

Unidad 1

a. Probabilidades y Estadística

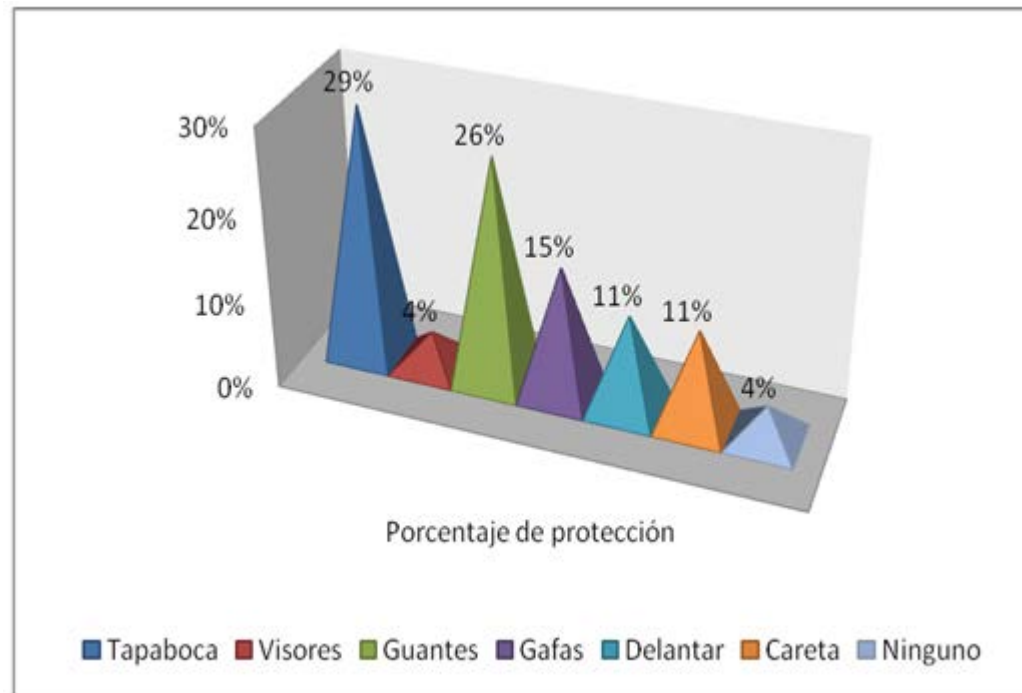
1

IN3401

SEMESTRE OTOÑO, 2012

ESTADÍSTICA DESCRIPTIVA (Continuación)

2



Medidas de Tendencia Central – Media Geométrica

3

- La *media geométrica* proporciona una medida precisa de un cambio porcentual promedio de una serie de números:

$$MG = \sqrt[n]{X_1 X_2 \cdots X_n}$$

- Es de utilidad en series económicas y ventas. Ej: tasa de crecimiento promedio en los ingresos (para tomar decisiones como campañas publicitarias)

Medidas de Tendencia Central – Media G.(2)

4

Año	Ingreso	Porcentaje del año anterior
1992	US\$ 50,000	-- --
1993	55,000	55/50 = 1.10
1994	66,000	66/55 = 1.20
1995	60,000	60/66 = 0.91
1996	78,000	78/60 = 1.30

$$MG = \sqrt[4]{(1.10)(1.2)(0.91)(1.3)} = 1.1179$$

Medidas de Dispersión

5

- Miden que tanto se dispersan o desvían los datos en torno a la media.
- El *rango* es la medida de dispersión más simple (y menos útil). El rango es simplemente la diferencia entre la observación más alta y la más baja.
- La desventaja es que sólo considera dos de los (posiblemente) cientos de observaciones, ignorando el resto de los datos.

Medidas de Dispersión - Varianza

6

- La varianza es el promedio de las desviaciones de las observaciones con respecto a su media al cuadrado.

$$\sigma^2 = \frac{\sum_{i=1}^N (X_i - \mu)^2}{N}$$

- La desviación estándar es la raíz de la varianza. Es una medida muy útil de dispersión ya que tiene las mismas unidades que la variable estudiada.

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

Medidas de Dispersión – Varianza(2)

7

- La *varianza muestral* sigue la misma lógica:

$$s^2 = \frac{\sum_{i=1}^N (X_i - \bar{X})^2}{n-1} \quad s = \sqrt{s^2}$$

- Llama la atención que se divide por $n-1$, lo que se debe a que este estadístico tiene $n-1$ *grados de libertad*.
- Los grados de libertad equivalen al número de observaciones menos el número de *restricciones* impuesta en tales observaciones.

Medidas de Dispersión – Varianza(3)

8

- Una restricción es cualquier valor que deba calcularse de dichas observaciones. En este caso la restricción es el cálculo de la media muestral.
- Ej: Se tienen $n=4$ obs. que producen una media de 10, la media de 10 sirve como restricción y hay $n-1=3$ g.l.
- Se pueden escoger 3 obs. Cualquiera, por ej. 8, 9 y 11.
- Después ya no hay libertad para escoger la última obs., que debe ser 12 si se quiere tener un promedio de 10.

Medidas de Dispersión

- Otras medidas de dispersión son los *cuartiles*, los *deciles* y los *percentiles*.
- Cada conjunto de datos tiene tres cuartiles que lo dividen en cuatro partes iguales. El primer cuartil (inferior) cuenta con el 25% de las observaciones. El segundo es justo la mitad (50% de los datos) y el tercero el 25% superior.
- Los deciles separan el conjunto de datos en 10 subconjuntos iguales, y los percentiles en 100 partes.

Medidas de Dispersión(2)

10

- Un percentil y su ubicación en un arreglo ordenado se identifica mediante subíndices. Por ejemplo, el 15vo percentil es indica como P_{15} , y su ubicación en la serie es L_{15}
- El lugar del P-ésimo percentil es:
$$L_p = (n + 1) \frac{P}{100}$$
- Ejemplo: número de acciones transadas en la Bolsa de Valores de Nueva York:

Medidas de Dispersión(3)

11

3	10	19	27	34	38	48	56	67	74
4	12	20	29	34	39	48	59	67	74
7	14	21	31	36	43	52	62	69	76
9	15	25	31	37	45	53	63	72	79
10	17	27	34	38	47	56	64	73	80

- Se desea calcular el percentil 25, P_{25} , para estas acciones. Primero debe hallarse su ubicación en la serie:

$$L_{25} = (50 + 1) \frac{25}{100} = 12.75$$

Medidas de Dispersión(4)

12

- El valor resultante (12.75) dice que el percentil 25 está ubicado al 75% del trayecto comprendido entre la doceava (20) y la treceava (21) observación, i.e.

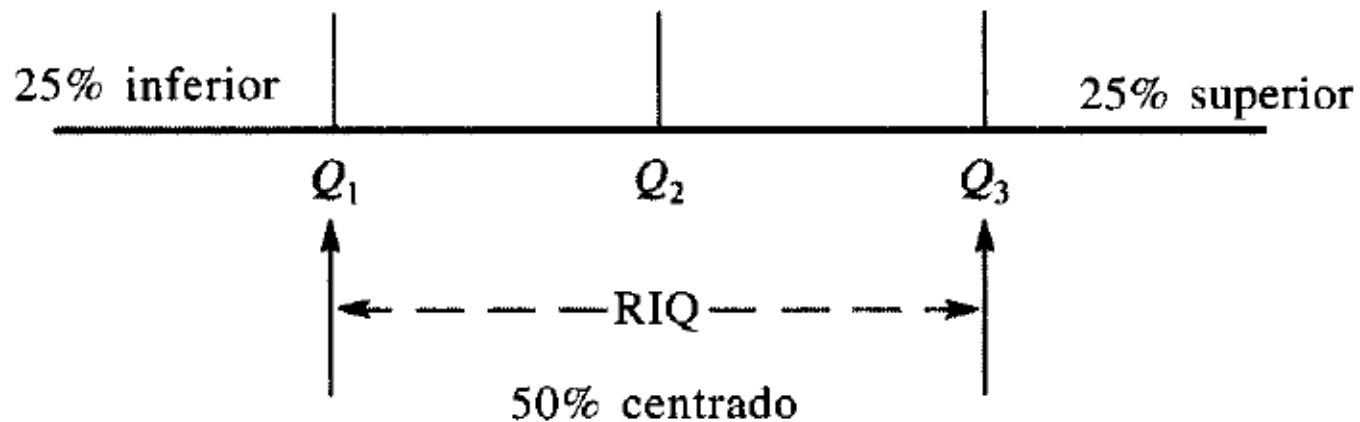
$$P_{25} = 20 + 0.75(21 - 20) = 20.75$$

- Notar que el primer decil es equivalente a P_{10} , el segundo P_{20} y así sucesivamente.
- El primer cuartil es igual a P_{25} , el segundo a P_{50} y el tercero a P_{75}

Medidas de Dispersión(5)

13

- Una medida única de dispersión es el rango intercuartílico (RIQ), la diferencia entre el tercer cuartil y el primer cuartil (50% de las obs.).

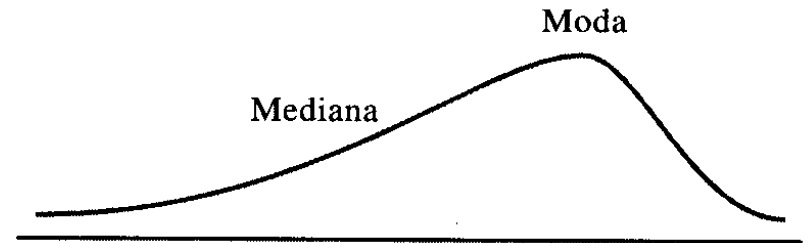
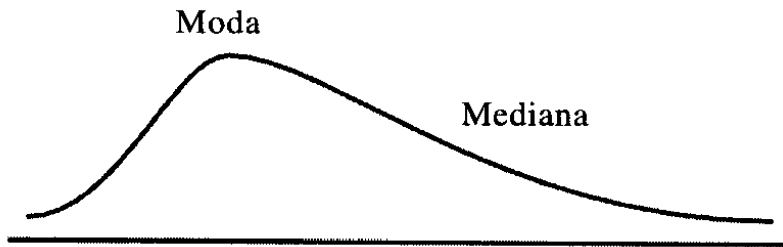


- Esta medida no está muy influenciada por observaciones extremas

Asimetría

14

- **Asimetría o Sesgo (*skewness*):** No todas las distribuciones son normales, algunas están sesgadas a la izquierda o derecha:



- Se define el coeficiente de Asimetría o Sesgo (Pearson):

$$A_p = \frac{\mu - moda}{\sigma},$$

Asimetría (2)

15

Interpretación:

- Si $P < 0$, los datos están sesgados a la izquierda (asimetría negativa)
- si $P > 0$ entonces están sesgados a la derecha (asimetría positiva)
- $P = 0$ implica que los datos se distribuyen normalmente.

Curtosis

16

- Se basa en el promedio de las desviaciones típicas a la cuarta potencia y representa el apuntalamiento de la distribución:

$$\sum_{i=1}^n \left[\frac{(x_i - \bar{X})^4 / S^4}{n} \right] - 3$$

- Se corrige en 3 que corresponde a la curtosis de la normal.
- Si vale cero es *mesocúrtica*
- Si es positiva es más apuntalada que la normal y se llama leptocúrtica.
- Si es negativa es más achatada que la normal y se llama platicúrtica.

Curtosis(2)

17

- La curtosis es uno de los conceptos peor comprendidos en la estadística. Se suele confundir con la varianza:

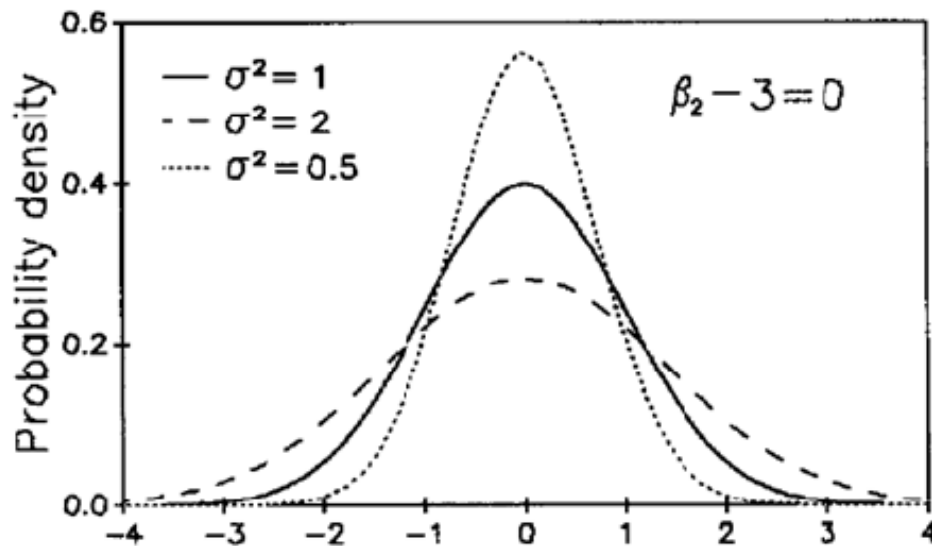
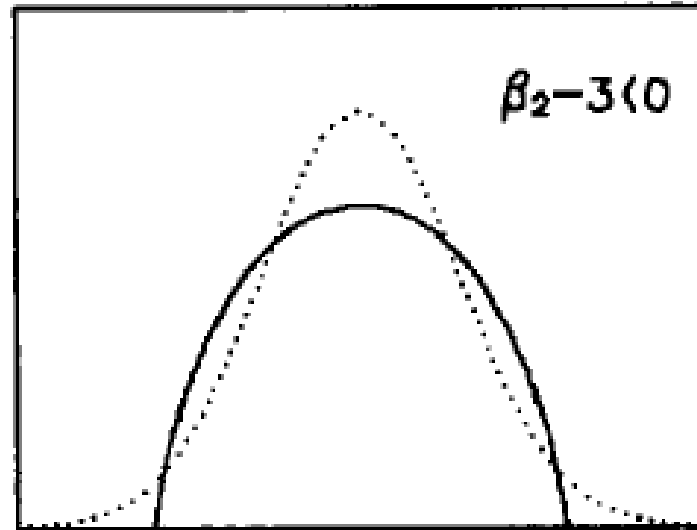
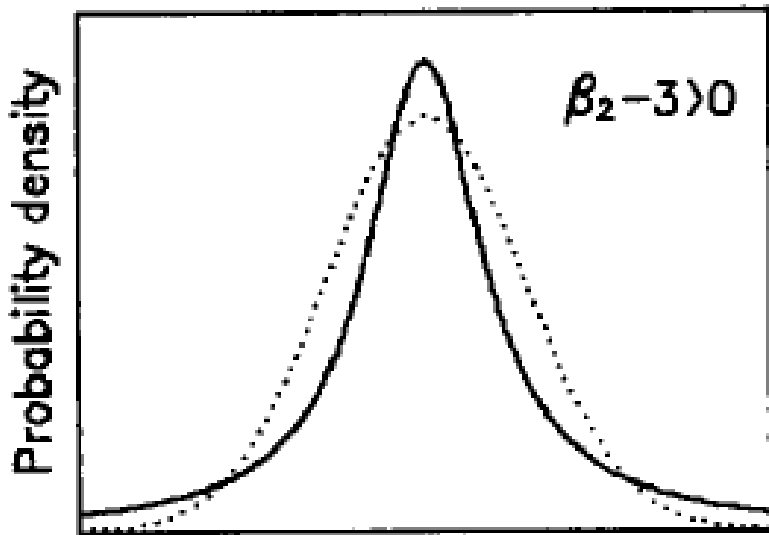


Figure 4. Three normal distributions with variances (σ^2) of 0.5, 1, and 2.

Curtosis(3)

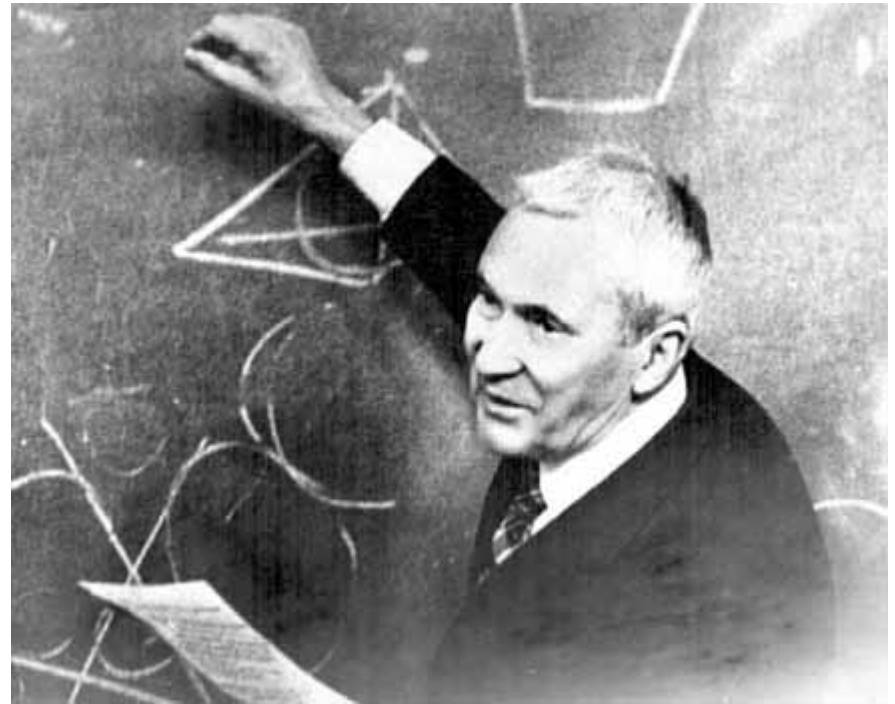
18

- La curtosis representa una medida adimensional que representa un movimiento de masa que no afecta la varianza.



PROBABILIDADES

19



Probabilidades

20

- La *probabilidad* es la posibilidad numérica de que ocurra un evento.
- La probabilidad de un evento es medida por valores comprendidos entre 0 (imposibilidad) y 1 (certeza).
- Un *experimento* (aleatorio) es una acción que puede tener distintos resultados posibles.
- El *espacio muestral* es el conjunto de resultados posibles de un experimento. Se suele representar por EM o Ω .

Probabilidades(2)

21

- Ejemplos:
 - Experimento 1: “tirar una moneda y ver que sale”
 $\Omega = \{C, S\}$
 - Experimento 2: “tirar dos monedas y ver que sale”
 $\Omega = \{CC, SS, CS\}$
- Un *suceso* es un subconjunto del espacio muestral
 $S_1 = \{C\}, S_2 = \{S\}, S_3 = \{CC, SS\}$

Probabilidades(3)

22

- **Modelo clásico (Laplace):**

$$P(E) = \frac{\text{Número de formas que puede ocurrir el evento E}}{\text{Número total de posibles resultados}}$$

- **Ejemplo: $P(\text{cara})=1/2$**

Probabilidades(4)

23

- **Modelo empírico:**

$$P(E) = \frac{\text{Número de veces que ha ocurrido el evento E}}{\text{Número total de experimentos}}$$

- **Ejemplo: P(cara)**

Definición Axiomática

24

- Axioma 1: la probabilidad no puede ser negativa:

$$P(A) \geq 0$$

- Axioma 2: la probabilidad del espacio muestral es uno

$$P(EM) = 1$$

- Axioma 3: dos conjuntos son disjuntos ssi la probabilidad de su unión es la suma de sus probabilidades:

$$A \cap B = \emptyset \leftrightarrow P(A \cup B) = P(A) + P(B)$$

Consecuencias

25

- Consecuencia 1: $P(A) \leq 1$
- Consecuencia 2: $P(A) + P(\bar{A}) = 1$
- Consecuencia 3: $P(\emptyset) = 0$
- Consecuencia 4: $A \subseteq B \rightarrow P(A) \leq P(B)$
- Consecuencia 5: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Probabilidad Condicional

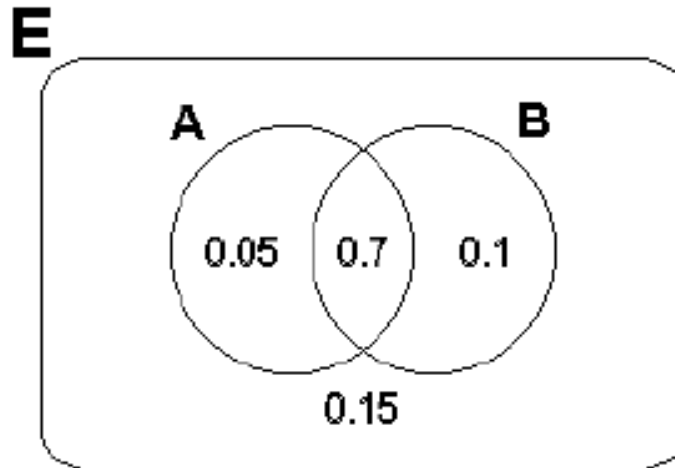
26

- *Probabilidad condicional* $P(A|B)$ es la probabilidad de que ocurra el evento A, dado que el evento B ya haya ocurrido.
- Ejemplo:
 - el 80% de los alumnos estudió para el examen
 - el 75% de los alumnos aprobó el examen
 - el 15% de los alumnos no estudió para el examen y no aprobó.
- Sea A el suceso “alumno aprobó examen” y B el suceso “el alumno estudió”. Se tiene que $P(A)=0.75$, $P(B)=0.8$ y
$$P(\bar{A} \cap \bar{B}) = 0.15$$

Probabilidad Condicional(2)

27

- Gráficamente:



- **Cual es la probabilidad de que un alumno que estudió haya aprobado el examen?**

Probabilidad Condicional(3)

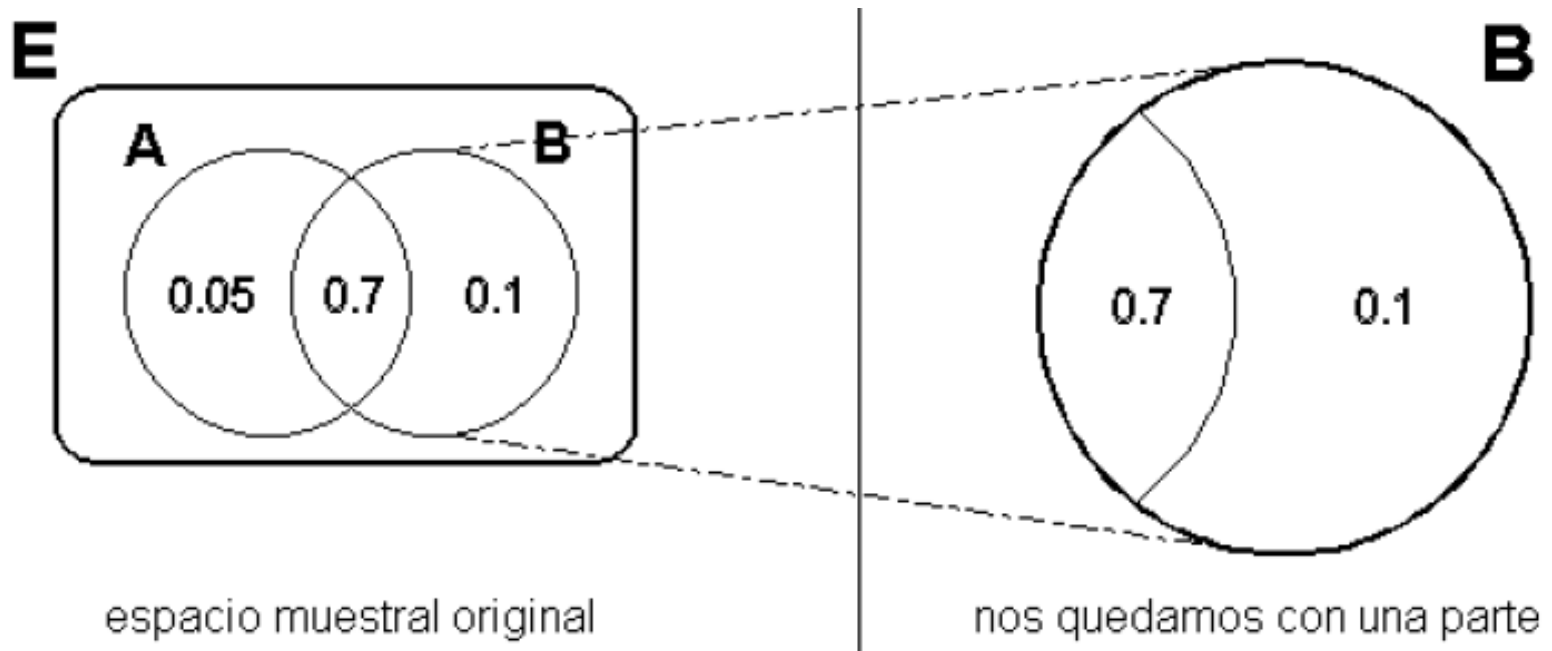
28

- Intuitivamente, los alumnos que estudiaron fueron el 80%
- Ese 80% está formado por un 70% que aprobó y un 10% que no aprobó. La probabilidad de aprobar es $70/80=0,88$
- Formalmente:
$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Probabilidad Condicional(4)

29

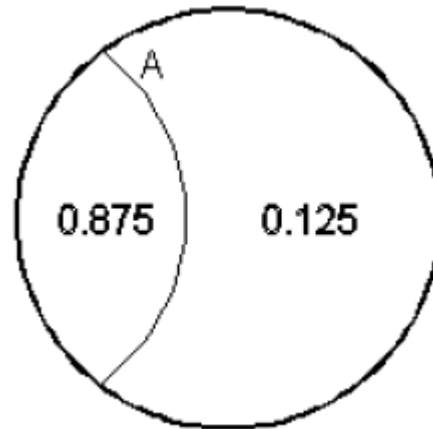
- Intuitivamente, $P(A|B)$ es la probabilidad de “estar parados en A, sabiendo que estamos parados en B”.



Probabilidad Condicional(5)

30

- Sin embargo, B no está listo para ser espacio muestral (probabilidades no suman 1)
- Es necesario dividir las probabilidades de B por un factor para que sea EM manteniendo la proporción relativa.
- Como las probabilidades contenidas en B suman $P(B)$, dividiendo por este factor se cumple lo anterior



el nuevo espacio muestral

Propiedades

31

- Conmutatividad intersección:

$$P(A \cap B) = P(A|B)P(B) = P(B|A)P(A) = P(B \cap A)$$

- Intersección 3 eventos:

$$P(A \cap B \cap C) = P(A)P(A|B)P(C|A \cap B)$$

- Principio Bayes:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Independencia

32

Dos sucesos A,B son independientes ssi:

- $P(A|B) = P(A)$



- $P(B|A) = P(B)$



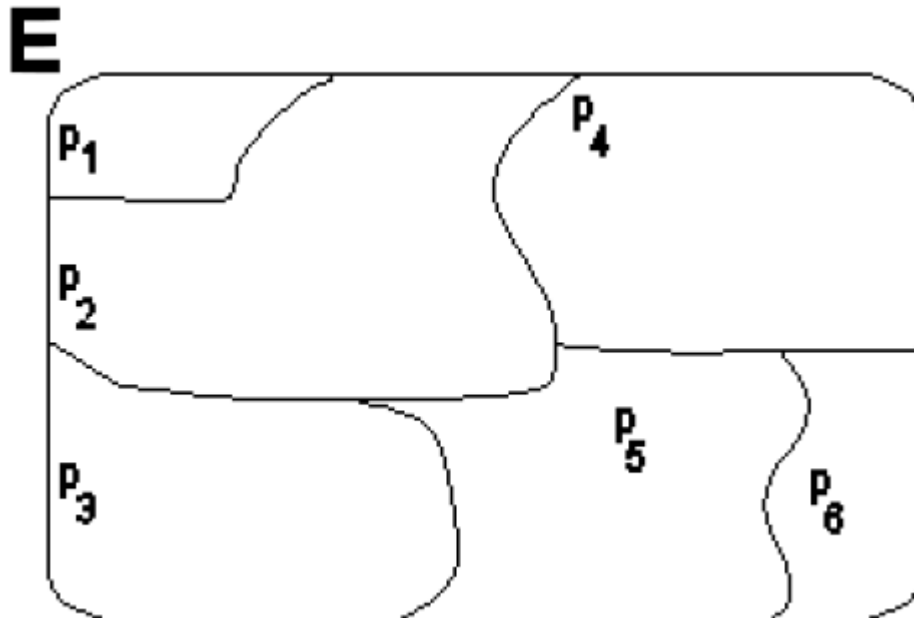
- $P(A \cap B) = P(A)P(B)$

- Advertencia: La independencia de dos sucesos no tiene nada que ver con que dos sucesos sean disjuntos. De hecho, si dos sucesos, con probabilidades no nulas, son independientes, entonces no pueden ser disjuntos, ya que $p(A \cap B) = p(A) \cdot p(B) \neq 0$.

Probabilidades Totales

33

- Consideremos un espacio muestral E , con la siguiente partición:



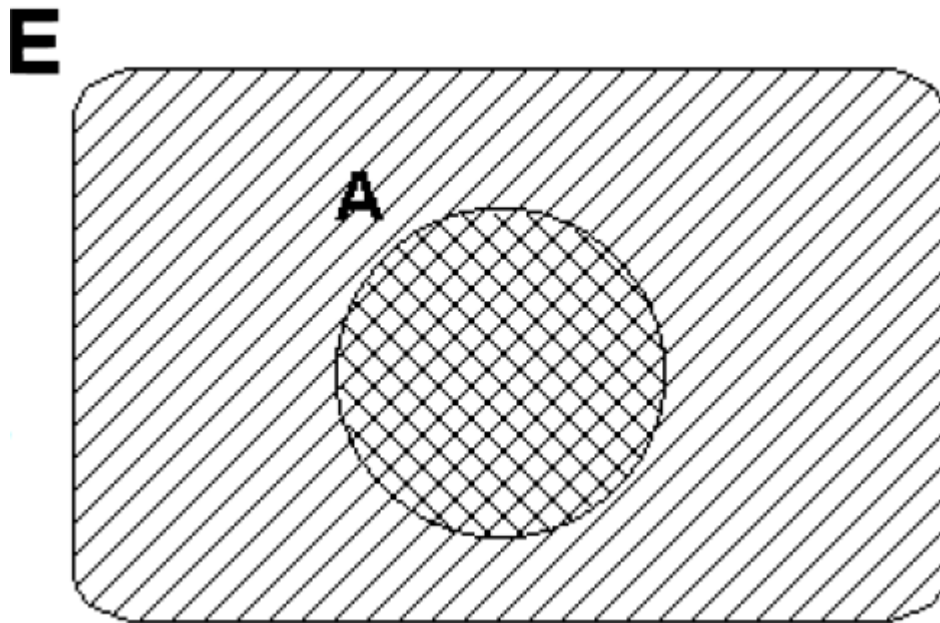
$$E = \sum_i p_i$$

$$p_i \cap p_j = \emptyset$$

Probabilidades Totales(2)

34

- Además se cuenta con el suceso A , que es subconjunto del espacio muestral:



$$P(A) = P(A \cap E)$$

Probabilidades Totales(3)

35

- Dado que E es la sumatoria de las probabilidades de la partición establecida:

$$P(A) = P(A \cap E) = P\left(A \cap \sum_i p_i\right)$$

- Aplicando la propiedad distributiva de conjuntos:

$$P\left(A \cap (p_1 \cup p_2 \cup \dots \cup p_n)\right) = P\left((A \cap p_1) \cup (A \cap p_2) \cup \dots \cup (A \cap p_n)\right)$$

Probabilidades Totales(4)

36

- Utilizando el tercer axioma podemos escribir la probabilidad de la suma (unión) como suma de probabilidades:

$$\begin{aligned} P((A \cap p_1) \cup \dots \cup (A \cap p_n)) \\ = P((A \cap p_1)) + \dots + P((A \cap p_n)) \end{aligned}$$

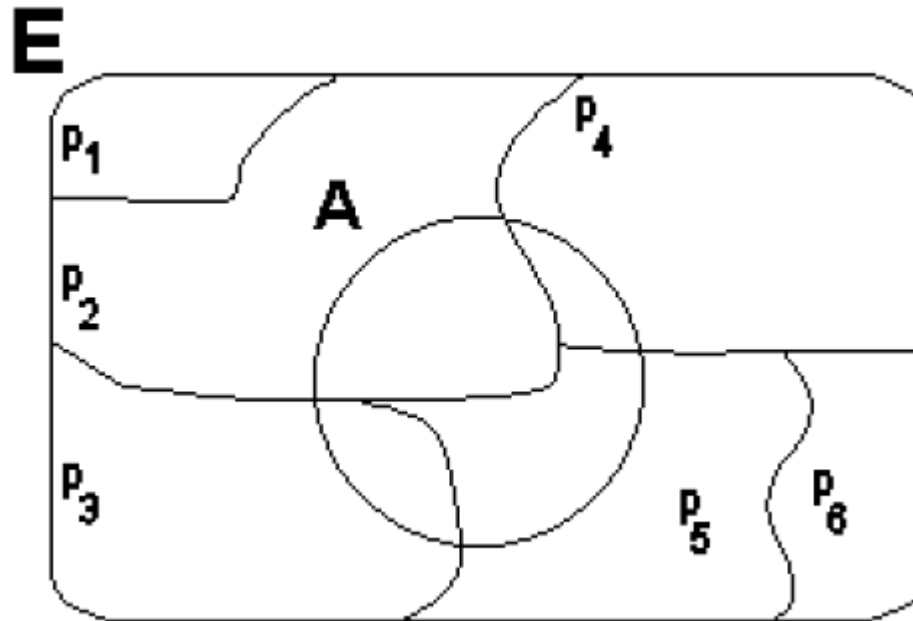
- En resumen, llegamos a lo que se conoce como *probabilidad total*:

$$\sum_i P(A \cap p_i)$$

Probabilidades Totales(5)

37

- Gráficamente:



Probabilidades Totales(5)

38

- En particular, para una partición de un suceso D y su complemento:

$$P(A) = P(A \cap D) + P(A \cap \bar{D})$$

- Utilizando ahora la fórmula de probabilidad condicional:

$$P(A) = P(A|D)P(D) + P(A|\bar{D})P(\bar{D})$$

- En general:

$$P(A) = \sum_i P(A \cap p_i) = \sum_i P(A|p_i) P(p_i)$$

Teorema de Bayes

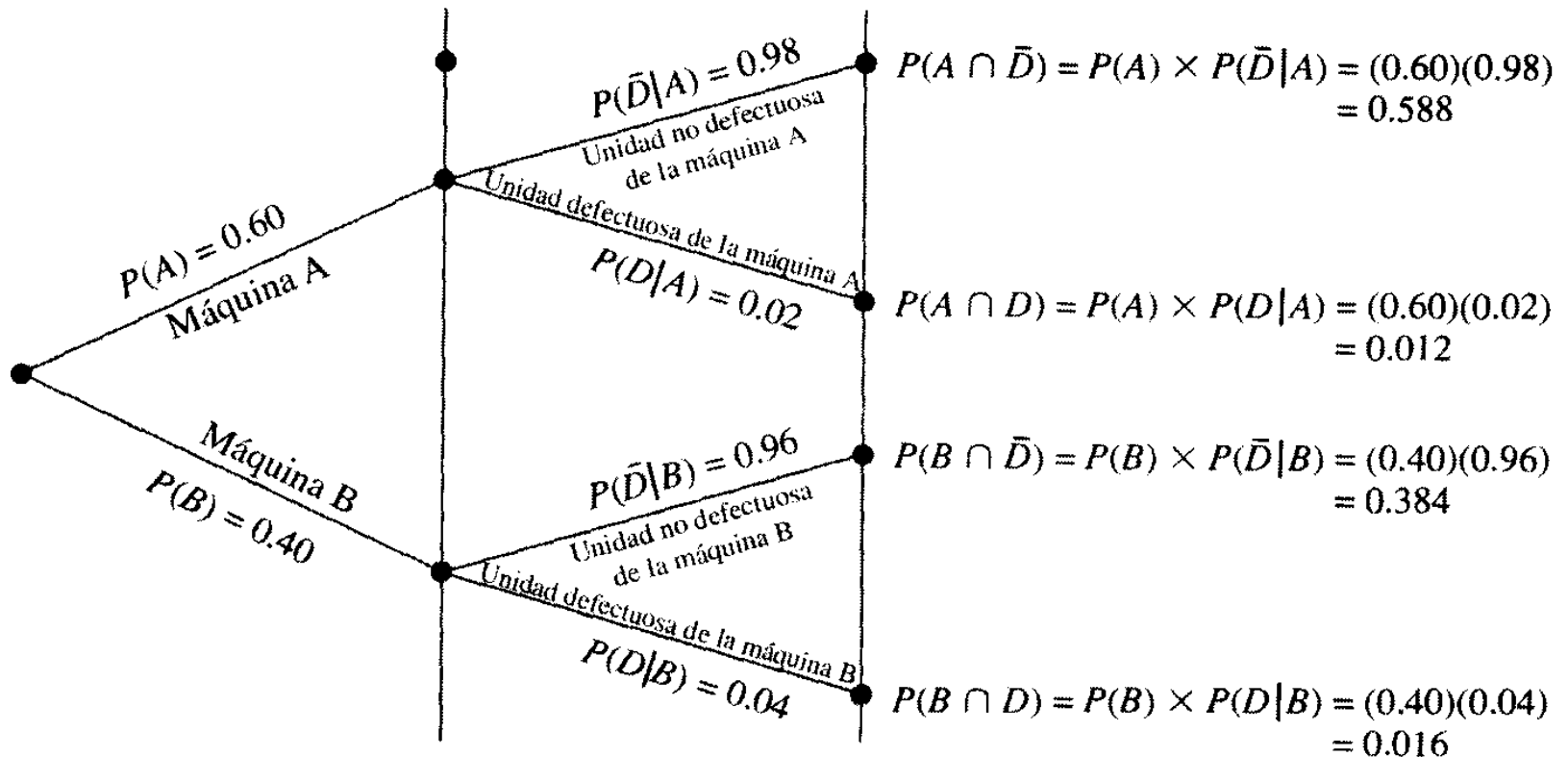
39

- **Ejemplo:** En una empresa manufacturera, una máquina A produce el 60% de la producción total, mientras que una máquina B el restante 40%.
- El 2% de las unidades producidas por A son defectuosas, mientras que B tiene una tasa de defectos del 4%.
- Se cuenta con una unidad defectuosa, se desea conocer la probabilidad de que venga de la máquina A.

Teorema de Bayes(2)

40

- Árbol:



Teorema de Bayes(3)

41

- La probabilidad $P(A|D)$ se puede obtener utilizando la tercera propiedad obtenida por la probabilidad condicional.
- Sin embargo, se desconoce $P(D)$. Necesitamos aplicar probabilidades totales:

$$P(D) = P(D|A)P(A) + P(D|B)P(B)$$

- Bayes:
$$P(A|D) = \frac{P(D|A)P(A)}{P(D|A)P(A) + P(D|B)P(B)}$$

Teorema de Bayes(4)

42

- Volviendo al problema:

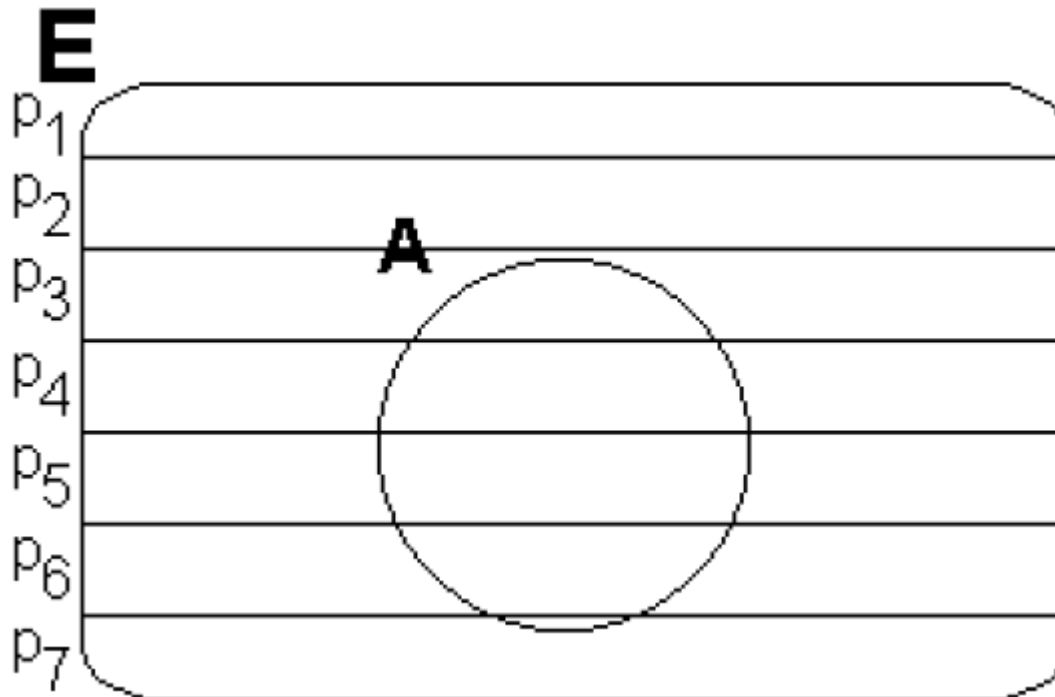
$$P(A|D) = \frac{0.012}{0.012 + 0.016} = 0.43$$

- Tiene sentido? $P(A|D) < 0.5$? $P(A|D) < P(A)$?

Teorema de Bayes(5)

43

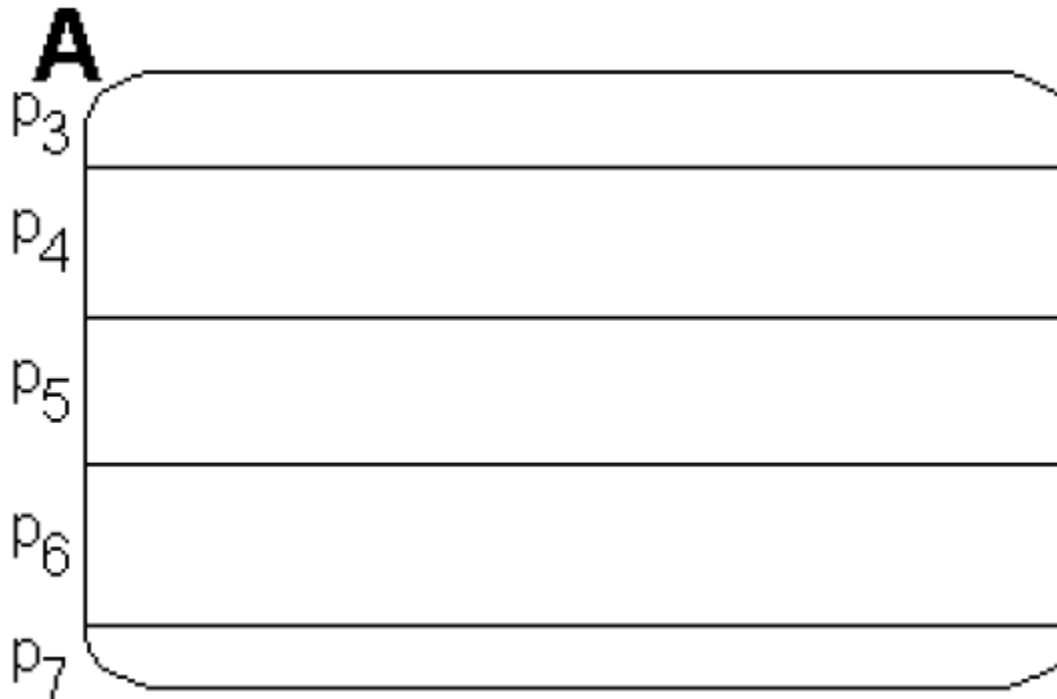
- Gráficamente, tenemos un suceso A en un espacio muestral particionado. Conocemos las *probabilidades a priori* o probabilidades de las partes **sabiendo que ocurrió A** :



Teorema de Bayes(6)

44

- Nos interesan las *probabilidades a posteriori* o probabilidades originales de las partes p_i :



Teorema de Bayes(7)

45

- La *probabilidades a posteriori* son:

$$P(p_i | A) = \frac{P(A|p_i)P(p_i)}{\sum_j P(A|p_j) P(p_j)}$$

- Esta última formula se conoce como la regla de Bayes.

Variables Aleatorias

46

- Vamos a llamar *variable aleatoria* a una variable cuyo valor sería el resultado de un determinado experimento.
- Por ejemplo, si el experimento consiste en arrojar un dado, podemos definir la variable aleatoria X cuyo valor será el número que salga en el dado.
- El conjunto de valores posibles de X es el espacio muestral.
- En general nos interesará cuál es la probabilidad de que X asuma cada valor.

Variables Aleatorias

47

- Formalmente, una variable aleatoria (v.a.) X es una función real definida en el espacio muestral asociado a un experimento aleatorio, Ω .

$$X: \Omega \rightarrow \mathbb{R}$$

- Se llama *rango* de una v.a. X y lo denotaremos R_X , al conjunto de los valores reales que ésta puede tomar, según la aplicación X . Dicho de otro modo, el rango de una v.a. es el *recorrido* de la función por la que ésta queda definida.

Variables Aleatorias

48

- Ejemplo: Supongamos que se lanzan dos monedas al aire. El espacio muestral (conjunto de resultados posibles) asociado al experimento, es:

$$\Omega = \{cc, cs, ss\}$$

- Podemos asignar entonces a cada suceso elemental del experimento el número de caras obtenidas. De este modo se definiría la variable aleatoria X como la función dada por

$$X: \Omega \rightarrow \mathbb{R} \qquad \{cc, cs, ss\} \rightarrow \{2, 1, 0\}$$

- El recorrido o rango de esta función, R_X , es el conjunto $R_X = \{0, 1, 2\}$

Variables Aleatorias(2)

49

- Para designar a las variables aleatorias se utilizan letras mayúsculas. Para designar a uno de sus valores posibles, se usan las letras minúsculas.
- Por ejemplo, si X es la variable aleatoria asociada a lo que sale al tirar un dado, podemos decir que $P(X = x) = 1/6, \forall x$.
- Las v.a. son consistentes con algunos conceptos introducidos anteriormente.

Variables Aleatorias(3)

50

- Ejemplo: Se tiene el experimento "tirar un dado y considerar el número que sale". El espacio muestral es
$$EM = \{1, 2, 3, 4, 5, 6\}$$
- Definiremos una variable aleatoria X : el número que sale al tirar el dado.
- Ahora usaremos esa variable aleatoria para calcular la probabilidad de que salga un número mayor que 3. Es decir:

$$P(X > 3)$$

- Observemos que " $X > 3$ " es un suceso. Ahora lo vamos a reemplazar por el suceso equivalente " $X=4 \vee X=5 \vee X=6$ "

Variables Aleatorias(4)

51

- **Ejemplo:**

$$P(X > 3) = P(X=4 \vee X=5 \vee X=6)$$

- Como los sucesos " $X=4$ ", " $X=5$ " y " $X=6$ " son disjuntos, podemos sumar sus probabilidades:

$$P(X=4 \vee X=5 \vee X=6) = P(X=4) + P(X=5) + P(X=6)$$

- Y ahora reemplazamos por las probabilidades que ya son conocidas:

$$P(X=4) + P(X=5) + P(X=6) = 1/6 + 1/6 + 1/6$$

- Con lo cual $P(X > 3) = 1/2$.

Tipos de Variables Aleatorias

52

- **Variable aleatoria discreta:** una v.a. es discreta si su recorrido es un conjunto discreto. La variable del ejemplo anterior es discreta. Sus probabilidades se recogen en la función de distribución.
- **Variable aleatoria continua:** una v.a. es continua si su recorrido no es un conjunto numerable. Intuitivamente esto significa que el conjunto de posibles valores de la variable abarca todo un intervalo de números reales.
- Por ejemplo, la variable que asigna la estatura a una persona extraída de una determinada población es una variable continua ya que, teóricamente, todo valor dentro de un rango es posible.

Función de probabilidades (v.a.d)

53

- Esta función le asigna a cada valor posible de la variable aleatoria un número real que consiste en la probabilidad de que ocurra, y debe cumplir con las 2 condiciones que enunciamos antes:

a) no puede ser negativa en ningún punto

$$\mathbb{P}(X = x_i) = p_i \geq 0$$

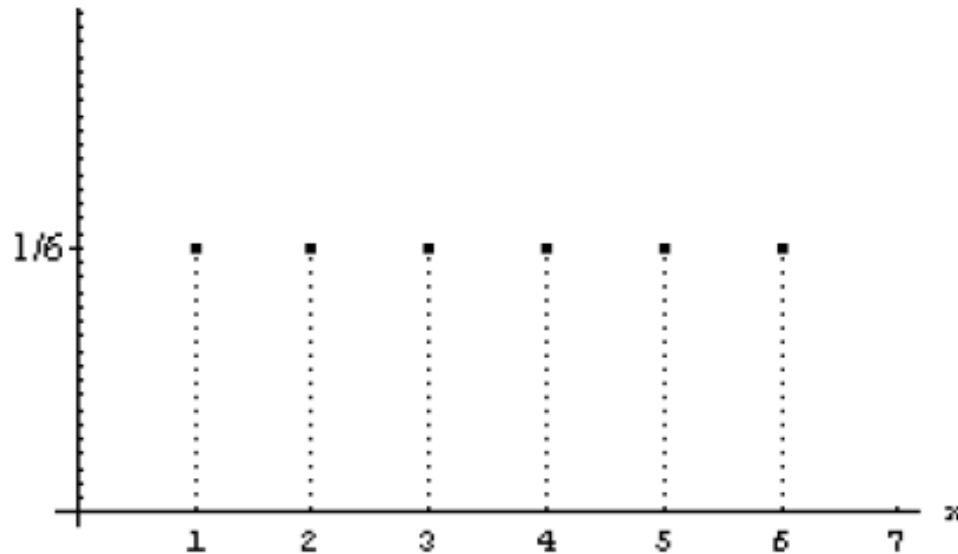
b) la suma de las probabilidades de todos los valores da 1.

$$\sum_{i \in \Omega} \mathbb{P}(X = x_i) = 1$$

Función de probabilidades(2)

54

• Ej:



$$P(X = x) = \left\{ \begin{array}{ll} 1/6 & x=1 \\ 1/6 & x=2 \\ 1/6 & x=3 \\ 1/6 & x=4 \\ 1/6 & x=5 \\ 1/6 & x=6 \\ 0 & \forall \text{ otro } x \end{array} \right\}$$

Función de distribución acumulada (v.a.d)

55

- Se la llama función de distribución acumulada porque indica la probabilidad "acumulada" por todos los valores con probabilidad no nula hasta x (partiendo de x_1):

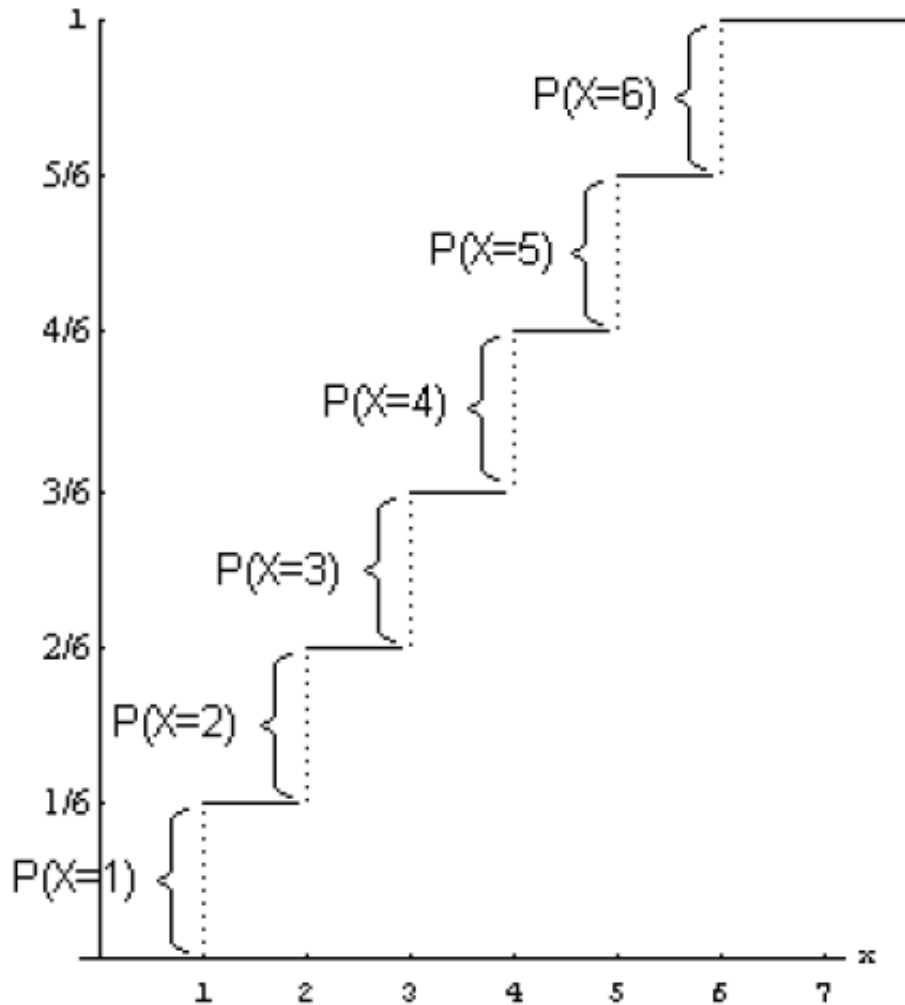
$$F(x) = \mathbb{P}(X \leq x) = \sum_{i=1}^x \mathbb{P}(X = x_i)$$

- La probabilidad acumulada comienza siendo cero (en $-\infty$) hasta que encuentra el primer valor con probabilidad no nula. A partir de ese valor, la probabilidad acumulada es la suma de las probabilidades de los puntos que encuentra hasta llegar al último valor con probabilidad no nula, a partir del cual la probabilidad acumulada vale uno.

Función de distribución acumulada (2)

56

• Ej:



$$F_X(x) = \left\{ \begin{array}{ll} 0 & x < 1 \\ 1/6 & 1 \leq x < 2 \\ 2/6 & 2 \leq x < 3 \\ 3/6 & 3 \leq x < 4 \\ 4/6 & 4 \leq x < 5 \\ 5/6 & 5 \leq x < 6 \\ 1 & x \geq 6 \end{array} \right.$$

Función de distribución acumulada(3)

57

- **Propiedades:**

$$\mathbb{P}(X < x) = F(x) - \mathbb{P}(X = x)$$

$$\mathbb{P}(X > x) = \sum_{i=x+1}^{\infty} \mathbb{P}(X = x_i) = 1 - F(x)$$

$$\mathbb{P}(a \leq X \leq b) = \sum_{i=a}^b \mathbb{P}(X = x_i) = F(b) - F(a) + \mathbb{P}(X = a)$$

Función de densidad de probabilidad (v.a.c)

58

- La **función de densidad de probabilidad (FDP)** en el caso continuo, representada como $f(x)$, se utiliza con el propósito de conocer cómo se distribuyen las probabilidades de un suceso o evento, en relación al resultado del suceso. Cumple:
- No negatividad: $f(x) \geq 0$
- El área bajo $f(x)$ es 1 en todo el EM: $\int_{\Omega} f(x) = 1$

Función de distribución acumulada (v.a.c)

59

- la función de distribución $F(x)$ es la integral de la función de densidad (de $-\infty$ hasta x):

$$F(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^x f(t)dt$$

- Notar que la probabilidad de que ocurra un valor en particular es cero:

$$\mathbb{P}(X \leq x) = \mathbb{P}(X < x)$$

Función de distribución acumulada (2)

60

Propiedades:

- Límites: $\lim_{x \rightarrow -\infty} F(x) = 0$ $\lim_{x \rightarrow \infty} F(x) = 1$
- $F(x)$ es monótonicamente creciente: si $b > a$ entonces $F(b) > F(a)$.
- Complemento: $\mathbb{P}(X > x) = 1 - F(x)$
- Segmento: $\mathbb{P}(a \leq X \leq b) = \int_a^b f(t) dt = F(b) - F(a)$