

ALGORITMO PARA EL CALCULO DE AREAS BAJO LA DISTRIBUCIÓN GAUSSIANA

Ing. JESUS RUBEN AZOR MONTOYA
Profesor Titular de Estadística Aplicada

Introducción

El presente trabajo tiene por objeto desarrollar un sencillo algoritmo que permite calcular áreas bajo la Distribución Normal o Gaussiana con un determinado nivel de exactitud.

Para lograr este cometido se hace uso del sencillo concepto de la Serie de Mac Laurin para desarrollar algebraicamente la función densidad de la distribución normal.

A partir de allí, integrando en forma definida la serie hallada a través del procedimiento anterior, se llega a otra serie que resuelve el problema.

Finalmente se indica un procedimiento algorítmico junto a un segmento de programa en C que permite la realización (y por ende verificación) de los resultados obtenidos.

Desarrollo teórico

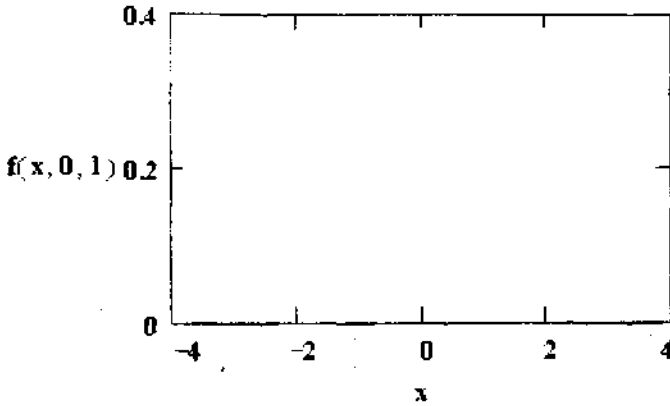
Quando se trata con distribuciones de probabilidad continuas, la más importante es la "distribución Gaussiana o normal", cuya función densidad está dada por la ecuación.

$$f(x, \mu, \sigma) = \frac{1}{\sqrt{2 \cdot \pi \cdot \sigma}} \cdot \exp \left[-\frac{1}{2} \cdot \left(\frac{x - \mu}{\sigma} \right)^2 \right]$$

en el rango:

$$-\infty < X < \infty$$

La cual, gráficamente, se ve del siguiente modo (cuando la media es nula y la varianza es uno, $\mu = 0$ y $\sigma^2 = 1$):



Si la curva representa una distribución de probabilidad, se debe cumplir que el área entre ella y el eje de las x debe ser 1.

De los datos más importantes que se solicitan cuando se trata con distribuciones en general, está el área bajo la "campana de Gauss" entre dos valores de abscisa determinados. Esto permite calcular la probabilidad que la variable aleatoria x se encuentre entre esos dos valores

Matemáticamente:

$$\text{Area} = \frac{1}{\sqrt{2 \cdot \pi} \cdot \sigma} \cdot \int_a^b f(x, \mu, \sigma) dx = \frac{1}{\sqrt{2 \cdot \pi} \cdot \sigma} \cdot \int_a^b \exp \left[-\frac{1}{2} \cdot \left(\frac{x - \mu}{\sigma} \right)^2 \right] dx$$

Vale decir, el área bajo la curva entre las abscisas a y b . Si se define la variable z como:

$$z = \frac{x - \mu}{\sigma}$$

La expresión anterior, para el caso de la Distribución Normal Standard, queda:

$$\text{Area} = \frac{1}{\sqrt{2 \cdot \pi}} \cdot \int_a^b \exp \left(-\frac{1}{2} \cdot z^2 \right) dz = \frac{1}{\sqrt{2 \cdot \pi}} \cdot \int_a^b f(z) dz \quad (1)$$

Reviste especial importancia, en temas tales como las pruebas de hipótesis, el valor del área bajo la campana entre dos valores de abscisa. Lo que se puede deducir fácilmente si se conoce el área entre $z=0$ y una abscisa genérica cualquiera z , que es lo que entregan generalmente las tablas utilizadas en estadística y probabilidad.

Es posible desarrollar en Serie de Mac Laurin una expresión para la ecuación (1). Para ello se deben hallar las derivadas de primero, segundo,... orden, las cuales son:

$$\frac{d}{dz} f(z) = -z \cdot \exp\left(-\frac{1}{2} \cdot z^2\right)$$

$$\frac{d^2}{dz^2} f(z) = (z^2 - 1) \cdot \exp\left(-\frac{1}{2} \cdot z^2\right)$$

$$\frac{d^3}{dz^3} f(z) = (-z^3 + z) \cdot \exp\left(-\frac{1}{2} \cdot z^2\right)$$

$$\frac{d^4}{dz^4} f(z) = (z^4 - 6 \cdot z^2 + 3) \cdot \exp\left(-\frac{1}{2} \cdot z^2\right)$$

$$\frac{d^5}{dz^5} f(z) = (-z^5 + 10 \cdot z^3 - 15 \cdot z) \cdot \exp\left(-\frac{1}{2} \cdot z^2\right)$$

$$\frac{d^6}{dz^6} f(z) = (z^6 - 15 \cdot z^4 + 45 \cdot z^2 - 15) \cdot \exp\left(-\frac{1}{2} \cdot z^2\right)$$

$$\frac{d^7}{dz^7} f(z) = (-z^7 + 21 \cdot z^5 - 105 \cdot z^3 + 105 \cdot z) \cdot \exp\left(-\frac{1}{2} \cdot z^2\right)$$

$$\frac{d^8}{dz^8} f(z) = (z^8 - 28 \cdot z^6 + 210 \cdot z^4 - 420 \cdot z^2 + 105) \cdot \exp\left(-\frac{1}{2} \cdot z^2\right)$$

Con estas primeras ocho derivadas, y por simple inspección de los coeficientes de la serie (valores de la función y de sus derivadas en $z=0$), se ve que siendo:

$$\begin{array}{llll} f(0)=1 & f'(0)=0 & f''(0)=0 & f'''(0)=3 \\ f^{(4)}(0)=0 & f^{(5)}(0)=-15 & f^{(6)}(0)=0 & f^{(7)}(0)=105 \end{array}$$

a) Los coeficientes de lugares impares en la serie (considerando al

primero como lugar 0) son todos nulos. Esto significa que la función es par.

b) Todos los coeficientes no nulos, excepto el primero, se obtienen multiplicando el anterior por los números 1;3;5;...; etc. alternando los signos.

En base a lo expuesto, el desarrollo en Serie de Mac Laurin será:

$$f(z) = 1 - \frac{z^2}{2!} + \frac{3 \cdot z^4}{4!} - \frac{15 \cdot z^6}{6!} + \frac{105 \cdot z^8}{8!} - \frac{945 \cdot z^{10}}{10!} + \dots$$

Explicitando operaciones:

$$f(z) = \left(1 - \frac{z^2}{2 \cdot 1} + \frac{3 \cdot 1 \cdot z^4}{4 \cdot 3 \cdot 2 \cdot 1} - \frac{5 \cdot 3 \cdot 1 \cdot z^6}{6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1} \right) + \frac{7 \cdot 5 \cdot 3 \cdot 1 \cdot z^8}{8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1} - \dots$$

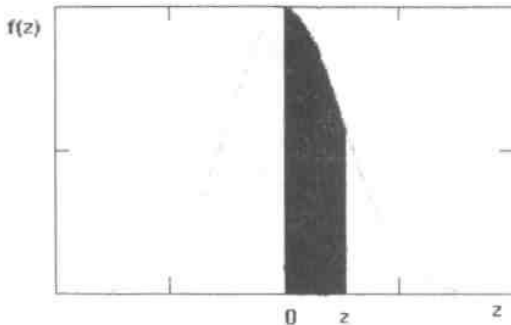
Si se define $n!_p$ (se lee "n factorial sub p") como $n!_p = n \cdot (n-2) \cdot (n-4) \cdot \dots \cdot 4 \cdot 2$ llamado "factorial de pares", la expresión anterior se reduce a:

$$f(z) = 1 - \frac{z^2}{(2!)_p} + \frac{z^4}{(4!)_p} - \frac{z^6}{(6!)_p} + \frac{z^8}{(8!)_p} - \frac{z^{10}}{(10!)_p} + \dots$$

Recuérdese que para obtener el valor correcto del área bajo la campana entre $z=0$ y z , se debe multiplicar a esta expresión por el factor:

$$\frac{1}{\sqrt{2 \cdot \pi}}$$

Para calcular el área bajo la curva que se indica en el siguiente gráfico:



$$\text{Area} = \frac{1}{\sqrt{2 \cdot \pi}} \int_a^b f(z) dz$$

El desarrollo es el siguiente:

$$\text{Area} = \frac{1}{\sqrt{2\pi}} \cdot \int_0^z \left[1 - \frac{z^2}{(2!)_p} + \frac{z^4}{(4!)_p} - \frac{z^6}{(6!)_p} + \frac{z^8}{(8!)_p} - \frac{z^{10}}{(10!)_p} + \dots \right] dz$$

$$\text{Area} = \frac{1}{\sqrt{2\pi}} \cdot \left[z - \frac{z^3}{3 \cdot (2!)_p} + \frac{z^5}{5 \cdot (4!)_p} - \frac{z^7}{7 \cdot (6!)_p} + \frac{z^9}{9 \cdot (8!)_p} - \frac{z^{11}}{11 \cdot (10!)_p} + \dots \right]$$

Algoritmo

De aquí el paso a un procedimiento algorítmico que mecanice el cálculo de áreas bajo la curva normal, es sencillo:

a) Se introducen como datos la abscisa z y el error máximo E con que se quiere calcular el área bajo la normal.

b) Se realizan las siguientes asignaciones de inicialización del proceso:

$$S=z \quad A=z \quad n=0$$

c) Se calculan:

$$A = A \cdot (-z \cdot z / (n+2)) \quad n = n+2 \quad A1 = A \cdot (n+1)$$

d) Si el valor absoluto de $A1$ es mayor que el error E , se calcula $S = S+A1$ y se repite la secuencia desde c)

e) Si no, se calcula el área del siguiente modo:

$$\text{Area} = \frac{1}{\sqrt{2\pi}} \cdot S$$

La convergencia de la expresión en serie desarrollada se mantiene en valores aceptables dentro del rango $-3 < z < 3$.

Programa

*/*El siguiente programa permite hallar áreas bajo la curva normal para abscisas comprendidas entre 0 y z, con un error E preestablecido */*

```
#include <stdio.h>
#include <math.h>
#include <conio.h>
main()
{
float z,E,A,S,A1,pi;
int n;
printf("introduzca valor de z:");
scanf("%f",&z);
printf("introduzca máximo error admitido:");
scanf("%f",&E);
printf("\n");
pi=atan(1)*4; S=z; A=z; n=0; /* inicialización */
do {
    A=-A*z*z/(n+2);
    n=n+2;
    A1=A/(n+1);
    S=S+A1;
}
while (fabs(A1)>E);
printf("El área entre 0 y %f",z);
printf(" es: ");
printf("%f",S/sqrt(2*pi));
}
```