Variables aleatorias. Distribución binomial y normal

Variable aleatoria

Def.- Al realizar un experimento aleatorio tenemos un espacio muestral E. A cualquier ley o aplicación que a cualquier suceso de E le asocie un número real se le llama **variable aleatoria**, y se suele escribir con letras mayúsculas: X, Y, Z

X: Suceso de E $\rightarrow$ R

♦♦ Se lanzan 2 dados, se define la variable aleatoria X como la suma de los números que aparecen arriba

E = {(1,1); (1,2);…(6,6)} hay 36 = 6x6 casos

Posibles valores de la variable aleatoria X: 2,3,4,5,6,7,8,9,10,11,12

♦♦ Se lanzan 2 dados y se suman los números que aparecen arriba, se define la variable aleatoria X salga un número múltiplo de 3

E = {(1,1); (1,2);…(6,6)} hay 36 = 6x6 casos

Posibles valores de X: 3, 6, 9, 12

♦♦ Lanzo 3 monedas y defino la variable aleatoria X como el nº de caras que salen

E tiene 2x2x2 = sucesos elementales

Posibles resultados de X: 0,1,2,3

♦♦ En un cultivo elijo 100 habichuelas con la vaina entera. Defino la variable aleatoria X como la longitud de la vaina.

♦♦ Alumnos de 2º de Bachillerato de un IES. Defino la variable aleatoria X como su altura

♦♦ Alumnos de 2º de Bachillerato de un IES. Defino la variable aleatoria X como su peso

♦♦ Levantamos una ficha de domino y defino la variable aleatoria X como la suma de los puntos de arriba

E tiene 28 elementos (la ficha 0-1 es la misma que la 1-0, y se cuenta sólo una vez)

fichas 0-0, 0-1, 0-2, …, 0-6
1-1, 1-2, …, 1-6
………………..
6-6

posibles resultados de X: 0,1,2,3,4,5,6,7,8,9,10,11,12

♦♦ Sacamos dos cartas de una baraja. Variable X nº de ases obtenidos; posibles resultados X: 0,1,2,

♦♦ Un alumno ha estudiado 12 temas de 30 de los que consta el examen. El examen consta de dos temas elegidos al azar. Variable X preguntas que se sabe posibles resultados de X: 0,1,2

Nota.- De los ejemplos anteriores hemos visto que hay variables aleatorias que se les asocia un número finito de valores y otras que en teoría todos los valores de un intervalo numérico. (ejercicios anteriores 4, 5 y 6)

Def.- Una **variable aleatoria** X decimos que es **discreta** si se le asocia un numero finito de valores.

Def.- Una **variable aleatoria** X diremos que es **continua** si se la asocia, en teoría, todos los valores de un intervalo.

Variable aleatoria discreta

Función de probabilidad

Suponemos que tenemos una variable aleatoria discreta **X** con valores x_1, x_2, …., x_n, y conocemos las siguientes probabilidades $p(X= x_1) = p_1$, $p(X= x_2) = p_2$, ….., $p(X= x_n) = p_n$.

Def.- Dada una variable aleatoria discreta **X** con valores x_1, x_2, …., x_n, de los cuales conocemos sus probabilidades p_1, p_2, p_3, …….., p_n. Se define su función de probabilidad como la ley o aplicación que asocia a cada valor $\mathbf{x}_i$ su probabilidad $p(X=x_i)=\mathbf{p_i}$, verificando:

1) $p_i > o$

2) $p_1+ p_2+ p_3+ …….+ p_n = 1$

Ejemplos

♦♦ Determinar la función de probabilidad de los ejemplos 1,2,3,7,8, 9

♦♦ Del 1) Se lanzan 2 dados, se define la variable aleatoria X como la suma de los números que aparecen arriba

E = {(1,1); (1,2);…(6,6)} hay 36 = 6x6 casos

Posibles valores de X: 2,3,4,5,6,7,8,9,10,11,12

$p(X=2)=p(X=12) = 1/36$
$p(X=3)=p(X=11) = 2/36$
$p(X=4)=p(X=10) = 3/36$
$p(X=5)=p(X=9) = 4/36$
$p(X=6)=p(X=8) = 5/36$
$p(X=7) = 6/36$

♦♦ Del 2.- Se lanzan 2 dados y se suman los números que aparecen arriba, se define la variable aleatoria X salga un número múltiplo de 3

E = {(1,1); (1,2);…(6,6)} hay 36 = 6x6 casos

Posibles valores de X: 3, 6, 9, 12

$p(X=3) = 2/36$
$p(X=6) = 5/36$
$p(X=9) = 4/36$
$p(X=12) = 1/36$

♦♦ Del 3.- Lanzo 3 monedas y defino la variable aleatoria X como el nº de caras que salen

E tiene 2x2x2 = sucesos elementales

Posibles resultados de X: 0,1,2,3

$p(X=0) = 1/8$
$p(X=6) = 3/8$
$p(X=9) = 3/8$

p(X=12) = 1/8

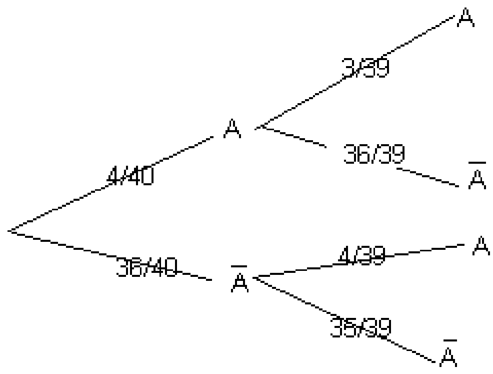
♦♦ Del 7.- Levantamos una ficha de domino y defino la variable aleatoria X como la suma de los puntos de arriaba

E 28 fichas 0-0, 0-1, 0-2, …, 0-6
1-1, 1-2, …, 1-6
………………..
6-6

posibles resultados de X: 0,1,2,3,4,5,6,7,8,9,10,11,12

♦♦ Dos cartas de una baraja X = nº de ases

X = 0,1,2, A = As



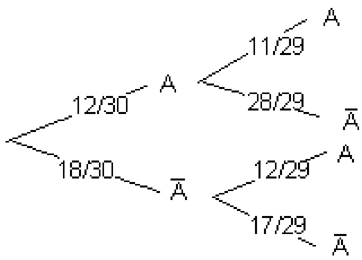
P(X=0) = ningún as = (36/40)(35/39). Sigo la línea donde no hay ases

P(X=1) = aparezca un as = (4/40)(36/39) + (36/40)(4/39)

P(X=2) = aparezcan dos ases = (4/40)(3/39).

♦♦ Se sabe 12 de 30 que consta el libro

X = 0,1,2; A = sabe



P(X=0) = (18/30)(17/29). Sigo la línea

P(X=1) = (12/30)(18/29) + (18/30)(12/29)

P(X=2) = (12/30)(11/29).

Función de distribución

Def.- La **función de distribución** de una variable aleatoria discreta X, que se escribe **F(x_i)**, es la suma de todas las probabilidades de la variable aleatoria X hasta el valor x_i, es decir

$F(x_i) = p(X \leq x_i) = p(X = x_1) + p(X = x_2) + p(X = x_3) + \ldots.. + p(X = x_i) =$

$= p_1 + p_2 + p_3 + \ldots.. + p_i$.

Es parecido a las frecuencias relativas acumuladas.

X	x_1 x_2 x_3 ………………….….. x_n
$p_n = p(X = x_i)$	p_1 p_2 p_3 ……………………...p_n

Su función de distribución es:

$$F(x)=\begin{cases} 0 & \text{si} \quad X<x_1 \\ p_1 & \text{si} \quad x_1 \le X<x_2 \\ p_1+p_2 & \text{si} \quad x_2 \le X<x_3 \\ \dots & \dots \quad \dots \\ p_1+p_2+....+p_i & \text{si} \quad x_i \le X<x_{i+1} \\ p_1+p_1+....+p_n=1 & \text{si} \quad x_n \le X<x_{n+1} \end{cases}$$

Propiedades:

1) F(x) es constante (toma el mismo valor) en cada intervalo $[x_i, x_{i+1})$

2) F(x) es creciente

3) $p(a < X \le b) = F(b) - F(a)$

X = Nº de caras al lanzar 3 monedas	0	1	2	3
$P_i = p(X = x_i)$	1/8	3/8	3/8	1/8

Su función de distribución es:

$$F(x)=\begin{cases} 0 & \text{si} \quad X<0 \\ 1/8 & \text{si} \quad 0 \le X<1 \\ 1/8+3/8=4/8 & \text{si} \quad 1\le X<2 \\ 4/8+3/8=7/8 & \text{si} \quad 2 \le X<3 \\ 7/8+1/8=1 & \text{si} \quad X \ge 3 \end{cases}$$

Se puede representar gráficamente y calcular algunas probabilidades

Media, varianza y desviación típica de una variable aleatoria discreta.

Nota: Suponemos una variable aleatoria discreta X que toma los valores $x_1, x_2, x_3, …., x_n$ con probabilidades $p_1, p_2, p_3, …, p_n$, donde $p_i = p(X = x_i)$

Def.- Se define la **media o esperanza matemática** de la variable aleatoria X, que se escribe μ_X, como el siguiente número:

$$\mu_X = p_1x_1 + p_2x_2 +\dots\dots\dots+ p_n x_n = \sum_{i=1}^{n} p_i x_i$$

Def.- Se define la **varianza** de la variable aleatoria X, que se escribe $(\sigma_X)^2$, como el siguiente número:

$$(\sigma_X)^2 = p_1(x_1 - \mu_X)^2 + p_2(x_2 - \mu_X)^2 +\dots\dots\dots+ p_n(x_n - \mu_X)^2 = \sum_{i=1}^{n} p_i(x_i-\mu_X)^2 = \sum_{i=1}^{n} p_i(x_i)^2-(\mu_X)^2$$

Def.- Se define la **desviación típica** de la variable aleatoria X, que se escribe (σ_X) , como el siguiente número:

$$(\sigma_X) = \sqrt{(\sigma_X)^2} = \sqrt{\sum_{i=1}^{n} p_i(x_i - \mu_X)^2} = \sqrt{\sum_{i=1}^{n} p_i(x_i)^2 - (\mu_X)^2}$$

♦♦ Calcular la media, varianza y desviación típica de la variable aleatoria X salir cara al lanzar 3 monedas

x_i	p_i	$p_i.x_i$	$p_i.(x_i)^2$
0	1/8	0	0
1	3/8	3/8	3/8
2	3/8	6/8	12/8
3	1/8	3/8	9/8
Σ	1	12/8	24/8

Media $\mu_X = \sum_{i=1}^{n} p_i x_i = 12/8$

Varianza $(\sigma_X)^2 = \sum_{i=1}^{n} p_i(x_i)^2 - (\mu_X)^2 = 24/8 - (12/8)^2 = 3/4$

Desviación típica $(\sigma_X) = \sqrt{(\sigma_X)^2} = \sqrt{\sum_{i=1}^{n} p_i(x_i)^2 - (\mu_X)^2} = \sqrt{\frac{3}{4}} = \frac{\sqrt{3}}{2} \cong 0'866$

♦♦ Hacerlo de los ejercicios resueltos anteriormente

Distribución Binomial (como ejemplo de variable aleatoria discreta)

Definición de distribución binomial:

Si realizamos **n** veces un experimento en el que podemos obtener éxito, A, con probabilidad **p** y fracaso, A^C, con probabilidad **q** (q = 1 − *p*), la obtención de éxito o fracaso es independiente de un experimento a otro.

La variable aleatoria discreta **X = nº de éxitos**, decimos que sigue una distribución binomial de parámetros **n** y **p**, y lo representaremos por X → **B(n;p)**.

Nota.- La **función de probabilidad** de la variable **X = nº de éxitos**, que sabemos sigue una distribución binomial de parámetros **n** y **p** (B(n,p)), viene dada por:

$$\mathbf{p(X = k) = (n\ sobre\ k).p^k.q^{(n-k)} = \binom{n}{k}.p^k.q^{(n-k)}.}$$

****** (n sobre k) = {número combinatorio} = $\binom{n}{k}$ = (n!)/(k!.(n-k)!)

En la calculadora (n sobre k) = " n tecla **nCr** k "

n! = factorial de un nº = n(n-1)(n-2)(n-3)….3.2.1

Por convenio 0! = 1, para poder demostrar las propiedades

(5 sobre 1) = $\binom{5}{1}$ = (5!)/(1!.(5-1)!) = 5

(7 sobre 3) = $\binom{7}{3}$ = (7!)/(3!.(7-3)!) =35

(8 sobre 4) = $\binom{8}{4}$ = (8!)/(4!.(8-4)!) = 70

(100 sobre 98) = $\binom{100}{98}$ = (100!)/(98!.(100-98)!) = 4950

***Nota**: Observar que las probabilidades de éxito y fracaso son complementarias, es decir,* **q** = 1 – p y **p** = 1 – q, por lo que basta saber una de ellas para calcular la otra.

♦♦ Antes teníamos B(7; 1/6) , y queríamos calcular p(X=3) (obtener 3 éxitos). Aplicando la fórmula: p(X = 3) = (7 sobre 3).$(1/6)^3.(5/6)^4$ = 0'0781

♦♦ Supongamos que la probabilidad de que una pareja tenga un hijo o una hija es igual. Calcular la probabilidad de que una familia con 6 descendientes tenga 2 hijos.

En este caso Éxito = A = "tener hijo" y p(A) = 0'5.

Fracaso = A^C "tener hija" y $p(A^C)$ = 0'5.

Estamos por tanto ante una binomial B(6;0'5) y nos piden p(X=2).

Si aplicamos la fórmula es: p(X = 2) = (6 sobre 2).$(0'5)^2.(0'5)^4$ = 0'2344

Probabilidades acumuladas

Nota.- Es posible que nos pidan no sólo la probabilidad de que ocurran un cierto numero de éxitos en concreto, sino que ocurran como mucho "k" éxitos o preguntas similares. En el ejemplo anterior, por ejemplo, podrían pedirnos:

♦♦ Cuál es la probabilidad de que aprueben como mucho 2 alumnos?.

Si éxito = aprobar y fracaso = suspender, p= 0'7 y q = 0'3, entonces nos piden p(X ≤ 2).

En este caso, basta pensar en que para que aprueben 2 alumnos como mucho, puede que aprueben 2, 1 o ninguno, es decir:

p(X ≤ 2) = p(X = 0)+p(X = 1)+p(X = 2) = 0'0001 + 0'0012 + 0'01 = 0'1013

***Nota.-** A veces es más fácil por la probabilidades contrario*

♦♦ Cuál es la probabilidad de que aprueben entre 3 y 6 alumnos (inclusive)?.

Del mismo modo:

p(3 ≤ X ≤ 6) = p(X = 3)+p(X = 4)+p(X = 5)+p(X = 6) =

= 0'0467 + 0'1361 + 0'2541 + 0'2965 = 0'7334

♦♦ Los alumnos de cierta clase se encuentran en una proporción del 67% que estudian inglés y el resto francés.

Tomamos una muestra de 15 alumnos de la clase, calcular:

a) Probabilidad de que al menos encontremos tres alumnos de inglés.

b) Probabilidad de que los 15 alumnos estudien inglés.

c) Probabilidad de que estudien inglés entre 7 y 10 alumnos.

Sol

Si éxito = estudiar inglés, p = 0'67 y fracaso = estudiar francés, q = 1-0'67 = 0'33.

Manejamos por tanto una Bin(15;0'67)

a) p(X ≥ 3) = p(X = 3)+p(X = 4)+p(X = 5)+p(X = 6)+. . .+ p(X = 15).

Una opción es calcular estas 13 probabilidades y sumarlas. Como hay que aplicar la fórmula para calcular cada una, la tarea se puede hacer bastante larga. Otra opción, mas sencilla, es pasar al complementario. El complementario de encontrar al menos 3 alumnos de inglés es encontrar como mucho 2 alumnos de inglés, $p(X \leq 2)$.

Es decir, $p(X \geq 3) = 1 - p(X < 3) = 1 - p(X \leq 2) = 1 - (p(X = 0)+p(X = 1)+p(X = 2))$

y sólo tenemos que calcular 3 probabilidades: $p(X = 0) \approx 0$, $p(X=1) = 0'000001$, $p(X=2) = 0'000026$ (!compruébalo!).

Por lo cual, $p(X \geq 3) = 1 - (0 + 0'000001 + 0'000026) = 1 - 0'000027 = 0'999973$

b) $p(X=15) = 0'0025$ (aplica la fórmula).

c) $p(7 \leq X \leq 10) = p(X = 7)+p(X = 8)+p(X = 9)+p(X = 10) =$

$= 0'0549 + 0'1114 + 0'1759 + 0'2142 = 0'5564$.

Media y desviación típica en una distribución binomial

Nota.- Aunque no se demostrará, en una distribución binomial B(n;p), el numero esperado de éxitos o **media**, viene dado por $\bar{x}$ **= n.p**. (Recordemos que la media es una medida de centralización).

La **varianza** es $\boldsymbol{\sigma^2 = (n.p.q)}$, y la **desviación típica**, $\boldsymbol{\sigma}$, s la raíz cuadra de la varianza, es decir $\boldsymbol{\sigma = \sqrt{(n.p.q)}}$.

Ejercicios

♦♦ En una B(7;0'4) determina: p(X=2), P(X=5), P(X=0), P(X>0), P(X>5). Determina su media y desviación típica.

♦♦ Lanzamos un dado 5 veces seguidas. Calcula la probabilidad de obtener:

a. Más de 3 unos.
b. Ningún uno.
c. Determina su media y desviación típica.

Sol (n = 5, éxito= obtener 1 al lanzar un dado., P (éxito) = 1/6)

♦♦ La probabilidad de que un cierto producto se rompa cuando es transportado es del 2%. Si se transportan 20 de éstos, calcula la probabilidad de que:

a. Se rompan más de 2.
b. No se rompa ninguno.
c. Determina su media y desviación típica.

Sol (n= 20, éxito= que se rompa un producto.; P (éxito) = 0,02 = 2% = 2/100 = 0,02)

♦♦ un examen tipo test tiene 10 preguntas, cada una de ellas con 3 opciones para elegir. Si un alumno contesta al azar; calcula la probabilidad de que:

a. Acierte más de 8 preguntas.
b. No acierte ninguna.
c. Acierte todas.
d. Determina su media y desviación típica.

Sol (n=10, éxito= acertar la pregunta. P (éxito) = 1/3)

♦♦ El 53% de los trabajadores de una empresa son mujeres. Si elegimos 8 personas de esa empresa al azar. Calcula la probabilidad de que haya:

a. Alguna mujer.
b. Más de 6 mujeres.
c. Determina su media y desviación típica.

Sol (n=8, éxito= obtener mujeres. P(éxito)= 0,53 ; fracaso=0,47)

Distribuciones Continuas. Distribución Normal

Introducción

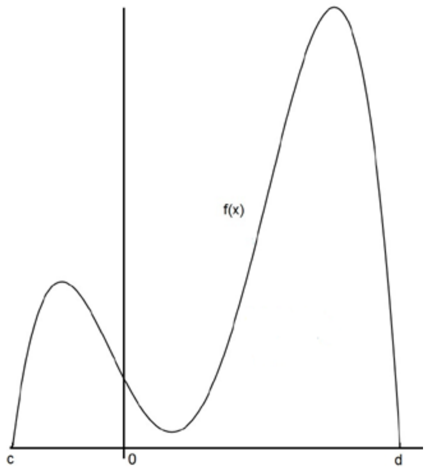
Ya hemos visto que la las variables aleatorias continuas X en teoría tomaban todos los posibles valores de un intervalo, que puede ser todo R.

En las variables aleatorias discretas teníamos el concepto de **función de probabilidad**, que verificaba las propiedades:

1) $p_i > 0$, con $p(X=x_i) = p_i$

2) $p_1 + p_2 + p_3 + \ldots\ldots + p_n = 1$

Sabemos que si una función f(x) es continua si su gráfica se dibuja sin levantar el lápiz del papel. Por ejemplo



El concepto de el concepto de **función de probabilidad**, en las variables aleatorias discretas, lo hará ahora la función f(x) que se llamará **función de densidad**, y las propiedades:

1) $p_i > 0$, con $p(X=x_i) = p_i$ de la discreta, se transformará en que $f(x) \geq 0$, es decir esté por encima del eje OX.

2) $p_1 + p_2 + p_3 + \ldots\ldots + p_n = 1$ de la discreta, se transformará en que el área bajo f(x), entre

$-\infty$ y $+\infty$ (en el dibujo es entre “c” y “d”) y el eje OX, valga 1.

Si se conoce el cálculo integral esto se expresa como $\int_{-\infty}^{+\infty} f(x)dx = \int_a^b f(x)dx = 1$.

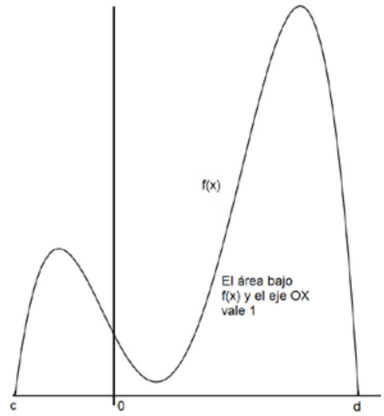
Función de densidad de una variable aleatoria continua.

Def.- Una función f(x) decimos que es una **función de densidad** de una variable aleatoria continua X si verifica:

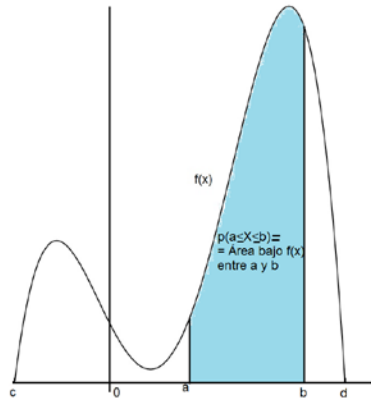
1º $f(x) \geq 0$, para todo punto x del intervalo en que está definida, es decir siempre es positiva (recordar que en las variables discretas $p_i \geq 0$)

2º El área encerrada bajo la curva y = f(x) y el eje OX es igual a la unidad, es decir

$\int_{-\infty}^{+\infty} f(x)dx = 1$. (recordar que en las variables discretas $p_1 + p_2 + .. + p_n = \sum_{i=1}^{n} p_i = 1$).



3º La probabilidad de f(x) en un intervalo [a,b] es el área bajo f(x), el eje OX y las abscisas x=a y x=b, es decir Área = $p(a \le X \le b) = \int_a^b f(x)dx$.



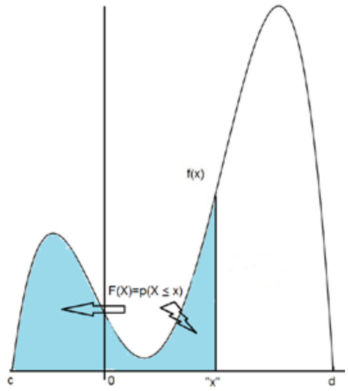
Propiedades

1.- No existe la probabilidad de un punto x= a, es decir p(X=a) = 0 = $\int_a^a f(x)dx = 0$

2.- Por la propiedad anterior las siguientes probabilidades son iguales
$p(a \le X \le b) = p(a < X \le b) = p(a \le X < b) = p(a < X < b)$.

Función de distribución de una variable aleatoria continua

Def.- La **función de distribución F(x)** de una función de densidad f(x) de una variable aleatoria continua (v.a.c.) es la función F(x) que verifica la siguiente condición de probabilidad F(X) = p(X ≤ x), es decir el área bajo f(x) entre -∞ y "x", y el eje OX.(si se conoce el cálculo integral F(X) = p(X ≤ x) = $\int_{-\infty}^{x} f(t)dt$), que es lo que se conoce como **función área**.



Si se conoce el cálculo integral la función la función de distribución F(x)se obtiene integrando la función de densidad f(x).

Función de densidad f(x) **Función de distribución F(x)**

$$f(x)=\begin{cases} 0 & \text{si} \quad x<0 \\ \frac{1}{3} & \text{si} \quad 0 \le x \le 3 \\ 0 & \text{si} \quad x>3 \end{cases} \qquad \text{Integrando} \qquad F(x)=\begin{cases} 0 & \text{si} \quad x<0 \\ \frac{x}{3} & \text{si} \quad 0 \le x \le 3 \\ 1 & \text{si} \quad x>3 \end{cases}$$

Propiedades de las funciones de distribución:

1. Puesto que F(x) es el valor de una probabilidad, se verifica: $0 \le F(x) \le 1$

2. La función de distribución F(x) es nula para todo valor de x anterior al menor valor de la variable aleatoria, es decir F(x) = 0, si x < a. (Suponiendo definida f(x) en [a,b])

3. La función de distribución F(x) es igual a la unidad para todo valor de x posterior al mayor de la variable aleatoria, es decir F(X) = 1, si x > a. (Suponiendo definida f(x) en [a,b])

4. La función *F(x) siempre es creciente*. (Se dibuja hacia arriba, de izquierda a derecha)

5.- Las probabilidades se calculan mediante la fórmula **p(a' ≤ X ≤ b') =**

$\int_{a'}^{b'} f(x)dx = [F(x)]_{a'}^{b'} =$ **F(b') – F(a')**, con F'(x) = f(x), si se conoce el cálculo integral

Media de una variable aleatoria continua

Recordemos las expresiones de la **media** de una variable estadística y de una variable aleatoria discreta:

Variable estadística *Variable aleatoria discreta*

$$\overline{x}=\frac{\sum_{i=1}^{n} n_i.x_i}{N} \qquad \mu=\sum_{i=1}^{n} p_i.x_i$$

Nota.- De forma similar a como definimos la media de una variable aleatoria discreta, definiremos ahora la media de una variable aleatoria continua. Únicamente tendremos en cuenta que en aquel caso se trataba de sumar un número finito de sumandos, en cambio ahora tratamos de «sumar» un número infinito de sumandos infinitesimales (integral definida).

Def.- Sea X una variable aleatoria continua cuyo recorrido es el intervalo [a, b] y sea f(x) su función de densidad. Se llama **media** de la variable continua X al valor de la siguiente integral: $\mu=\int_a^b x.f(x)dx$ (si se conoce el cálculo integral), si nó diremos que es un valor sobre el centro de los datos.

Nota.-La media de una variable aleatoria continua también recibe el nombre de **esperanza matemática o valor esperado.**

Varianza y desviación típica de una variable aleatoria continua X

Siguiendo con el paralelismo establecido con las otras variables ya estudiadas, recordemos ahora la expresión de la varianza y desviación típica para el caso de variable estadística y para el caso de variable aleatoria discreta.

Variable estadística

$$\sigma^2=\frac{\sum_{i=1}^{n} n_i.(x_i-\overline{x})^2}{N}=\frac{\sum_{i=1}^{n} n_i.(x_i)^2}{N}-(\overline{x})^2$$

$$\sigma=\sqrt{\sigma^2}$$

Variable aleatoria discreta

$$\sigma^2=\sum_{i=1}^{n} p_i.(x_i-\mu)^2=\sum_{i=1}^{n} p_i.(x_i)^2-(\mu)^2$$

$$\sigma=\sqrt{\sigma^2}$$

Def.- Sea X, una variable aleatoria continua cuyo recorrido es el intervalo [a, b] y sea f(x) su función de densidad. Se llama **varianza de la variable aleatoria continua X** al valor de la siguiente integral: $\sigma_X^2=\int_a^b (x-\mu)^2.f(x)dx=\int_a^b x^2.f(x)dx-(\mu)^2$

Def.- Se llama **desviación típica de la variable aleatoria continua X** a la raíz cuadra a positiva de la varianza, es decir : $\sigma_X=\sqrt{\sigma_X^2}=\sqrt{\int_a^b (x-\mu)^2.f(x)dx}=\sqrt{\int_a^b x^2.f(x)dx-(\mu)^2}$.

Ejercicios resueltos de distribuciones continuas

1. Hallar la media, la varianza y la desviación típica de una variable aleatoria continua que tiene la siguiente función de densidad: $f(x)=\begin{cases} 0 & \text{si} \quad x<0 \\ 1/3 & \text{si} \quad 0\leq x\leq 3 \\ 0 & \text{si} \quad x>3 \end{cases}$

Sol

a) Media: $\mu=\int_0^3 x.f(x)dx=\int_0^3 x.\frac{1}{3}dx=\left[\frac{x^2}{6}\right]_0^3=\frac{9}{6}=1'5$

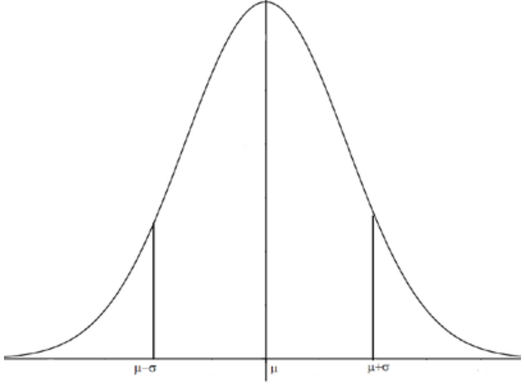
b)Varianza: $\sigma_X^2=\int_0^3 (x-\mu)^2.f(x)dx=\int_0^3 (x-1'5)^2.\frac{1}{3}dx=\left[\frac{(x-1'5)^2}{9}\right]_0^3=0'75$

c) Desviación típica: $\sigma=\sqrt{\sigma_X^2}=\sqrt{0'75}=0'866$

Se supone que se conoce el cálculo integral, para resolver el ejercicio.

Distribución Normal (Como ejemplo de Distribución Contínua)

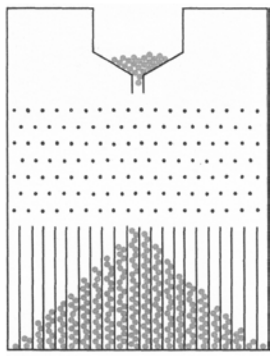
Nota.- La gráfica de una distribución normal tiene forma de campana. Es simétrica y centrada en la media μ, y es de la forma:



Experimento de Galton

Sobre un tablero inclinado están distribuidos regularmente un sistema de clavos. Éstos permiten deslizar un gran número de bolas que proceden de un depósito superior del aparato. Las bolas, al chocar con los clavos, se alejan en mayor o menor medida de la línea central de caída, según la ley del azar.

Recogiendo estas bolas en compartimientos estrechos, distribuidos a lo largo del borde inferior del tablero, las alturas que alcanzan las bolas en las distintas columnas dan una idea, bastante clara, de la distribución normal.



Dispositivo de Galton.

Nota.- Hay muchos fenómenos que siguen distribuciones son normales, por ejemplo: los pesos, las alturas, etc de los individuos de cualquier especie.

Nota.- No todas las distribuciones son normales, por ejemplo: El nivel de renta de los españoles, porque hay pocos con renta muy alta, y muchos con rentas bajas

Definición de la distribución normal de media μ y desviación típica σ .- Decimos que una variable aleatoria continua X **sigue una distribución normal de media μ y desviación típica σ**, y se designa por **N(μ, σ)**, si se cumplen las siguientes condiciones:

1ª La variable recorre toda la recta real, es decir $(-\infty,+\infty) = \Re$.

2ª La función de densidad, que es la expresión en términos de ecuación matemática de la curva de Gauss, es: $f(x)=\frac{1}{\sigma\sqrt{2\pi}}.e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$

donde:

e = 2,7182..., constante: base de los, logaritmos neperianos

π = 3,1415... relación de la longitud de una circunferencia a su diámetro

x = abscisa, valor cualquiera de un punto del intervalo.

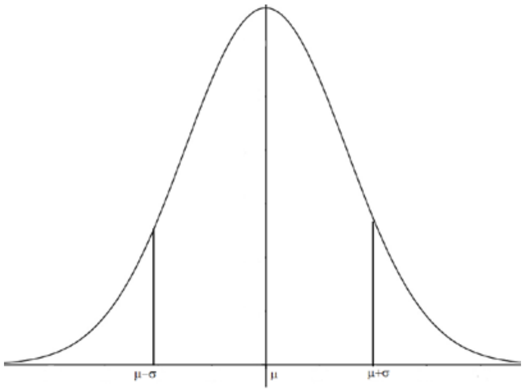
μ = media de la variable aleatoria X (parámetro).

σ = desviación típica de la variable aleatoria X (parámetro).

f(x) = ordenada de la curva.

A los valores **μ** y **σ** se los denomina **parámetros** de la distribución normal.

Nota.- Algo sobre la gráfica de la función de densidad de la **N(μ, σ)**.



1- La curva tiene forma campaniforme y es simétrica respecto a la recta vertical x = μ.

2- La ordenada máxima se obtiene precisamente para el valor de x = μ.

3- El área del recinto encerrado bajo la campana y el eje X es igual a la unidad, es decir $\int_{-\infty}^{+\infty} f(x)dx = 1$, por tratarse de una función de densidad. Al ser simétrica respecto al eje vertical que pasa por μ, dicho eje de simetría deja un área igual a 0'5 a la izquierda y otra igual a 0'5 a la derecha.

4. A ambos lados del valor x = μ las ordenadas de la curva decrecen, primero lentamente, y después con mayor rapidez, hasta hacerse su ordenada nula a medida que la abscisa se aleja hacia la derecha o hacia la izquierda del valor central. En la práctica, no es necesario alejarse mucho de los valores centrales para que la ordenada sea casi nula. (Posteriormente veremos que fuera del intervalo (μ - 3σ, μ + 3 σ) es muy improbable encontrar una observación).

La comprensión de estas propiedades facilitará mucho el manejo de las tablas.

Media y varianza de la distribución normal

Por la propia definición de la distribución se tiene que las características estadísticas de una distribución N(μ, σ) son:

a) Media: **= μ**

b) Varianza: **= σ^2**

c) Desviación típica: **= σ**

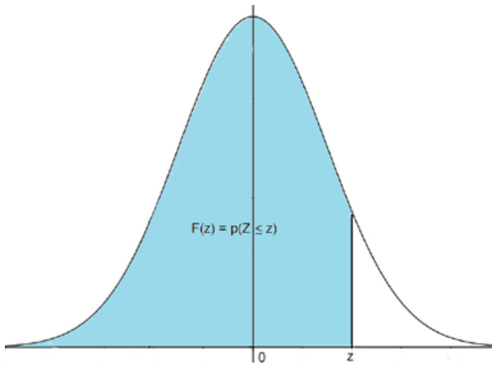
Distribución normal estándar

Def.- De las infinitas distribuciones **N(μ, σ)**, tiene especial interés **la distribución N(0,1);** es decir, aquella que tiene por media el valor cero (**μ** = 0) y por desviación típica la unidad (**σ** = 1). Esta distribución se llama **ley normal estándar,** o bien **distribución normal reducida.**

La función de densidad para μ = 0 y σ = 1 es: $f(x)=\frac{1}{\sqrt{2\pi}}.e^{-\frac{x^2}{2}}$

cuya representación gráfica es parecida a la de la normal N(μ; σ), pero un poco mas achatada.

Nota.- La función de distribución es F(x) = p(X ≤ x), en el caso de la normal, la variable siempre es “z” no “x”, y se escribe F(z) = p(Z ≤ z) = Φ(z). Recordamos que la función de distribución de la ley normal estándar proporciona el área del recinto sombreado de la figura.



Recordamos que la función de distribución es F(x) = p(X ≤ x), en el caso de la normal, la variable siempre es “z” no “x”, y se escribe F(z) = p(Z ≤ z) = Φ(z).

Nota.- Recordamos que las propiedades de la función de distribución eran:

1. Puesto que F(z) es el valor de una probabilidad, se verifica: 0 ≤ F(z) ≤ 1

2. La función de distribución F(z) es nula para todo valor de x anterior al menor valor de la variable aleatoria, es decir F(z) = 0, si z < a. (Suponiendo definida f(z) en [a,b])

3. La función de distribución F(z) es igual a la unidad para todo valor de x posterior al mayor de la variable aleatoria, es decir F(z) = 1, si x > a. (Suponiendo definida f(z) en [a,b])

4. La función *F(z) siempre es creciente*. (Se dibuja hacia arriba, de izquierda a derecha)

5.- Las probabilidades se calculan mediante la fórmula **p(a’ ≤ X ≤ b’) =**

$\int_{a'}^{b'} f(x)dx = [F(x)]_{a'}^{b'} =$ **F(b’) – F(a’)**, con F’(x) = f(x), si se conoce el cálculo integral.

Como el cálculo de esta integral es complicada, se ha realizado una tabla con los valores más usuales de la normal N(0,1), que es lo que se conoce como **tabla de la normal.**

Tipificación de la variable

Nota.- Todas las distribuciones normales N(μ,σ) se pueden transformar en la distribución N(0,1), se encuentra tabulada (tablas), mediante un cambio de variable que se denomina **tipificación de la variable.** Dicho cambio de variables es:

Si X sigue una N(μ,σ), entonces $Z = \frac{X-\mu}{\sigma}$ sigue una N(0,1).

Manejo de tablas

Sea **Z** una variable que sigue una distribución normal **N(0,1)**. Veamos con ejemplos, y en orden creciente de dificultad, los casos más frecuentes que se suelen presentar. (Utilizar la tabla anterior que es la que dan en Selectividad.)

Caso 1. **p(Z ≤ 1,45)** (Se mira directamente en las tablas)

Recuerda que **p(Z ≤ 1,45) = F(z) = Φ(z)**

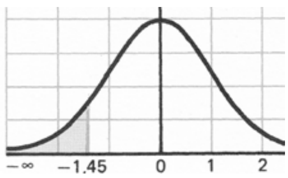
La probabilidad pedida es igual al área sombreada (mira la gráfica de la tabla), y se encuentra directamente en la tabla, sin más que buscar 1,4 en la columna 1ª y 0,05 en la fila 1ª; su intersección nos da la probabilidad

p(Z ≤ 1,45) = 0,9265

Esto quiere decir que el 92,65 % de las observaciones se distribuye entre - ∞ y 1,45.

Caso 2. **p(Z ≤ -1,45)** (Simetría y probabilidad del contrario)

La probabilidad pedida es igual al área sombreada de la figura de abajo.



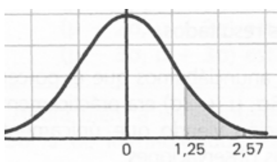
La tabla sólo proporciona probabilidades para valores de Z positivos. Pero teniendo en cuenta la simetría de la función de densidad f(x), y que el área encerrada por toda la curva es igual a la unidad (probabilidad del suceso contrario $p(A^C) = 1 - p(A)$), resulta:

p (Z < -1,45) = p (Z > 1,45) = 1 - p (Z ≤ 1,45) = 1 - 0,9265 = 0,0735

Este resultado indica que únicamente un 7,35 % de las observaciones se distribuye entre -∞ y -1,45.

Caso 3.- **p(1,25 < Z ≤ 2,57)** [F(2'57) – F(1'25), se mira directamente en tabla]

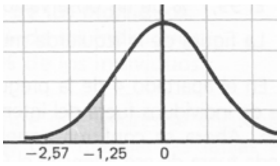
La probabilidad pedida es el área sombreada que muestra la figura de abajo. Su cálculo lo realizaremos restando al área mayor la menor:



p(1,25 < Z - 2,57) = p(Z < 2,57) - p(Z < 1,25) = 0,9949 - 0,8944 = 0,100 5

Caso 4. **p(-2,57 < Z ≤ -1,25)** (Simetría con probabilidad del contrario)

La probabilidad pedida es igual al área sombreada (figura de abajo), y como consecuencia de la simetría de la función de densidad se tiene:



$$p(-2{,}57 < Z \leq -1{,}25) = p(1{,}25 < Z - 2{,}57) = 0{,}1005$$

También puede hacerse por la definición de probabilidad y por la probabilidad del suceso contrario, es decir

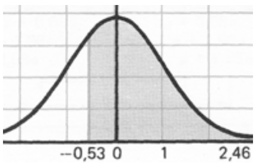
$p(-2{,}57 < Z \leq -1{,}25) = p(Z \leq -1{,}25) – p(Z \leq (-2{,}57) = [1- p(Z \leq 1{,}25)] - [1- p(Z \leq 2{,}57)] =$

$= p(Z \leq 2{,}57) - p(Z \leq 1{,}25) = p(1{,}25 < Z - 2{,}57) = 0{,}1005$

Es decir, únicamente el 10’05 % de las observaciones se distribuye entre 1,25 y 2,57, o bien entre -2’57 y -1’25.

Caso 5. **p(-0,53 < Z ≤ 2,46)**

La probabilidad pedida es igual al área sombreada (figura de abajo). En este caso, uno de los valores de la abscisa Z es negativo y el otro es positivo.



Teniendo en cuenta todo lo anterior, resulta:

$p(-0{,}53 < Z \leq 2{,}46) = p(Z \leq 2{,}46) - p(Z \leq -0{,}53) = p(Z \leq 2{,}46) - p(Z > 0{,} 53) =$

$= p(Z \leq 2{,}46) – [1 - p(Z \leq 0{,} 53)] = 0{,}9931 - (1 - 0{,}7019) = 0{,}695$

Lo que quiere decir que el 69,5 % de las observaciones se encuentra entre -0,53 y 2,46.

Nota.- Cualquier otro caso que se pueda presentar cabe reducirlo, adecuadamente, a los que acabamos de exponer.

Nota.- El 68,26% de las observaciones (o de los individuos) se encuentra en el intervalo $(\mu - \sigma, \mu + \sigma)$.

El 95,4% de las observaciones está en $(\mu - 2\sigma, \mu + 2\sigma)$ y el 99,7% de las observaciones está en $(\mu - 3\sigma, \mu + 3\sigma)$.

Ejercicios resueltos de la normal

♦♦ Los pesos de los individuos de una población se distribuyen normalmente con media 70 kg y desviación típica 6 kg. De una población de 2000 personas, calcular cuántas personas tendrán un peso entre 64 y 76 kg.

Sol

Se trata de una distribución N(70,6). Como el 68,2% de los individuos está en el intervalo (70-6,70+6), habrá que calcular el 68,2% de 2000; es decir:

68,2.(2000/100) = 1364

También se puede hacer calculando la probabilidad p(64 < X < 76) y después multiplicarlo por el nº de personas que es 2000.

Luego se espera que haya 1364 personas con pesos comprendidos entre 64 y 76 kg.

♦♦ Se ha aplicado a 300 alumnos de 4º de ESO un test de agresividad y se ha observado que se distribuyen normalmente con media 30 y desviación típica 12. Se pide:

a) ¿Qué proporción de alumnos tendrá una puntuación en dicho test entre 20 y 35?
b) ¿Cuántos alumnos tendrán una puntuación superior a 42?
Sol

Se trata de una distribución N(30,12). Calculemos las probabilidades pedidas:

a) $p(20 < X \leq 35) = p\left(\frac{20 - 30}{12} < Z \leq \frac{20 - 30}{12}\right) = p(-0'83 < Z \leq 0'42) =$

$= p(Z \leq 0,42) - [1 - p(Z \leq 0,83)] = = 0,6628 - (1 - 0,7967) = 0,4595$

Es decir, aproximadamente el 46% de los alumnos tiene una puntuación de agresividad entre 20 y 35.

b) $p(X > 42) = 1 - p(X \leq 42) = 1 - p\left(Z \leq \frac{42 - 30}{12}\right) = 1 - p(Z < 1) = 1 - 0,8413 = 0,1587$

Es decir, el 15,87% de los individuos tiene puntuaciones superiores a 42. El número de individuos se obtendrá multiplicando el total de alumnos por la proporción; es decir: 300.(15,87/100) = 48 alumnos.

Aproximación de la binomial por la normal

Teorema.- Si X, variable aleatoria discreta, sigue una B(n,p), verificando que **n** es muy grande y que **np** y **nq** ≥ 5, con **q** = 1 –**p**, entonces la variable X se puede aproximar por una nueva variable X' continua que sigue una normal N(np, √(npq)). Esta nueva variable X', después por tipificación $Z = \frac{X' - np}{\sqrt{npq}}$ seguirá una normal N(0;1)

Nota.- Gracias a esta aproximación es fácil hallar probabilidades binomiales pues la probabilidad de la distribución binomial para valores grandes de **n** resulta muy complicada de calcular y, en consecuencia, prácticamente inutilizable.

Nota.- Recordemos que la distribución **binomia**l es de variable aleatoria **discreta X** y, por tanto, tiene sentido calcular probabilidades puntuales; por ejemplo, p(X = 2). En cambio, la variable aleatoria **continua X'** sigue una distribución **normal** y, por tanto, no tiene sentido calcular probabilidades puntuales, pues son todas nulas; en este caso p(X' = 2) = 0.

Para resolver este problema se introducen unas correcciones (*correcciones de Yates*), que consiste en considerar los valores de la variable aleatoria discreta X como marcas de clase de intervalos del siguiente modo:

X sigue una B(n,p) y X' sigue una N(np,√(npq)).

$$p(X = a) = p(a – 0'5 < X' \leq a + 0'5)$$

$$p(X \leq a) = p(X' \leq a - 0'5)$$

$$p(X < a) = p(X' < a - 0{,}5)$$

Ejercicios resueltos de aproximación de una binomial a una normal

♦♦ Se sabe, por una estadística sociológica realizada recientemente, que el *nivel de aceptación* de un determinado partido es del 25% de la población. De un muestreo aleatorio realizado sobre 40 personas, se desea saber cuál es la probabilidad de que 15 de ellos acepten a dicho partido.

Sol

Es obvio que en principio se trata de una distribución binomial:

A = Éxito = «aceptar el partido» $\Rightarrow$ $p = p(A) = 0{,}25$

A^C = «no aceptar el partido» $\Rightarrow$ $q = p(A^C) = 0{,}75$

Sea X la variable que expresa el número de personas de la muestra que aceptan el partido. Se verifica entonces que X sigue una distribución B(40; 0'25).

Como se verifica que: $np = 40.0'25 \geq 5$ y $nq = 40.0'75 \geq 5$, por tanto podemos aproximarlo por una normal, es decir aproximamos la variable X que sigue una binomial B(40; 0'25) por la variable X' que sigue una distribución normal $N(np, \sqrt{(npq)}) =$
$= N(10;2'74)$.

Calculemos la probabilidad pedida:

$p(X=15) = p(14{,}5 < X' \leq 15{,}5)=p\left((14{,}5-10)/ 2{,}74 < Z \leq (15{,}5 -10)/2{,}74 \right) =$

$= p(1{,}64 < Z \leq 2) = p(Z\leq 2) - p(Z\leq 1{,}64) = 0{,}9772 - 0{,}9495 = 0{,}0277$

♦♦ Después de realizar varios sondeos sobre una población con escasa cultura, se ha conseguido averiguar que únicamente el 15% de la misma es favorable a los tratamientos de psicoterapia. Elegida al azar una muestra de 50 personas de dicha población, se desea saber:

a) La probabilidad de que haya más de cinco personas favorables a dichos tratamientos.
b) La probabilidad de que a lo sumo haya seis personas favorables.

Sol

Se trata de una distribución binomial:

A = Éxito = «ser favorable» $\Rightarrow$ $p = p(A) = 0{,}15$

A^C = «ser desfavorable» $\Rightarrow$ $q = p(A^C) = 0{,}85$

Sea X la variable aleatoria que expresa el número de personas de la muestra favorables a los tratamientos de psicoterapia, sigue una distribución B(50;0'15).

Como se cumple que: $np = 50.0'15 \geq 5$ y $nq = 50.0'85 \geq 5$, por tanto podemos aproximarlo por una normal, es decir aproximamos la variable X que sigue una binomial B(50;0'15) por la variable X' que sigue una $N(np, \sqrt{(npq)}) = N(50.0'15; \sqrt{(50.0'15.0'85)}) =$
$= N(7'5;2'52)$.

Calculemos las probabilidades pedidas:

a) $p(X>5)=p(X'>4{,}5)=p(Z>(4{,}5-7{,}5)/ 2{,}52))=p(Z>-1'19)=1- p(Z\leq 1'19) = 1- 0{,}8830 =0{,}1170$

b) $p(X\leq 6)=p(X'\leq 6{,}5)=p(Z\leq(6{,}5-7{,}5)/ 2{,}52))=p(Z\leq - 0{,}4)=1-p(Z\leq 0{,}4)=1- 0{,}6554=0{,}3446$.