9704	.9711	.9717	.9723	.9728
1.50	.9732	.9736	.9740	.9743	.9745	.9747	.9748	.9749	.9749	.9748

**DATOS**

COEFICIENTES	ABSCISAS
.52175600	.26356000
.39866700	1.41340300
.07594240	3.59642600
.00361176	7.08581000
.00002337	12.64080100

**FUNCION DISTRIBUCION NORMAL**

X	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.00	.5041	.5082	.5124	.5166	.5207	.5249	.5290	.5332	.5373	.5415
.10	.5456	.5498	.5539	.5581	.5622	.5663	.5704	.5745	.5787	.5828
.20	.5868	.5909	.5950	.5991	.6031	.6072	.6112	.6152	.6192	.6232
.30	.6272	.6312	.6351	.6391	.6430	.6469	.6508	.6547	.6585	.6624
.40	.6662	.6700	.6738	.6775	.6813	.6850	.6887	.6924	.6960	.6997
.50	.7033	.7068	.7104	.7139	.7174	.7209	.7244	.7278	.7312	.7346
.60	.7379	.7412	.7445	.7478	.7510	.7542	.7574	.7605	.7636	.7667
.70	.7697	.7727	.7757	.7786	.7815	.7844	.7873	.7901	.7928	.7956
.80	.7983	.8009	.8036	.8062	.8087	.8112	.8137	.8162	.8186	.8210
.90	.8233	.8256	.8279	.8301	.8323	.8344	.8365	.8386	.8407	.8427
1.00	.8446	.8465	.8484	.8503	.8521	.8539	.8556	.8573	.8590	.8606
1.10	.8622	.8638	.8653	.8668	.8682	.8696	.8710	.8724	.8737	.8750
1.20	.8762	.8774	.8786	.8797	.8808	.8819	.8830	.8840	.8850	.8859
1.30	.8869	.8878	.8886	.8895	.8903	.8911	.8918	.8926	.8933	.8939
1.40	.8946	.8952	.8959	.8965	.8970	.8976	.8981	.8986	.8991	.8996
1.50	.9000	.9005	.9009	.9013	.9017	.9021	.9025	.9028	.9032	.9035

## KOMPUTATION VON VERTEILUNGSFUNKTIONEN

### Zusammenfassung

In konkreter Anwendung werden die Möglichkeiten der GAUSS-LAGUERRE Methode der angenäherten Integration bei halbunendlichen Intervallen gezeigt.

Es wird eine Tabelle der normalen Verteilungsfunktion in ihren charakteristischsten Anordnung unter Benutzung von nur 4 und 5 Ordinaten aufgestellt. Es wird das bei der I.B.M.-Rechenmaschine 1620 der Universität La Plata benutzte Programm beigelegt. In weniger als drei Minuten wurden 150 Integrale berechnet. Die relativen Maximalfehler betragen 4 % für 4 Ordinaten und 3% für 5 Ordinaten.

## COMPUTATION DES FONCTIONS — DISTRIBUTION

### Résumé

On montre, dans la pratique, les possibilités de la méthode GAUSS-LAGUERRE d'intégration approximative dans des intervalles semi-infinites.

Il faut d'abord étalonner la fonction-distribution normale dans son rang le plus significatif, en utilisant seulement quatre ou cinq ordonnées. On ajoute après le programme employé avec la I.B.M. 1620 de l'Université Nationale de La Plata. On a calculé 150 intégrales en moins de trois minutes. Les erreurs relatives maxima on été de l'ordre du 4 % pour quatre ordonnées, et du 3 % pour cinq ordonnées.

## COMPUTING A DISTRIBUTION FUNCTION

### Summary

We show, through concrete application, the possibilities of the GAUSS-LAGUERRE method of approximate integration at the semi-infinite intervals.

The normal distribution function was tabulated in its most significant range by means of only 4 and 5 ordinates. Attached is the program used on the I.B.M. 1620 computer belonging to the National University of La Plata (Argentina). In less than three minutes 150 integrals were calculated. The relative maximum errors were of 4 % in 4 ordinates and 3 % in 5 ordinates.

## COMPUTO DI FUNZIONI DISTRIBUZIONE

### Riassunto

Si mostra, in una applicazione concreta, le possibilità del metodo di GAUSS-LAGUERRE d'integrazione approssimativa in intervalli semiinfiniti.

Si computa la funzione distribuzione normale nel suo livello più significativo per mezzo dell'uso di soltanto 4 e 5 ordinate. Si aggiunge il programma usato nella computeratrice I.B.M. 1620 dell'Università Nazionale di La Plata. In meno di tre minuti furono calcolati 150 integrali. Gli errori relativi massimi furono dell'ordine del 4 % per 4 ordinate e del 3 % per 5 ordinate.