

Variables Aleatorias

Sea el siguiente experimento: “Extraer 2 artículos de una producción de 1.000 artículos entre los cuales hay 50 defectuosos”. Si se controla la calidad de los casos extraídos, el Espacio Muestral será:

$$S=\{BB, BD, DB, DD\}$$

Si en cambio en el mismo experimento, el observador tuviera como objetivo del experimento contar la cantidad de defectuosos, alguien podría plantear al Espacio Muestral de la siguiente manera:

$$S=\{0, 1, 2\}$$

Las diferencias entre los espacios muestrales estarán dadas por los sucesos elementales y sus probabilidades y las mismas son evidentes. Pero un experimento bien definido, tiene un único Espacio Muestral. No podemos cambiar el Espacio Muestral cada vez que deseamos otro objetivo. Vemos que, en el ejemplo planteado, cuando vemos el Espacio Muestral puro, sin introducir un objetivo, cada suceso elemental estará dado por una letra que identifica el estado de la observación, pero en cambio en el segundo caso teniendo en cuenta el objetivo del observador vemos el número que indica la cantidad de defectuosos.

Frecuentemente, cuando experimentamos, lo hacemos para algo, por algo, teniendo un objetivo o varios. Hay experimentos que las observaciones son números y otros no, pero los objetivos de los observadores muchas veces son números. Por ejemplo, el resultado de los siguientes experimentos aleatorios se denota con un número:

- el resultado de lanzar un dado,
- el tiempo que hay que esperar para que se presente una falla en un circuito,
- el número de estaciones de una red de computadoras que requieren la atención del servidor de la red en un momento dado,
- el número de personas en una comunidad que requieren atención médica en un día especificado,
- el peso sumado de las personas que están en un elevador en un momento determinado del día,
- la cantidad en dinero transportado en un camión antes de que éste sufra una avería,
- la cantidad de personas que han decidido votar a un determinado candidato, etc.

Al número que identifica el resultado del experimento o bien al objetivo del observador en la experimentación se lo llama **variable aleatoria**.

Estas variables aleatorias no son variables en el sentido usual, como las que estamos acostumbrados a manejar en matemática, que toman valores o varían en forma predecible a partir de leyes. El valor de una variable aleatoria no se puede conocer con exactitud de antemano, sino luego de la realización del experimento. El azar juega un papel importante en la medición de la variable aleatoria y su valor no es predecible.

Las variables aleatorias son una función del experimento, o sea son una función del Espacio Muestral, y llevan en su definición el objetivo que tiene el observador en la experimentación. Cuando hablamos de función, lo hacemos en el sentido estricto de la palabra y con las características de una función: para cada suceso elemental del Espacio Muestral existe uno y sólo

un número que lo vincula con la variable. Al conjunto de todos los valores que una variable aleatoria puede tomar se lo llama Recorrido.

Las variables aleatorias se clasifican usualmente en discretas y en continuas, de acuerdo con el número de valores que pueden asumir:

- **Discretas:** pueden adoptar un número finito, o una infinidad contable de valores. Es un conjunto discreto de valores posibles. Estas variables son el resultado de **contar**, varían en forma *discreta* (a saltos).
- **Continuas:** pueden adoptar infinitos valores. Los valores posibles se encuentran en cualquier parte de un intervalo. Estas variables son el resultado de **medir**.

¿Cuáles de las variables aleatorias mencionadas arriba son discretas? ¿y cuáles son continuas? Ciertamente el peso de las personas en el elevador no es discreto, pero la cantidad de pasajeros en el mismo elevador sí lo es.

Una variable aleatoria *nos interesa aún cuando no tiene un valor específico*. Denotamos con letras mayúsculas a las variables aleatorias ($X, Y, Z, \dots$) y con minúsculas a los valores posibles que contemplamos para ellas.

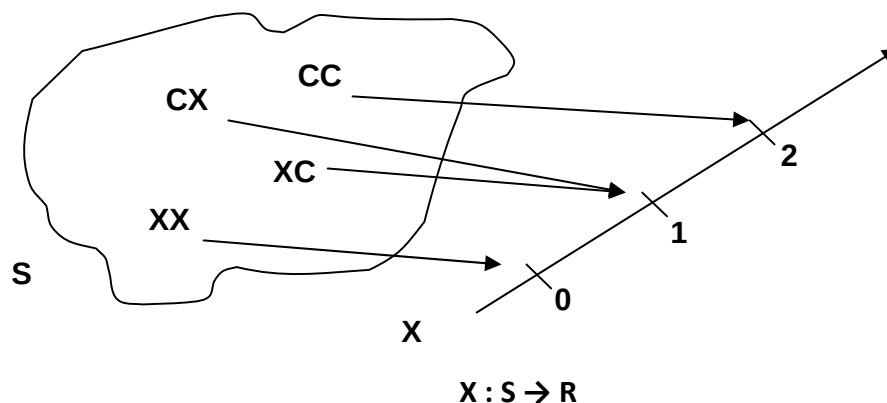
Ejemplo: Si se arroja una moneda dos veces y se registra que salió, obtenemos que el Espacio Muestral es:

$$S = \{CC, CX, XC, XX\}$$

Teniendo en cuenta que nuestro objetivo es ver cuántas caras han salido. Definimos a nuestra variable aleatoria X como el número de caras al arrojar dos veces una moneda. Por lo que el recorrido de la variable será:

$$R_x = \{0, 1, 2\}$$

Nuestra variable aleatoria X : cantidad de caras al arrojar una moneda, es una función del Espacio Muestral:



De acuerdo con el Espacio Muestral y las Probabilidades que tiene cada suceso elemental podemos definir la probabilidad de que la variable tome un valor determinado dentro del recorrido. Siendo el recorrido $R_x = \{0, 1, 2\}$ la probabilidad de que la variable tome cada uno de esos valores será:

$$P(X=0) = 1/4$$

$$P(X=1) = 2 \cdot 1/4 = 1/2$$

$$P(X=2) = 1/4$$

Teniendo en cuenta que la variable es un número, podríamos plantear por ejemplo la probabilidad de que la variable sea mayor o igual a 1:

$$P(X \geq 1) = P(X=1) + P(X=2) = 3/4$$

Tener en cuenta que:

$$P(X \geq a) + P(X < a) = P(-\infty < x < \infty) = 1$$

Una variable aleatoria presenta dos características importantes:

1. Un conjunto de valores posibles es el conjunto **imagen** al que llamaremos **recorrido** de la variable aleatoria.
2. Una probabilidad asociada a cada uno de los posibles resultados de la variable, la cual queda expresada mediante lo que llamamos una **función de probabilidad**.
3. Una **Función de distribución** que es el acumulado de la función de probabilidad.

O sea, una variable aleatoria X , tiene un R_x , para cada $x_i \in R_x$ existe una probabilidad asociada:

$$f(x_i) = P(X = x_i) \quad \forall x_i \in R_x$$

También existe una función de Distribución asociada a cada uno de los valores del recorrido:

$$F(x_i) = P(X \leq x_i) \quad \forall x_i \in R_x$$

Si el **recorrido** de X es un conjunto finito o infinito numerable implica que X es una **V.A.D.** (variable aleatoria discreta). Caso contrario será **V.A.C.** (variable aleatoria continua).

Si tomamos como variable aleatoria la suma de los valores resultados de lanzar dos dados, desde el punto de vista de la probabilidad, el número resultante nos interesa antes de que realicemos el experimento. Conocido el resultado, ya no es interesante (al menos, no para la probabilidad).

La imagen de esta variable aleatoria es $R_x = \{ 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 \}$.

Llamando **X** al número resultante de sumar los dos números posibles de salir en el experimento, podemos contemplar de que **X** sea igual a **x**, donde **x** es cualquier número del **recorrido**.

Claro que nos resultará muy interesante saber como es $P(X = x)$ para los diferentes valores de X por eso es importante conocer cuál es el Espacio Muestral y cuál es la probabilidad de cada uno de los sucesos elementales dentro de ese Espacio Muestral.

En nuestro ejemplo de lanzar dos dados el Espacio Muestral es:

$S = \{ (1,1) ; (1,2) ; (1,3) ; (1,4) ; (1,5) ; (1,6) ; (2,1) ; (2,2) ; (2,3) ; (2,4) ; (2,5) ; (2,6) ;$
 $(3,1) ; (3,2) ; (3,3) ; (3,4) ; (3,5) ; (3,6) ; (4,1) ; (4,2) ; (4,3) ; (4,4) ; (4,5) ; (4,6) ;$
 $(5,1) ; (5,2) ; (5,3) ; (5,4) ; (5,5) ; (5,6) ; (6,1) ; (6,2) ; (6,3) ; (6,4) ; (6,5) ; (6,6) \}$

$\#S = 36$

Entonces, por ejemplo, $P(X = 6) = 5/36$ (esto se lee: "la probabilidad de que X sea igual a 6 es 5 sobre 36").

Este valor se corresponde con el experimento, ya que corresponde a la probabilidad del suceso A que es que la suma de los dos dados sea 6.

Dicho suceso es $A = \{ (1,5) ; (5,1) ; (2,4) ; (4,2) ; (3,3) \}$ $\#A = 5$ y $P(A) = 5/36$

Así como $P(X = 8) = 5/36$ pues corresponde al suceso B que la suma de los dos dados sea 8.

$B = \{ (3,5) ; (5,3) ; (2,6) ; (6,2) ; (4,4) \}$ $\#B = 5$ y $P(B) = 5/36$

Siguiendo con el ejemplo de los dos dados, el lanzar dados para ver que número cae no es muy apasionante que digamos, acompañemos los dados con un tablero de oca o de backgammon o

algún otro juego interesante. Si fuéramos a jugar Monopolio, es natural que nos interesen otro tipo de eventos. Podríamos estar interesados en saber

- Cuál es la probabilidad que el resultado sea menor que 8: $P(X < 8) = 21/36$ ¿por qué?;
- Tal vez deseamos saber si el resultado está entre 3 y 9: $P(3 < X < 9) = 23/36$ ¿por qué?;
- Podemos estar interesados en que X sea distinta de 7: $P(X \neq 7) = 1 - P(X = 7) = 1 - 1/6 = 5/6$. ¿Por qué?

Naturalmente, esta notación se extiende de manera natural a todo tipo de intervalos y desigualdades.

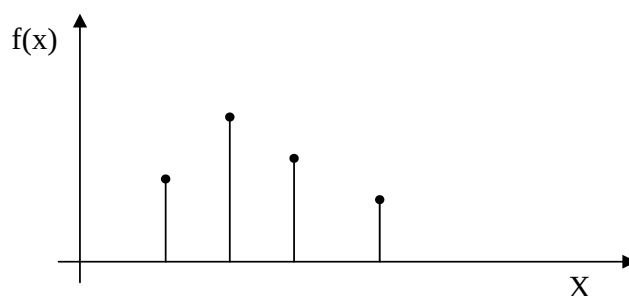
Regresando a donde estábamos,

A la función: $f(x) = P(X = x)$ se le llama **función de probabilidad** de X.

Esta función es una función común no tiene nada de aleatorio. Dicho de otra forma, una vez determinados los valores de las probabilidades, la función de probabilidad es una función común y corriente, tiene su dominio, su codominio o imagen, su gráfica, puede ser inyectiva, etc.

Hay algunos hechos importantes respecto a esta función:

1. Para una variable aleatoria discreta los valores posibles son los únicos para los cuales esta probabilidad es diferente de cero. O sea que, si el dominio de la función de probabilidad se ampliara a todos los reales, la función va a valer cero casi siempre, excepto en un conjunto discreto de puntos.
2. El valor de la función de probabilidad depende esencialmente de la variable aleatoria a la que nos referimos, pero es siempre positivo y menor que uno, cuando tengamos varias variables y puede no quedar claro a cuál variable nos referimos, es conveniente poner el símbolo de la variable como subíndice para la función: $f_X(x)$, $f_Y(y)$
3. La gráfica de una función de probabilidad se presenta de la siguiente forma:



Regresando al ejemplo de los dados, cambiemos el juego. En este juego se gana si el resultado es par y se pierde si es impar; la cantidad que pierde o gana será el doble del resultado. Aquí, su ganancia (positiva si Ud. gana, negativa si pierde) es una variable aleatoria Y, su imagen es

$$R_Y = \{-22, -18, -14, -10, -6, 4, 8, 12, 16, 20, 24\}$$

Completar la siguiente tabla con la función de probabilidad de Y:

y	-22	-18	-14	-10	-6	4	8	12	16	20	24
f(y)	2/36	4/36	6/36	4/36	2/36	1/36	3/36	5/36	5/36	3/36	1/36

Ahora con la tabla completada, calculemos:

- la probabilidad de ganar más de 8 pesos; $P(Y > 8) = 5/36 + 5/36 + 3/36 + 1/36$
- la probabilidad de que se pierda más de 10 pesos; $P(Y < -10) = 2/36 + 4/36 + 6/36$
- la probabilidad de que la ganancia esté entre dos valores, por ejemplo, entre -10 y +16 inclusive; $P(-10 < Y < 16) = 2/36 + 1/36 + 3/36 + 5/36$

Una función de probabilidad de una variable aleatoria discreta, para ser correcta, debe satisfacer dos propiedades:

- **$f(x)$ debe ser siempre mayor o igual a 0**
- **la suma de $f(x)$ para todos los valores de x debe dar 1**

Si volviéramos al ejemplo de arrojar una moneda dos veces:

$X=x$	0	1	2
$f(x)=P(X=x)$	1/4	1/2	1/4

La suma $P(X=0) + P(X=1) + P(X=2) = 1$

En nuestro caso de que se gana si el resultado de la suma de los dados es par y se pierde si es impar; la cantidad que pierde o gana será el doble del resultado será:

y	-22	-18	-14	-10	-6	4	8	12	16	20	24
$f(y)$	2/36	4/36	6/36	4/36	2/36	1/36	3/36	5/36	5/36	3/36	1/36

$$P(Y=-22)+P(Y=-18)+P(Y=-14)+P(Y=-10)+P(Y=-6)+P(Y=4)+P(Y=8)+P(Y=12)+P(Y=16)+P(Y=20)+P(Y=24)=1$$

FUNCIÓN DE DISTRIBUCIÓN

Cuando la imagen de una variable aleatoria es un intervalo Real decimos, que la variable es continua. La matemática que utilizamos para las variables continuas es diferente a la de las discretas.

Una función muy útil en el cálculo de probabilidades de una variable aleatoria (**VA**) es la función de distribución que surge de acumular la función de probabilidad. *Esta función se define de igual forma para continuas y discretas:*

$$F(x) = P(X \leq x)$$

Para no confundirnos con la notación, suele escribirse también así:

$$F(t) = P(X \leq t) \text{ siendo } F: \mathbb{R} \rightarrow [0,1]$$

La función de distribución tiene las siguientes propiedades:

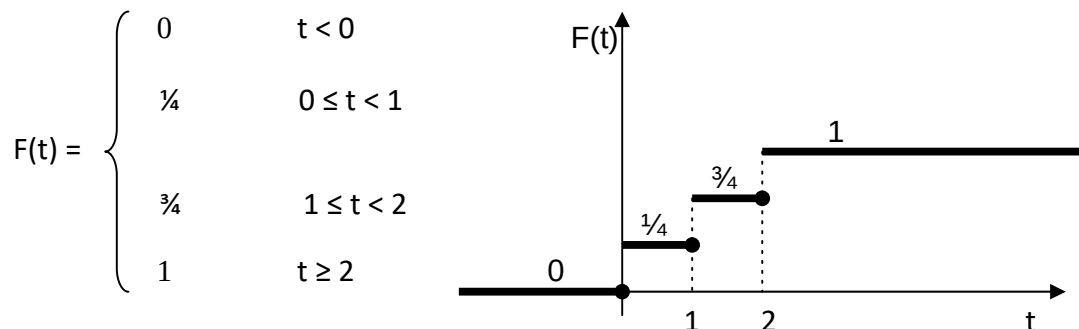
1. $F(\text{en menos infinito}) = 0$
2. $F(\text{en más infinito}) = 1$
3. F es una función es monótona creciente, siempre crece, es no decreciente.
4. Es una función positiva.
5. F sirve para calcular la probabilidad de que la variable esté dentro de un intervalo, de la siguiente forma: $P(a < X \leq b) = F(b) - F(a)$

La función de distribución para variables aleatorias discretas es monótona creciente, no decreciente, y escalonada.

Para que esta última propiedad nos sea de utilidad deberíamos tener la distribución ya calculada. Para muchas variables aleatorias de uso común, las distribuciones ya están calculadas y tabuladas. Por lo que se han armado tabas de los diferentes valores de la variable, variables muy frecuentes dentro de nuestra vida. Por eso más adelante deberemos identificar a qué tipo de variable se asociará la variable que tenemos en estudio para no tener que hacer los cálculos, sino que podamos tomar los resultados de la Tabla de la variable a la cual asociamos nuestro problema. A esta asociación de un problema real a una variable tipificada o tabulada se lo llama **modelizar**, y se basa en tomar un modelo que se le parezca a nuestro problema de forma de sólo analizar los valores de nuestras tablas y dedicarnos no a hacer cuentas, sino a tomar conclusiones en nuestro problema.

Las 4 propiedades que se señalaron arriba *definen* a una función de distribución, de modo que para saber si una función es una distribución o no, basta ver que las cumpla.

En el ejemplo de la moneda la Función de Distribución será:



VALOR ESPERADO o ESPERANZA

Hay un par de cantidades importantes asociadas con cada variable aleatoria:

- el valor esperado, esperanza matemática o simplemente esperanza o media.
- la varianza y
- su raíz cuadrada, la desviación estándar.

La **esperanza** para una **VAD** es:

$$E(X) = \sum x_i \cdot P(X=x_i) = \sum x_i \cdot f(x_i) \quad \forall x_i \in R_X$$

En nuestro ejemplo de la moneda:

$$E(X) = 0 \cdot \frac{1}{4} + 1 \cdot \frac{2}{4} + 2 \cdot \frac{1}{4} = 1$$

Propiedades:

Si una distribución de probabilidad es simétrica respecto de un valor μ , dicho valor es el valor esperado $E(X)$

La **varianza** de una variable aleatoria discreta (VAD) mide cuánto se dispersan los valores respecto del valor esperado $E(X)$.

$$V(X) = \sum (x_i - E(X)) \cdot f(x_i) = \sum (x_i - \mu)^2 \cdot f(x_i) \quad \forall x_i \in R_X$$

$$V(X) = \sum (x_i^2 \cdot f(x_i) - 2\mu x_i \cdot f(x_i) + E^2(X) \cdot f(x_i)) =$$

$$= \sum x_i^2 \cdot f(x_i) - 2\mu \sum x_i \cdot f(x_i) + \mu^2 \cdot \sum f(x_i) =$$

$$= \sum x_i^2 \cdot f(x_i) - 2\mu^2 + \mu^2$$

$$= \sum x_i^2 \cdot f(x_i) - \mu^2$$

$$V(X) = E(X^2) - E^2(X)$$

Fórmula cómoda!!!

En nuestro ejemplo de la moneda:

$$E(X^2) = 0^2 \cdot 1/4 + 1^2 \cdot 2/4 + 2^2 \cdot 1/4 = 3/2$$

$$V(X) = \sigma^2 = E(X^2) - E^2(X) = 3/2 - 1^2 = 1/2$$

El desvío estándar o dispersión es $\sqrt{V(X)} = \sigma \Rightarrow \sigma = 0,7071$

Ejercicio:

El señor A le propone a B jugar a un juego a cambio de un pago de \$40. B debe arrojar un dado y A le pagará \$10 si sale 1 ó 2; \$20 si sale 3 ó 4; \$40 si sale 5; \$80 si sale 6. ¿Considera que A es un "tonto"?

x	1	2	3	4	5	6
f(x)	1/6	1/6	1/6	1/6	1/6	1/6

Valor esperado = $\$10 \cdot 1/6 + \$10 \cdot 1/6 + \$20 \cdot 1/6 + \$20 \cdot 1/6 + \$40 \cdot 1/6 + \$80 \cdot 1/6 = \$30$ $E(X) = \$30$

El valor que es más probable que gane B es \$30 y habiendo pagado \$40 por jugar, lo más probable es que en total pierda \$10 (y que el señor A gane \$10).

Ejemplo: Si X es una VA. discreta cuya función de probabilidad es

x	1	2	3	4	5	6
f(x)	1/21	2/21	3/21	4/21	5/21	6/21

Calculemos $E(X) = 13/3$ y $V(X) = 20/9$ entonces $\sigma(X) = 1.4907$.

Ejemplo de Esperanza: Se tira un dado y se gana si el resultado es par y se pierde si es impar; la cantidad que pierde o gana será el doble del resultado. Llamando X a la VA. que señala su ganancia, X tiene la siguiente función de probabilidad.

X	-10	-6	-2	4	8	12
$f(x)$	1/6	1/6	1/6	1/6	1/6	1/6

Si jugamos una vez, no sabemos si vamos a perder o a ganar. Pero si jugamos muchas veces, las probabilidades nos dicen que 1/6 de las veces que juguemos perderemos 10 pesos, 1/6 perderemos 6 y así sucesivamente. Nuestra ganancia final será:

$$E(X) = (1/6)(-10) + (1/6)(-6) + (1/6)(-2) + (1/6)(4) + (1/6)(8) + (1/6)(12) = 1$$

esto es, la ganancia será: \$ 1.

Esta ganancia no es segura, ¿qué sentido tiene? Al aumentar el número de veces que jugamos, las frecuencias de los diferentes resultados del dado se acercarán a las probabilidades y la ganancia se estabilizará cerca de 1 peso.

Ejemplo de Varianza: Se lanza una moneda una sola vez, si cae cara se gana 4 pesos, si no, se pierde 2 pesos. La ganancia esperada ahora es:

$$E(X) = (1/2)(4) + (1/2)(-2) = 1$$

es decir, también \$1.

¿Son iguales los juegos? Desde el punto de vista de la ganancia esperada son iguales. Pero es claro que el juego con el dado involucra mayores cantidades, es decir mayores riesgos. Es más fácil que en un juego con el dado terminemos más alejados de la ganancia esperada que en el juego con la moneda. Esto se refleja en la varianza. Veamos, la varianza del juego anterior es:

$$E(X^2) = (1/6)(-10)^2 + (1/6)(-6)^2 + (1/6)(-2)^2 + (1/6)(4)^2 + (1/6)(8)^2 + (1/6)(12)^2 = 60,67$$

$$V(X) = E(X^2) - E^2(X) = 60,67 - 1 = 59,67$$

Por otra parte, la varianza del juego con la moneda es:

$$E(X^2) = (1/2)(3)^2 + (1/2)(-3)^2 = 9$$

$$V(X) = E(X^2) - E^2(X) = 9 - 1 = 8$$

El segundo juego es más conservador, las discrepancias entre los resultados y el resultado esperado son menores.

Juego limpio

En el juego con la moneda sabemos cómo volverlo equitativo, basta que hagamos que se gane y pierda lo mismo. Pero ¿cómo volver equitativo el juego con el dado? Si mantenemos la misma cantidad en caso de ganar, es decir 2 pesos por cada punto, cuánto hay que poner en la pérdida para equilibrar el juego. Llamemos k a la cantidad que equilibra el juego. El valor esperado es ahora:

$$(1/6) \cdot (-5k) + (1/6) \cdot (-3k) + (1/6) \cdot (-k) + (1/6) \cdot (4) + (1/6) \cdot (8) + (1/6) \cdot (12)$$

y la suma es: $(1/6)(24 - 9k)$. Como queremos equidad en el juego, esta esperanza debe ser cero. O sea, $k = 24/9$