DISTRIBUCIONES DE PROBABILDAD

INTRODUCCIÓN

Como ya habíamos mencionado, en el tema de estadística descriptiva, la estadística actúa como disciplina puente entre los modelos matemáticos y los fenómenos reales. El estudio de una determinada característica de una población, se hace tomando una muestra para inferir, posteriormente, los resultados a la población mediante la elaboración de un modelo teórico que se ajuste a esta lo más posible. La probabilidad es el modelo matemático de las frecuencias relativas cuando se pasa del estudio de la muestra al de la población. De igual forma las variables aleatorias serán el modelo matemático de las variables estadísticas, las distribuciones de probabilidad constituyen el modelo matemático de las distribuciones de frecuencia. Y en general, a lo largo de este tema, desarrollaremos un modelo para los principales conceptos estudiados en el tema de estadística descriptiva.

Estudiaremos modelos teóricos para las poblaciones y por tanto las nociones que se tratarán serán de naturaleza teórica, aun cuando los nombres de algunos de estos conceptos coincidan con sus análogos empíricos estudiados anteriormente.

VARIABLE ALEATORIA

Cuando realizamos un experimento aleatorio es usual que no nos interese el resultado completo del mismo sino observar una característica del mismo que puede ser medible numéricamente. Así, al observar las familias de una población, puede interesarnos únicamente conocer el número de hijos varones o únicamente la edad de la esposa,etc. El concepto de variable aleatoria va a permitirnos asociar a cada suceso del experimento un número real.

Comencemos analizando algunos ejemplos:
1. Consideremos el experimento consistente en lanzar dos monedas al aire. Su espacio muestral será: E={ (x,x), (x,c), (c ,x), (c,c)} .
Supongamos ahora que solo nos interesa estudiar la variable X={número de caras obtenidas}, los resultados que se podrían dar toman los valores {0,1,2}, asociando a cada elemento del espacio muestral un número: el de caras de ese suceso.

X: E--------------------→ R
(c,c)------------------→ 2
(c,x)------------------→ 1
(x,c)------------------→ 1
(x,x)------------------→ 0

A esta aplicación le llamaremos variable aleatoria asociada al experimento.
NOTA:Es evidente que a un mismo experimento pueden asociarse distintas variables aleatorias, según lo que estemos estudiando. En el ejemplo anterior si hubiésemos estudiado el número de cruces la aplicación, es decir la variable aleatoria, variaría.

2.- Consideremos el experimento consistente en lanzar dos dados, cuyo espacio muestral viene dado por: E= {(1,1),(1,2),..............,(6,1),........,(6,6) }. La ley que asocia a cada suceso de dicho espacio la suma de los puntos obtenidos es una variable aleatoria que toma los valores:X={2,3,4,5,6,7,8,9,10,11,12}.
La ley que asocia a cada suceso el mayor de los números obtenidos, es otra variable aleatoria, asociada al mismo experimento, que toma valores Y={1,2,3,4,5,6 }.

3.- Consideremos el experimento que consiste en observar si las piezas fabricadas, en una determinada fábrica, son o no defectuosas. Su espacio muestral es E={ D, $\overline{D}$ }. En este caso la observación que realizamos es cualitativa y no cuantitativa, pero aún así podemos asociar a cada uno de estos resultados un número, por ejemplo: 0 si es defectuosa y 1 si es no defectuosa, dando lugar a una variable aleatoria que toma los valores X={0,1}

$$
\begin{aligned}
X: E &\text{------------------}\rightarrow R\\
D &\text{------------------}\rightarrow 0\\
\overline{D} &\text{------------------}\rightarrow 1
\end{aligned}
$$

Vemos por lo tanto, que en todo experimento podemos asociar a cada uno de los resultados obtenidos un número real. A esta asociación o aplicación se le llama variable aleatoria.

<u>Definición:</u>

Dado un experimento aleatorio S cuyo espacio muestral es E, llamaremos variable aleatoria asociada a dicho experimento a toda aplicación X: E------------→ R que asocia a cada suceso del espacio muestral un número real.

Por abuso del lenguaje es frecuente confundir la variable aleatoria como ley o función con el conjunto de valores que toma y así diremos: dada la variable X={ $x_1,x_2,.....x_n$} en lugar de decir la variable X que asocia a los sucesos {$S_1,S_2,.....S_n$} valores {$x_1,x_2,....x_n$} respectivamente

La variable aleatoria es el modelo matemático de la variable estadística cuando se pasa del estudio de una muestra al de la población. Así "al lanzar tres veces dos monedas al aire" y observar el número de caras obtenidas podemos obtener como resultados de nuestro experimento los sucesos {(c,c),(c,x),(x,c)} y la variable estadística tomaría por tanto los valores {1,2}; sin embargo la variable aleatoria (asociada a toda la población) toma los valores {0,1,2}.

La variable se llamará discreta si sólo puede tomar unos ciertos valores que no admiten subdivisiones sucesivas. Se llamará continua si puede tomar todos los valores correspondientes a un intervalo de la recta real, o lo que es lo mismo si sus valores admiten subdivisiones sucesivas.

Ejemplos:

1.- La variable que representa el número de caras obtenidas en el lanzamiento de tres monedas al aire toma los valores {0,1,2,3}. Es discreta

2.- La variable que representa el perímetro craneal de una serie de individuos toma los valores [40,100] (medida en cm.). Es continua.

3.- La variable que asocia al lanzamiento de dos dados los números 0 si la suma obtenida es par y 1 si es impar, toma los valores {0,1}. Es discreta.

4. La variable que asocia al lanzamiento de dos dados el menor de los números obtenidos es discreta y toma los valores {1,2,3,4,5,6} Nota: observar que en los dos últimos ejemplos el experimento es el mismo pero la variable aleatoria no.

5. La variable aleatoria X={tiempo invertido por un alumno en ir a su casa desde el instituto} es continua y toma los valores [t_m , t_M] entre los valores mínimo y máximo de tdos los alumnos de dicho instituto.

NOTA: hay variables discretas que pueden tomar infinitos valores distintos pero sin subdivisiones sucesivas. Por ejemplo: lanzamos un dado y observamos el número obtenido, si es un 6 paramos el experimento, si no lo es volvemos a tirar. La variable X={número de veces que hemos lanzado el dado antes de detenernos} es discreta pero su recorrido es todos los números naturales.

Vamos a continuación a hacer el estudio de las variables aleatorias discretas.

FUNCIÓN DE PROBABILIDAD

Consideremos el siguiente ejemplo: lanzamos 240 veces un dado obteniendo los siguientes resultados:

CARA	1	2	3	4	5	6
FRECUENCIA	40	39	42	38	42	39

Construyamos la distribución de frecuencias relativas, a la izquierda, y a la derecha representemos los resultados esperados, mediante el cálculo de probabilidades, si imaginamos que ralizásemos un número infinito de pruebas. La consideración de los resultados posibles y sus probabilidades da lugar a la noción de función de probabilidad.

CARA	f_r
1	0'1667
2	0'1625
3	0'1750
4	0'1583
5	0'1750
6	0'1625

Distribución de frecuencias

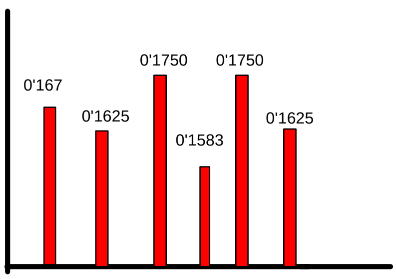
CARA	PROBABILIDAD
1	1/6
2	1/6
3	1/6
4	1/6
5	1/6
6	1/6

Distribución de probabilidad

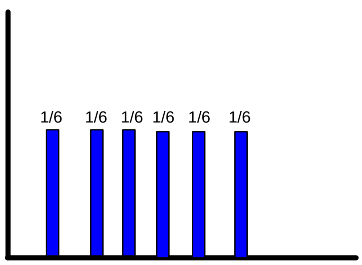
Si nos fijamos en la tabla de la derecha., observamos que a cada valor de la variable le hacemos corresponder su probabilidad. A esa aplicación se le llama función de probabilidad o función de masa de probabilidad o distribución de probabilidad. La distribución de probabilidad es el modelo matemático de la distribución de frecuencias relativas.

Tenemos así los conceptos de muestra, variable estadística, frecuencia relativa y distribución de frecuencias como nociones concretas; de los cuales, por un proceso de abstracción, surgen los conceptos teóricos de población, variable aleatoria, probabilidad y función de probabilidad.

Una función de probabilidad como la anterior, es susceptible de una representación, mediante un diagrama de barras, análoga a la que habíamos hecho para las distribuciones de frecuencias.



Modelo experimental



Modelo teórico

<u>Definición</u>
Sea X una variable aleatoria discreta que toma los valores $\{x_1,x_1,.........,x_n\}$ *. LLamamos función de probabilidad o función de masa de probabilidad asociada a X a la aplicación:*
f: R--------------------------→[0,1] *definida como :* f(x)=P(X=x)

Obsérvese que si $x \notin \text{Im}(X)$, $f(x)= P(X=x)=P(\emptyset)=0$; en caso de que $x=x_i \in \text{Im}(X)$, $f(x_i)=P(X=x_i)$ será la suma de las probabilidades de todos los sucesos elementales cuya imagen por X sea x_i.
Tambien es importante observar que $p(x_1)+p(x_2)+....+p(x_n)=1$ ya que es la probabilidad del suceso seguro
Ejemplo:
Consideremos el experimento consistente en lanzar dos monedas y observemos el número de caras obtenidas. E={(c,c),(c,x),(x,c),(x,x)} y la variable X=" número de caras" toma los valores {0,1,2}
la función de probabilidad vendrá dada por $f(x)=0$ si $x\notin\{0,1,2\}$; $f(0)= P(X=0)=1/4$
$f(1)=P(X=1)=1/2$; $f(2)=P(X=2)=1/4$

Ejercicios (solución al final del tema)
1.- Dado el experimento aleatorio consistente en lanzar dos dados y la variable aleatoria :
X=" suma de los valores de sus caras". Construir su función de probabilidad y representarla gráficamente.
2.- Una ruleta está dividida en 36 sectores iguales. Consideremos la ruleta dividida en tres partes: A que incluye los sectores del 1 al 25, B que incluye del 26 al 35 y C el sector número 36. Si el resultado del juego es un número de A pagas 50 pts., si es de B ganas 100 y si es de C ganas 200pts. Haz la función de probabilidad y el diagrama de barras correspondiente. ¿ Te conviene jugar?
3.- Realizamos el experimento consistente en lanzar dos dados y anotar el mayor de los dos números que han salido. Estudia la variable aleatoria a la que da lugar, su función de probabilidad y el diagrama de barras asociado.
4.- Una urna tiene 25 bolas de las que: 4 tienen el número 1, 5 el nº 2, 1 el nº 3, 1 el nº 4, 2 el nº 5, 3 el nº 6, 2 el nº 7, 3 el nº 8, y 4 el nº 9. Realizamos el experimento consistente en extraer una bola de la urna y anotar el número que ha salido. Estudia su variable aleatoria y su función de probabilidad
5. Un jugador tira dos veces seguidas a una diana. Sabemos que la probabilidad de que acierte en cada tirada es de 0’6. Halla la función de probabilidad de la variable aleatoria X={ nº de aciertos obtenidos} y represéntala gráficamente.

<u>FUNCIÓN DE DISTRIBUCIÓN</u>

Sea X una variable aleatoria discreta, definida sobre un espacio muestral E, llamamos función de distribución de X a una aplicación que asigna a cada número real x la probabilidad de que la variable tome valores menores o iguales que él.

F : R-------------------------→[0,1]
x-------------------------→$P(X \le x)$

Es el modelo teórico de la distribución de frecuencias relativas acumuladas.

Ejemplo:
Lanzamos un dado y observamos la variable X=" número de la cara obtenida"
la función de probabilidad viene dada por f : R----------------→[0,1] tal que: f(1)=f(2)=f(3)=f(4)=f(5)=f(6)=1/6 ; f(x)=0 $\forall x \notin \{1,2,3,4,5,6\}$

Su función de distribución será: F : R-------------------→[0,1] tal que:

$$F(x)=\begin{cases} 0 \text{ si } x<1 \\ 1/6 \text{ si } 1 \le x < 2 \\ 2/6 \text{ si } 2 \le x < 3 \\ 3/6 \text{ si} 3 \le x < 4 \\ 4/6 \text{ si } 4 \le x < 5 \\ 5/6 \text{ si } x \le 5 < 6 \\ 1 \text{ si } x \ge 6 \end{cases} \quad ; F(x)=P(X\le x)$$

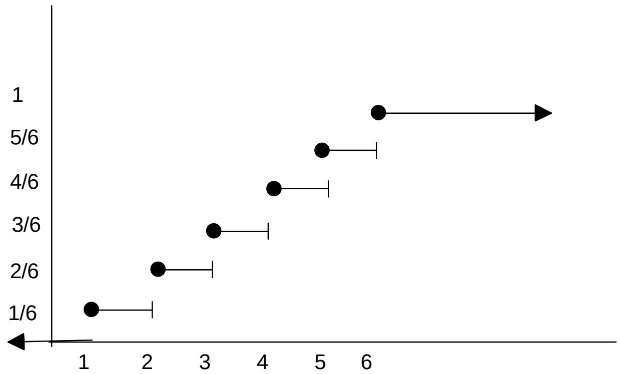
Obsérvese que si queremos hallar, por ejemplo $P(X > 2) = 1-P(X \le 2)=1-F(2)=1-2/6=4/6$

Ejercicio: Hallar la función de distribución de la variable asociada al experimento de lanzar dos dados y anotar la suma de las caras obtenidas

<u>Propiedades</u>

1. Es una función definida positiva ($F(x) \ge 0 \ \forall x \in R$)
2. Es creciente ($\forall x, y \in R \ \ x \le y \Rightarrow F(x) \le F(y)$)
3. La gráfica de F es escalonada, toma valores constantes en los intervalos : $(-\infty,x_1)$, $[x_1,x_2),[x_2,x_3),.......[x_n,\infty)$
4. $\lim_{x\to -\infty} F(x)=0$; $\lim_{x\to \infty} F(x)=1$

Ejemplo: Gráfica de la función de distribución considerada en el ejemplo anterior



CARACTERÍSTICAS DE UNA VARIABLE ALEATORIA DISCRETA

De igual forma que lo hacíamos en las variables estadísticas, vamos a definir para las variables aleatorias unos parámetros o medidas que sirvan de promedios y de medidas de dispersión. Estas medidas características serán: Esperanza matemática (E(X)); varianza (σ^2) y desviación típica (σ)

ESPERANZA MATEMÁTICA

La esperanza matemática o valor promedio será el modelo matemático de la media aritmética; se llama tambien media de la variable aleatoria.

Dado que $\bar{x} = \frac{\Sigma x_i f_i}{\Sigma f_i} = \frac{x_1 f_1}{\Sigma f_i} + \frac{x_2 f_2}{\Sigma f_i} + .. + \frac{x_n f_n}{\Sigma f_n} = x_1 f_{r1} + x_2 f_{r2} + .. + x_n f_{rn} = \Sigma x_i\ f_{ri}$

y teniendo en cuenta que el modelo matemático de las frecuencias relativas es la probabilidad, parece lógico definir E(X)= $\Sigma x_i p_i$ siendo $p_i=P(x_i)$

Se llama esperanza matemática de una variable aleatoria X que toma los valores $x_1,x_2,....,x_n$ con probabilidades respectivas $p_1,p_2,.....,p_n$ a la media ponderada de dichos valores, es decir:
$E(X)=x_1p_1+x_2p_2+....+x_np_n=\Sigma x_i\ p_i$

VARIANZA Y DESVIACIÓN TÍPICA

Recordemos que la expresión de la varianza de una variable estadística es:

$$s^2=\frac{\Sigma(x_i-\bar{x})^2 f_i}{\Sigma f_i} = \Sigma(x_i-\bar{x})^2 f_{ri} \qquad \text{o bien} \quad s^2=\Sigma x_i^2 f_{ri} - \bar{x}^2$$

De la misma forma, utilizando probabilidades, definiremos la varianza para una variable aleatoria discreta como:

$$\sigma^2=\Sigma(x_i-E(X))^2 p_i \qquad \text{o bien} \qquad \sigma^2=\Sigma x_i^2 p_i - (E(x))^2$$

LLamaremos desviación típica de una variable aleatoria discreta X a la raiz cuadrada positiva de su varianza $\sigma=\sqrt{\sigma^2} = \sqrt{\Sigma(x_i-E(X))^2 p_i}$

Ejemplo: Consideremos el experimento de lanzar un dado y anotar el número de la cara que sale. Hallar su Esperanza matemática, varianza y desviación típica.

x_i	p_i	x_ip_i	x_i^2	$x_i^2p_i$
1	1/6	1/6	1	1/6
2	1/6	2/6	4	4/6
3	1/6	3/6	9	9/6
4	1/6	4/6	16	16/6
5	1/6	5/6	25	25/6
6	1/6	1	36	36/6
	$\Sigma=1$	$\Sigma=\frac{21}{6}$		$\Sigma=\frac{91}{6}$

$E(X)=\Sigma x_i p_i = \frac{21}{6} = 3'5$; $\sigma^2 = \Sigma x_i^2 p_i - (E(X))^2 = \frac{91}{6} - 3'5^2 = 2'916$;
$\sigma = \sqrt{\sigma^2} = \sqrt{2'916} = 1'707$

Ejercicios (solución al final del tema)

1.- Un miembro del consejo de administración de una empresa ha comprobado, que si bien todos los años tienen una junta, ha habido años que tienen hasta 5. Por la experiencia acumulada durante años, sabe que el número de juntas anuales se distribuye de la siguiente forma:

Nº de juntas	1,00	2,00	3,00	4,00	5,00
probabilidad	2/15	5/15	1/15	3/15	4/15

Se pide:
a) Comprobar que la anterior es una función de probabilidad y representarla gráficamente
b) La función de distribución y su representación
c) La esperanza, varianza y desviación típica
d) La probabilidad de que en un año, elegido al azar, se celebren más de 3 juntas

2.- Calcula la media, desviación típica y varianza de la distribución de probabilidad correspondiente a la suma de las puntuaciones obtenidas al lanzar dos dados. Halla: $P(X < 5)$; $P(X \geq 10)$; $F(4)$; $F(-2)$; $F(19)$

3.- Sacamos dos cartas de una baraja y anotamos el número de ases obtenidos.
a) ¿ Cuál es la función de distribución de probabilidad?
b) ¿ Cuál es la función de distribución?
c) Calcula la media y desviación típica

4.- Una moneda se lanza cuatro veces. Si X denota el número de cara obtenidas, hallar $E(X)$

5.- Se lanza un dado, designemos por X el número doble del que aparezca. Hallar su media y desviación típica.

6.- Sea X una variable aleatoria discreta, cuya función de probabilidad es:

X	0	1	2	3	4	5
P	0'1	0'2	0'1	0'4	0'1	0'1

a) Calcular y representar gráficamente su función de distribución
b) Calcular $P(X < 4'5)$; $p(3 \leq X < 4'5)$
c) Hallar su valor esperado, varianza y desviación típica.

Pasaremos a continuación al estudio de las variables aleatorias continuas.

DISTRIBUCIÓN DE PROBABILIDAD CONTINUA

Consideremos el siguiente ejemplo:
Tratamos de pesar, una tras otra, todas las naranjas puestas a la venta en un almacen, y supongamos que disponemos de una balanza con precisión de 2 en 2 Dg., esto quiere decir que las medidas que obtendremos serán del tipo;10 Dg, 12 Dg etc., medidas que no admiten subdivisiones sucesivas; sin embargo sabemos que la variable peso no es discreta sino continua y por tanto no admite únicamente como valores los números reales de los tipos propuestos.

La única razón de poder obtener sólo un número finito de valores radica en la limitación del instrumento de medida, de ahí que si en la balanza nos da un valor cualquiera por ejemplo entre 10 y 12 Dg atribuimos a dicha naranja un peso en el intervalo [10,12], no conociendo su valor exacto. (No es posible conocer el valor exacto de una variable continua, de ahí que medirla consista en clasificarla dentro de un intervalo. los modelos descriptivos de variables aleatorias continuas se basan en este principio)

Una vez pesadas todas las naranjas del almacen podemos calcular la proporción de naranjas con peso dentro de cada intervalo de longitud 2 Dg , y levantar rectángulos con base igual a dicha amplitud y con área igual a la correspondiente proporción de naranjas. Así construimos el histograma de frecuencias relativas.(figura 1)

Supongamos ahora que compramos una balanza de mayor precisión, que mida de un en un Dg, con ella aumenta la fiabilidad de las medidas que tomamos, no obstante, aunque la variable peso de forma teórica es continua, los datos que obtenemos siguen siendo los de una variable discreta. Construimos el nuevo histograma de frecuencias relativas, en el cual las bases serán la mitad de las anteriores, y la frecuencia relativa de cada intervalo se distribuirá ahora en los 2 subintervalos a los que da lugar.(figura 2)

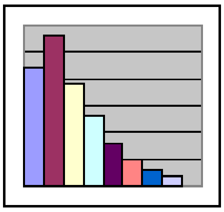
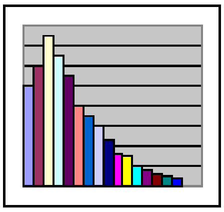
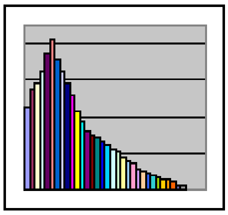
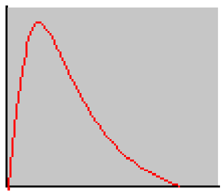





Figura 1   Figura 2   Figura3   Figura4

Podemos seguir repitiendo el experimento utilizando balanzas cada vez más precisas y seguir construyendo los histogramas correspondientes, compuestos por rectángulos cada vez más estrechos (figura 3). Sin embargo, llegará un momento en que no podremos disponer físicamente de una balanza más precisa que la última utilizada. Esto significa que la V.A. peso, aunque teóricamente continua, seguirá siendo físicamente discreta debido a los instrumentos de medida.

Podemos concebir idealmente un aumento indefinido del número de naranjas y del número de señales de la balanza, cada vez más próximas entre sí (estamos pasando de la muestra a la población, lo que nos exije un proceso de abstracción), los correspondientes histogramas se irían adelgazando más y más, puesto que la longitud de los intervalos tendería a cero, hasta que la línea quebrada escalonada se transformaría en una linea continua que describirá el comportamiento de la variable estimada aplicado a toda la población (en nuestro caso distribución del peso de las naranjas) (Figura 4)

De esta forma pasamos del histograma de frecuencias, que es la representación de una distribución de frecuencias continua (modelo experimental), a la curva de probabilidad, que será la representación de una distribución de probabilidad continua (modelo teórico). Llamaremos función de densidad a esa función límite.

Analicemos el histograma para ver las propiedades que tendrá la función de densidad:

a) La linea quebrada no tiene punto alguno debajo del eje OX, por lo tanto, en el modelo teórico, f(x) $\geq 0 \; \forall x \in R$

b) El área total bajo el histograma vale 1 (por ser un histograma de frecuencias relativas). Por tanto en el modelo teórico el área total bajo la curva tambien ha de ser 1; esto es:

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

c)Tanto en el histograma como en el caso de la curva, el eje OX representa la variable. Para el histograma, la frecuencia de que la variable tome un valor comprendido entre dos números a y b, viene representada por el área de los rectángulos entre a y b. Por lo tanto para la curva y=f(x), la probabilidad de que la variable tome un valor comprendido entre a y b tendrá que venir tambien dada por el área encerrada por dicha función y el eje OX entre a y b ; es decir: P(a $\leq x \leq$ b) = $\int_a^b f(x)dx$. El conocimiento de la función de densidad nos permite hallar probabilidades por integración. (Ver figuras en la página anterior)

Definición

Sea X una variable aleatoria continua. Una función f: R $-----------\rightarrow R$ *es una función de densidad de X si y solo si:*

1) f(x) $\geq 0 \forall x \in$R

2) $\int_{-\infty}^{\infty} f(x)dx = 1$

La función de densidad permite definir la probabilidad en cada intervalo mediante el área correspondiente P(a $\leq x \leq$ b) = $\int_a^b f(x)$ dx

Observaciones

1.- Ya hemos dicho que la probabilidad de que un modelo de variable continua asigne a la observación un valor exacto (es decir medido con infinita precisión) es cero. Es decir, no tiene sentido hablar de la probabilidad de un valor puntual. Utilizando la función de densidad observamos que P(X=a) = $\int_a^a f(x)$ dx =0 (representaría el área de una linea). Como consecuencia de ello P(a$\leq x \leq b) = P(a < x \leq b) = P(a \leq x < b)$ = P(a < x < b)

2.- f(x) representa la densidad de probabilidad en cada punto, no la probabilidad; de ahí que reciba el nombre de función de densidad y no de función de probabilidad como en el caso discreto. (P(X=a)=0 y sin embargo f(a) puede ser distinto de cero, luego f(x) no representa la probabilidad)

Ejemplos:

1.- Sea X una variable aleatoria continua con función de densidad definida por:

$$f(x)=\begin{cases} 1/6\,x+k \text{ si } 0 \leq x \leq 3 \\ 0 \text{ para todos los demás valores de x} \end{cases}$$

Se pide: a) Calcular k ; b) Hallar P(1< x < 2)

a) Si f(x) es una función de densidad ha de verificar que $\int_{-\infty}^{\infty} f(x)dx$ =1. En nuestro caso

$$1= \int_{-\infty}^{\infty} f(x)dx = \int_0^3 (1/6x + k)\,dx = 1/12x^2+kx]_0^3 = 3/4+3k=1 \text{ de donde } k=1/12$$

b) $P(1 < x < 2)= \int_1^2 (1/6x+1/12)dx=1/12x^2+1/12x]_1^2=1/2-1/6=1/3$

2.- Supongamos que la variable continua X tiene por función de densidad f(x)= $|x|e^{-x^2}$

a) comprobar que f(x) es una función de densidad

b) ¿cuál es la probabilidad de que la variable tome un valor en el intervalo (0,1)?

A) $f(x)=\begin{cases} xe^{-x^2} \text{ si } x \geq 0 \\ -xe^{-x^2} \text{ si } x<0 \end{cases}$ por lo que $f(x)\geq 0 \forall x \in R$

$$\int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{0} -xe^{-x^2}dx + \int_0^{\infty} xe^{-x^2}dx = (1/2\ e^{-x^2}]_{-\infty}^{0} + (-1/2e^{-x^2})]_0^{\infty} =$$

$$=1/2- \lim_{x\to -\infty} 1/2e^{-x^2} + (\lim_{x\to \infty} -1/2e^{-x^2})+1/2 =1$$

b) $P(0 < x < 1)=\int_0^1 xe^{-x^2}dx = 1/2(1-e^{-1})$

FUNCIÓN DE DISTRIBUCIÓN

Del mismo modo que del histograma de frecuencias relativas se pasa a la función de densidad , pasaremos del diagrama acumulativo de frecuencias relativas a la función de distribución

Sea X una variable aleatoria continua con función de densidad y=f(x). Definimos su función de distribución como: $F\ : R \longrightarrow R$

$x \longrightarrow P(X \leq x)$ *por tanto* $F(x)= \int_{-\infty}^{x} f(t)\,dt$

Es la generalización al caso continuo del polígono de frecuencias relativas acumuladas cuando la longitud del intervalo tiende a cero. Como en el caso discreto proporciona la probabilidad acumulada hasta un determinado valor de la variable

Al igual que en el caso discreto la función de distribución verifica las siguientes propiedades:

1. $F(x)\geq 0\ \forall x \in R$
2. F(x) es creciente en R
3. $\lim_{x\to \infty} F(x) =1$ y $\lim_{x\to -\infty} F(x) =0$

Además si f es continua F ' (x) =f(x) (por el teorema fundamental del cálculo integral)

Ejemplos:

1.- Sea X una variable aleatoria continua con la siguiente función de densidad (comprueba que lo es)

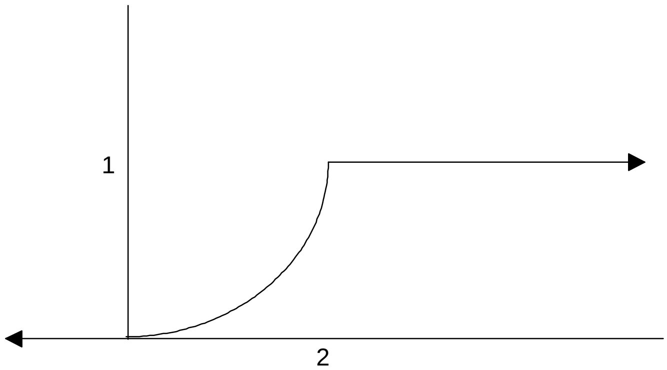
$$f(x)=\begin{cases} 1/2x \text{ si } 0 \leq x \leq 2 \\ 0 \text{ en cualquier otro caso} \end{cases}$$

Hallar su función de distribución y representarla gráficamente.

SOLUCIÓN

$$F(x)=\begin{cases} 0 \text{ si } x<0 \\ \int_0^x 1/2tdt = 1/4x^2 \text{ si } 0 \le x \le 2 \\ 1 \text{ si } x>2 \end{cases}$$

Gráfica de F(x)



2.-La función de densidad de una V. A. continua X es:

$$f(x)=\begin{cases} 0 \text{ si } x<0 \\ \text{Asenx si} 0 \le x \le \pi \\ 0 \text{ si } x>\pi \end{cases}$$

a) ¿ Cuál es el valor de A ?; b) Hallar su función de distribución.

a) $1=\int_{-\infty}^{\infty} f(x)dx = \int_0^{\pi} Asenxdx = -\text{Acosx}]_0^{\pi}=2A$, por lo tanto 2A=1; A=1/2

b) $$F(x)=\begin{cases} 0 \text{ si } x<0 \\ \int_0^x \frac{1}{2}\text{sentdt}=\frac{1}{2}(1-\cos x) \text{ si } 0 \le x \le \pi \\ 1 \text{ si } x>\pi \end{cases}$$

3. Sea X una variable aleatoria continua cuya función de distribución viene dada por:

$$F(x)=\begin{cases} 0 \text{ si } x \le \pi \\ \cos x \text{ si } \pi < x < 3\pi/2 \\ 1 \text{ si } x \ge 3\pi/2 \end{cases}$$

a)Halla su función de densidad y comprueba que lo es. b) Calcula P($\pi/4 < X < 5\pi/4$)

a) Por el teorema fundamental del cálculo $f(x)=F'(x)=\begin{cases} 0 \text{ si } x \le \pi \\ -\text{senx si } \pi<x<3\pi/2 \\ 0 \text{ si } x>3\pi/2 \end{cases}$. Comprobamos ahora que es una función de densidad: 1) $f(x)\ge 0 \forall x \in R$; 2) $\int_{-\infty}^{\infty} f(x)dx = \int_{\pi}^{3\pi/2} -senxdx = \cos x]_{\pi}^{3\pi/2} = \cos(3\pi/2) - (\cos\pi) = -(-1) = 1$ Por tanto es una función de densidad

b) $P(\pi/4 < X < 5\pi/4) = \int_{\pi/4}^{5\pi/4} f(x)dx = \int_{\pi}^{5\pi/4} -senxdx = \cos x]_{\pi}^{5\pi/4} = \cos(5\pi/4) - \cos(\pi) = -\frac{\sqrt{2}}{2} + 1$

COMPARACIÓN DEL CASO DISCRETO Y CONTINUO

Siendo f(x) la función de probabilidad de una variable aleatoria discreta y f(x) la función de densidad de una v.a. continua, nos encontramos con que ambas son las funciones que acumuladas, en forma de sumatorio en el caso discreto y de integral en el caso continuo, dan como resultado la función de distribución. Hay sin embargo diferencias notables entre ellas y,en consecuencia, entre el cálculo de probabilidades para las variables discretas y continuas:

1. V.A.D. f(x) representa una probabilidad, P(X=x) y por tanto $f(x) \leq 1 \forall x \in R$; en el caso de la V.A.C. f(x) no representa una probabilidad y por tanto puede ser >1
2. V.A.D. $P(X=x) \geq 0 \forall x \in R$; V.A.C $P(X=x)=0 \quad \forall x \in R$
3. En el caso discreto nos valemos de puntos para definir la probabilidad; P(X=x)=f(x). En el caso continuo la probabilidad hallarse en intervalos, debido a ser nula la probabilidad en cualquier punto, y dicha probabilidad viene dada por el área limitada por la función de densidad en ese intervalo $P(a<x<b)=\int_a^b f(x)\, dx$
4. En el caso discreto cualquier probabilidad es la suma de probabilidades puntuales, en el caso continuo cualquier probabilidad es una integral definida asociada a un intervalo.

CARACTERÍSTICAS DE UNA VARIABLE ALEATORIA CONTINUA

Análogamente a como se hizo para variables estadísticas y para las aleatorias discretas, se definen para una variable aleatoria continua los siguientes parámetros:

ESPERANZA MATEMÁTICA

Recordemos que: $\bar{x} = \Sigma x_i f_{ri}$ (V. estadística) y $E(X)=\Sigma x_i p_i$ (V.A. discreta)

Si tenemos en cuenta que en el caso discreto estamos sumando un número finito de sumandos y en el continuo estamos "sumando" un número infinito de sumandos infinitesimales (la longitud de los intervalos tiende a cero), y teniendo en cuenta también la definición de integral definida, se sigue que: $E(X)=\int_{-\infty}^{\infty} x f(x)\, dx$ (V.A. continua)

VARIANZA Y DESVIACIÓN TÍPICA

Siguiendo con el paralelismo entre V. estadísticas, V. aleatorias discretas y V. aleatorias continuas, se sigue que:

$S^2=\Sigma(x_i - \bar{x})^2 f_{ri}$ (V. estadísticas); $\sigma^2 = \Sigma(x_i - E(X))^2 p_i$ (V.A. discretas)

y por tanto $\sigma^2 = \int_{-\infty}^{\infty}(x-E(X))^2 f(x)dx$ (V.A: continua)

En cuanto a la desviación típica será como en todos los casos: $\sigma = \sqrt{\sigma^2}$

NOTAS:

-La mediana será la solución de la ecuación F(x)=1/2; es decir la solución de la ecuación $\int_{-\infty}^{x} f(t)dt = 1/2$

- La moda se definirá como el valor de la variable al que corresponde un máximo en la función de densidad y=f(x) y por tanto, si esta es derivable, la moda es una de las soluciones de la ecuación f ' (x) =0

Ejemplos:

1.- Sea X una variable aleatoria continua cuya función de densidad viene dada por:

$$f(x)=\begin{cases} 0 \text{ si } x \le 0 \text{ ó } x \ge \pi \\ 1/2\,senx \text{ si } 0<x<\pi \end{cases}$$

Hallar su esperanza, desviación típica y varianza.

$$E(x)=\int_{-\infty}^{\infty} xf(x)dx=\int_0^{\pi} x\tfrac{1}{2}senxdx=\tfrac{1}{2}(-x\cos x+\text{senx}\,]_0^{\pi}=\frac{\pi}{2}$$

$$\sigma^2=\int_{-\infty}^{\infty}(x-E(x))^2 f(x)dx=\int_0^{\pi}(x-\tfrac{\pi}{2})^2\tfrac{1}{2}senxdx=\tfrac{1}{2}\int_0^{\pi}(x^2senx-\pi xsenx+\frac{\pi^2}{4}senx)dx$$

$$=1/2(2xsenx-x^2cosx+2cosx-\pi senx+\pi xcosx-\pi^2/4cosx]_0^{\pi}=(\pi^2/4)-2$$

$$\sigma=\sqrt{\frac{\pi^2}{4}-2}$$

2.-Sea $f(x)=\begin{cases} e^{-x} \text{ si } x>0 \\ 0 \text{ si } x \le 0 \end{cases}$ Se pide su función de distribución, esperanza y mediana.

$$F(x)=\begin{cases} 0 \text{ si } x \le 0 \\ \int_{-\infty}^{x} f(t)dt=\int_0^x e^{-t}dt=[-e^{-t}]_0^x=-e^{-x}+1=1-e^{-x} \text{ si } x>0 \end{cases}$$

$$E(x)=\int_{-\infty}^{\infty} xf(x)dx=\int_0^{\infty} xe^{-x}dx=-xe^{-x}-e^{-x}\;]_0^{\infty}=1$$

Me= solución de la ecuación F(x)=0'5; por lo tanto $1-e^{-Me}=0'5 \Rightarrow Me=-L(0'5)=0'69$

El 50% de la población está por debajo de 0'69; si tenemos en cuenta que la media es 1, la distribución es muy asimétrica (piénsese que en una distribución simétrica la media es igual a la mediana). Para comprobarlo con más claridad veamos cual es la imagen de la media en la función de distribución: $F(1)=1-e^{-1}=0'63$; esto significa que por debajo de la media se encuentra el 63% de la población.

EJERCICIOS (solución al final del tema)

1.- Dada la función de distribución de una variable continua X:

$$F(x)=\begin{cases} 0 \text{ si } x \le 0 \\ \text{senx si } 0<x<\pi/2 \\ 1 \text{ si } x>\pi/2 \end{cases}$$ Hallar su función de densidad.

2.-La variable X tiene por función de densidad: $f(x)=\begin{cases} 0 \text{ si } x<1 \\ 1/8x+a \text{ si } 1 \le x \le 5 \\ 0 \text{ si } x>5 \end{cases}$

a) Calcular a para que sea función de densidad, b) Representar las funciones de densidad y distribución; c) Hallar la media y varianza.

3.- La variable aleatoria X tiene la siguiente función de distribución:

$$F(x)=\begin{cases} 0 \text{ si } x<0 \\ x^2 \text{ si } 0 \le x \le 1 \\ 1 \text{ si } x>1 \end{cases}$$

a) Calcular su función de densidad; b) representar las funciones de densidad y de distribución; c) calcular su media y varianza d) calcular P(1/4 < x < 1/2)

4.-Dada la función: $f(x)=\begin{cases} \dfrac{C}{1+x^2} \text{ si } 0 \le x \le 1 \\ 0 \text{ para cualquier otro valor de x} \end{cases}$

a) Hallar C para que sea una función de densidad; b) Calcular la esperanza de la variable X que tiene a f por función de densidad

5.- Se lanzan un par de dados corrientes. Sea X la variable que denota el menor de los dos números que aparezcan. Hallar las funciones de probabilidad y distribución. la media y desviación típica.

6.-Sea X una variable aleatoria cuya función de probabilidad viene dada por:

P(X=r)=1/8 (r=2,3,4,5,6,7.8.9). Se pide:

a) Representar la función de probabilidad
b) Hallar la función de distribución y representarla
c) Hallar la esperanza y desviación típica
d) Las probabilidades $P(X \geq 6)$; $P(4 < X < 7'5)$; $P(X < -3)$

7.- La función de densidad de una V.A.continua viene dada por:

$$f(x)=\begin{cases} c(1-x^2)^{1/2} \text{ si } |x|<1 \\ 0 \text{ en cualquier otro caso} \end{cases}$$

a)Determinar c; b) Determinar E(X); c) Calcular $P(0 < X < \pi/4)$(* dificultad integral)

8.-Si una V.A.Continua tiene por función de densidad: $f(x)=\begin{cases} 0 \text{ si } x<0 \\ ae^{-ax} \text{ si } x \geq 0 \end{cases}$, siendo a > 0

a) Comprobar que f es función de densidad ; b) ¿ puede ser a cualquier número positivo?
c) Hallar su función de distribución F(x). ¿ Es F ' (x) = f(x) ?
d) determinar E(X) y Var(X)

9.-La función de distribución de una variable aleatoria es:

$$F(x)=\begin{cases} 0 \text{ si } x \leq 0 \\ \dfrac{x^2}{1+x^2} \text{ si } x>0 \end{cases}$$ Hallar su función de densidad