Tema 5. Probabilidad

5.1. Definición de probabilidad. Problema de la asignación de probabilidades.

La distinción entre *fenómenos determinísticos y aleatorios* conduce a plantearse el problema de la medida de la incertidumbre mostrada por estos últimos.

Conviene recordar algunas de las propiedades de las *frecuencias relativas* de *variables estadísticas* que van a proporcionar un excelente apoyo intuitivo para la **definición de la probabilidad**.
Consideremos una tabla estadística en la que se recogen los resultados del lanzamiento de un dado en 300 ocasiones:

x_i	n_i	f_i
1	45	0,15
2	57	0,19
3	51	0,17
4	48	0,16
5	54	0,18
6	45	0,15
total	*n=300*	1

1.- La frecuencia relativa de cualquier modalidad x_i es un número comprendido entre 0 y 1,

$$0 \le f_i \le 1.$$

2.- La frecuencia relativa de dos o más modalidades es la suma de las frecuencias relativas de cada una de las modalidades. Por ejemplo, la frecuencia relativa de obtener un número par es

$$f(par) = f\left(x_2 \cup x_4 \cup x_6\right) = f_2 + f_4 + f_6 = 0{,}19 + 0{,}16 + 0{,}15 = 0{,}50$$

3.- La frecuencia relativa de todas las modalidades (suma de todas las frecuencias relativas) es 1,

$$\sum_{i=1}^{6} f_i = 1$$

Llamaremos **suceso** a cada uno de los posibles resultados de un *experimento o fenómeno aleatorio*. Utilizaremos letras mayúsculas para referirnos a ellos.

En el experimento de lanzar el dado, son ejemplos de sucesos: *A*={*obtener par*}, *B*={*obtener impar*}, *C*={*obtener un 3*}, *D*={*obtener un número mayor o igual que 4*}, *E*={*obtener un número menor que 4*}... Utilizaremos estos sucesos como ejemplos para ilustrar las siguientes definiciones y operaciones sobre sucesos.

Al suceso formado por un único resultado se denomina **suceso elemental.** Para estos sucesos, además de la notación general con letras mayúsculas, utilizaremos la notación ω_i. Un ejemplo de suceso elemental sería *C*={*obtener un 3*}, ω_3 según esta última notación (notación análoga a la utilizada en variables estadísticas, x_i).

Se llama **suceso seguro** al suceso que siempre ocurre, está formado por todos los sucesos elementales, se nota Ω. En el ejemplo del dado $\Omega = \{1,2,3,4,5,6\}$.

Al conjunto vacío, que se nota $\varnothing$, se denomina **suceso imposible**.

Operaciones y relaciones con sucesos:

El suceso *A* **implica** el suceso *B* si siempre que ocurre *A* ocurre *B*. Lo notaremos $A \subseteq B$. En el ejemplo del dado: *C* implica *B*, $C \subseteq B$, $\{3\} \subseteq \{impares\}$.

Se define la **unión de dos sucesos** *A* y *B*, $A \cup B$, como el suceso que ocurre cuando ocurre el suceso *A* **o** el suceso *B*. En el ejemplo del dado: $B = \{impar\}$, $D = \{mayor\ o\ igual\ que\ 4\}$, $B \cup D = \{1,3,4,5,6\}$.

La unión de sucesos cumple las **propiedades conmutativa, asociativa e idempotente**:

$$A \cup B = B \cup A$$

$$A \cup (B \cup C) = (A \cup B) \cup C$$

$$A \cup A = A$$

Se define la **intersección de dos sucesos** *A* y *B*, $A \cap B$, como el suceso que ocurre cuando ocurre el suceso *A* **y** el suceso *B*. En el ejemplo del dado: $B = \{impar\}$, $D = \{mayor\ o\ igual\ que\ 4\}$, $B \cap D = \{5\}$.

La intersección de sucesos cumple las **propiedades conmutativa, asociativa e idempotente**:

$$A \cap B = B \cap A$$

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

$$A \cap A = A$$

Las operaciones de unión e intersección conjuntamente cumplen las **propiedades distributivas:**

$$A\cap(B\cup C)=(A\cap B)\cup(A\cap C)$$

$$A\cup(B\cap C)=(A\cup B)\cap(A\cup C)$$

y de **absorción:**

$$A\cup(A\cap B)=A$$

$$A\cap(A\cup B)=A$$

Se define la **diferencia de dos sucesos**, $A-B$, como el suceso formado por los sucesos elementales que son de *A* pero no de *B*. Para poder definir correctamente esta operación se ha de cumplir que $B\subseteq A$. En el ejemplo del dado: $B=\{impar\}$, $C=\{3\}$, $B-C=\{1,5\}$.

El **suceso contrario (o complementario)** de *A* es el que ocurre cuando no ocurre *A*. Se nota como $\overline{A}$. $\overline{A}=\Omega-A$. En el ejemplo del dado: $A=\{par\}$, $\overline{A}=\{impar\}$, $D=\{mayor\ o\ igual\ que\ 4\}$, $\overline{D}=\{menor\ que\ 4\}$.

Esta operación tiene las siguientes **propiedades**:

$$\overline{\overline{A}}=A$$

$$A\subseteq B \quad\Rightarrow\quad \overline{B}\subseteq\overline{A}$$

$$\overline{\varnothing}=\Omega \qquad \overline{\Omega}=\varnothing$$

$$A\cap\overline{A}=\varnothing \qquad A\cup\overline{A}=\Omega$$

$$\overline{A_1\cup\ldots\cup A_n}=\overline{A_1}\cap\ldots\cap\overline{A_n}$$

$$\overline{A_1\cap\ldots\cap A_n}=\overline{A_1}\cup\ldots\cup\overline{A_n}$$

Los sucesos *A* y *B* son **incompatibles** si no pueden ocurrir ambos simultáneamente, $A\cap B=\varnothing$. Un suceso y su complementario son siempre sucesos incompatibles. En el ejemplo del dado: $A=\{par\}$ y $C=\{3\}$ son incompatibles.

Representaremos los resultados de un experimento aleatorio mediante el par $(\Omega,\mathcal{A})$, donde Ω es el conjunto de todos los sucesos elementales de un experimento aleatorio y $\mathcal{A}$ el conjunto de todos los sucesos (elementales y no elementales).

Definición de probabilidad.

Definimos una medida de probabilidad sobre $(\Omega,\mathcal{A})$ como una aplicación:

$$(\Omega, \mathscr{A}) \xrightarrow{\mathsf{P}} \mathrm{R}$$

$$A \in \mathscr{A} \rightarrow \mathsf{P}(A) \in \mathrm{R}$$

Verificando las siguientes **condiciones**:

1.- $\mathsf{P}(A) \geq 0 \qquad \forall A \in \mathscr{A}$

2.- Sean $A_1, ..., A_n$ sucesos incompatibles, $\mathsf{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mathsf{P}\left(A_i\right)$

3.- $\mathsf{P}\left(\Omega\right) = 1$

Obsérvese la analogía entre estas tres condiciones y las propiedades de las frecuencias relativas que destacábamos al comienzo de la lección.

Propiedades:

1.- $\mathsf{P}(\overline{A}) = 1 - \mathsf{P}(A)$

2.- Un caso particular de la propiedad anterior es $\mathsf{P}(\varnothing) = 1 - \mathsf{P}(\Omega) = 1 - 1 = 0$

3.- Si $A \subseteq B \quad \Rightarrow \quad \mathsf{P}(A) \leq \mathsf{P}(B)$

4.- Probabilidad de la unión de sucesos compatibles: $\mathsf{P}(A \cup B) = \mathsf{P}\left(A\right) + \mathsf{P}\left(B\right) - \mathsf{P}\left(A \cap B\right)$

5.- Consecuencia de la anterior propiedad es la subaditividad de P: $\mathsf{P}\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} \mathsf{P}\left(A_i\right)$

Problema de la asignación de probabilidades.

Estas son las propiedades que tiene la medida de la probabilidad pero la teoría no especifica nada más sobre los valores numéricos concretos que toma la función $\mathsf{P}(A)$.

Si el número de sucesos elementales, ω_i, es finito, es suficiente asignar probabilidades a los sucesos elementales, $\mathsf{P}\left(\omega_i\right)$, y definir la probabilidad de un suceso *A* como

$$\mathsf{P}\left(A\right) = \sum_{\omega_i \in A} \mathsf{P}\left(\omega_i\right)$$

Si además de ser finito el conjunto de sucesos elementales, $\Omega = \left\{\omega_1, ..., \omega_i, ..., \omega_n\right\}$, suponemos que los sucesos elementales son *equiprobables*, $\mathsf{P}\left(\omega_i\right) = \frac{1}{n} \quad \forall \omega_i$, obtenemos la conocida **regla de Laplace**

$$\mathsf{P}\left(A\right) = \sum_{\omega_i \in A} \mathsf{P}\left(\omega_i\right) = \sum_{\omega_i \in A} \frac{1}{n} = \frac{k}{n} = \frac{casos\ favorables}{casos\ posibles}$$

siendo *k* el número de sucesos elementales en *A*.

Es muy importante no confundir el anterior razonamiento con un *concepto universal de probabilidad*. A esta **concepción de la probabilidad** se le denomina **clásica o de Laplace.**

Según esta concepción de la probabilidad, en el ejemplo del dado se supondría que todas las caras (sucesos elementales) tienen la misma probabilidad de aparición y esta es:

$$\mathrm{P}(\omega_i) = \frac{1}{6} = 0{,}1\widehat{6}$$

En la semejanza entre las propiedades de las frecuencias relativas y la definición de probabilidad se basa la **concepción frecuentista** de la probabilidad. Considera que la probabilidad de un suceso es el valor límite en torno al cual se estabiliza la frecuencia relativa de ese suceso cuando se observa indefinidamente el fenómeno aleatorio. En el ejemplo del dado, si continuamos observando los resultados cuando lo lanzamos muchas veces (más de las 300 iniciales), se observará que el valor de las frecuencias relativas ya no cambia apenas y se estabiliza en torno a un valor que tomaremos como medida de la probabilidad de aparición de dicho suceso.

x_i	n_i	f_i
1	45	0,15
2	57	0,19
3	51	0,17
4	48	0,16
5	54	0,18
6	45	0,15
total	*n=300*	1

→

x_i	n_i	f_i
1	456	0,152
2	561	0,187
3	516	0,172
4	483	0,161
5	534	0,178
6	450	0,150
total	*n=3000*	1

Cada una de estas concepciones tiene sus ventajas e inconvenientes. Por ejemplo, la concepción clásica no nos valdría para modelar el comportamiento de un dado, moneda, … que estuviesen trucados para conseguir con mayor probabilidad determinados resultados o sencillamente mal construidos. En contra de la concepción frecuentista se podría esgrimir el argumento de que multitud de fenómenos aleatorios no pueden repetirse de forma indefinida, especialmente fenómenos de tipo social y económico (elecciones, valores bursátiles, competiciones deportivas, …). Esta y otras dificultades han llevado a considerar interpretaciones alternativas de la probabilidad como la **concepción subjetiva** en la que se considera un grado de creencia o confianza en la ocurrencia de un hecho (la probabilidad de que la selección española gane el próximo mundial de fútbol es de 4/5).

►***EJEMPLO 5.1***

Calcule las probabilidades de los siguientes sucesos en el lanzamiento de un dado según la concepción clásica de Laplace y según la concepción frecuentista (tomando para este último caso como valores estables de las frecuencias los observados en los 3000 lanzamientos). Expréselo de distintas formas donde se pueda.

A={obtener par}, B={obtener impar}, C={obtener un 3}, D={obtener un número mayor o igual que 4}, E={obtener un número menor que 4}.

Solución:

Concepción clásica:

Concepción frecuentista:

◄

5.2. Definición de probabilidad condicionada. Sucesos dependientes e independientes.

Probabilidad condicionada.

En determinadas situaciones es necesario calcular la probabilidad de sucesos dada la condición adicional de que cierto suceso ha ocurrido. A este tipo de **probabilidad** se denomina **condicionada.**

Se define la probabilidad del suceso *B* condicionado a *A*, $\mathrm{P}\left(B/A\right)$, como la probabilidad de que ocurra *B* supuesto que ha ocurrido *A*. Dicha *probabilidad condicionada* del suceso *B* al suceso *A* es igual a:

$$\mathrm{P}\left(B/A\right)=\frac{\mathrm{P}\left(B\cap A\right)}{\mathrm{P}\left(A\right)} \qquad \mathrm{P}\left(A\right)>0$$

Como en la probabilidad no condicionada, aquí se sigue tomando como modelo las frecuencias relativas. Supongamos el siguiente ejemplo sobre datos de los estudios completados por una población de 1000 habitantes

	Enseñanza obligatoria	Bachillerato	Universitarios	TOTAL
hombres	250	160	40	450
mujeres	350	140	60	550
TOTAL	600	300	100	1000

Entre los hombres (*H* sería el suceso conocido o que ha ocurrido) la frecuencia relativa de los que completaron los estudios universitarios (*U* sería el suceso sobre el que se quiere calcular la probabilidad, en este caso la frecuencia) es:

$$f\left(U/H\right)=\frac{40}{450}=\frac{\frac{40}{1000}}{\frac{450}{1000}}=\frac{f\left(U\cap H\right)}{f\left(H\right)}$$

Mientras que la frecuencia relativa de los que completaron los estudios universitarios (sin imponer ninguna condición sobre el sexo) es:

$$f\left(U\right)=\frac{100}{1000}$$

De forma análoga a como se ha definido $\mathrm{P}\left(B/A\right)$ se define $\mathrm{P}\left(A/B\right)$ como:

$$\mathrm{P}\left(A/B\right)=\frac{\mathrm{P}\left(A\cap B\right)}{\mathrm{P}\left(B\right)} \qquad \mathrm{P}\left(B\right)>0$$

De las definiciones de $\mathrm{P}\left(B/A\right)$ y $\mathrm{P}\left(A/B\right)$ se deduce la **fórmula de la probabilidad compuesta** o **fórmula del producto de probabilidades:**

$$\mathrm{P}\left(A\cap B\right)=\mathrm{P}\left(A\right)\mathrm{P}\left(B/A\right)=\mathrm{P}\left(B\right)\mathrm{P}\left(A/B\right)$$

Para tres sucesos el anterior resultado se puede enunciar de 3!=6 formas alternativas, una de ellas es:

$$\mathrm{P}\left(A\cap B\cap C\right)=\mathrm{P}\left(A\right)\mathrm{P}\left(B/A\right)\mathrm{P}\left(C/A\cap B\right)$$

y para *n* sucesos una de sus posibles formas es:

$$\mathrm{P}\left(A_1\cap A_2\cap\ldots\cap A_n\right)=\mathrm{P}\left(A_1\right)\mathrm{P}\left(A_2/A_1\right)\mathrm{P}\left(A_3/A_1\cap A_2\right)\ldots\mathrm{P}\left(A_n/A_1\cap A_2\cap\ldots\cap A_{n-1}\right)$$

Sucesos dependientes e independientes.

En general $\mathrm{P}\left(B/A\right) \neq \mathrm{P}(B)$ y decimos en tal caso que *B* **depende** de *A*. Si $\mathrm{P}\left(B/A\right) = \mathrm{P}(B)$ diremos que *B* es **independiente** de *A*, es decir, que haya ocurrido *A* no modifica la probabilidad de *B*.

En el caso de **independencia:**

$$\mathrm{P}\left(B/A\right) = \frac{\mathrm{P}(B \cap A)}{\mathrm{P}(A)} = \mathrm{P}(B) \qquad \Rightarrow \qquad \mathrm{P}(A \cap B) = \mathrm{P}(A)\mathrm{P}(B)$$

Y como consecuencia se sigue que también:

$$\mathrm{P}\left(A/B\right) = \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)} = \frac{\mathrm{P}(A)\mathrm{P}(B)}{\mathrm{P}(B)} = \mathrm{P}(A)$$

luego si *B* es independiente de *A*, entonces *A* es independiente de *B*.

Según todo lo anterior, se toma como definición de que los sucesos *A* y *B* son **independientes** la siguiente expresión:

$$\mathrm{P}(A \cap B) = \mathrm{P}(A)\mathrm{P}(B)$$

La noción de independencia es fundamental en la teoría de la probabilidad. La mayoría de los resultados en probabilidad se obtienen bajo la hipótesis de independencia.

►***EJEMPLO 5.2***

En un dado con igual probabilidad de aparición para todas sus caras comprobar que los siguientes sucesos, dos a dos, son dependientes.

A={obtener par}, B={obtener impar}, C={obtener un 3}, D={obtener un número mayor o igual que 4}, E={obtener un número menor que 4}.

◄

5.3. Fórmula de la probabilidad total. Fórmula de Bayes.

Sean los sucesos $A_1, ..., A_n$ una *partición* de Ω, es decir

$$A_i \cap A_j = \varnothing \qquad \forall i \neq j \qquad ; \qquad \bigcup_{i=1}^{n} A_i = \Omega$$

en este contexto, la **fórmula de la probabilidad total** afirma que:

$$\mathrm{P}(B) = \sum_{i=1}^{n} \mathrm{P}\left(B/A_i \right) \mathrm{P}(A_i)$$

Las probabilidades condicionadas de un suceso cualquiera *B* sobre cada uno de los elementos de la partición, $\mathrm{P}\left(B/A_i \right)$, son normalmente conocidas. No es así con las probabilidades $\mathrm{P}\left(A_i/B \right)$. Veamos cómo resolver este problema:

por la fórmula de la probabilidad compuesta $\mathrm{P}(A_k \cap B) = \mathrm{P}(B)\mathrm{P}\left(A_k/B \right) = \mathrm{P}(A_k)\mathrm{P}\left(B/A_k \right)$

y según la fórmula de la probabilidad total $\mathrm{P}(B) = \sum_{i=1}^{n} \mathrm{P}\left(B/A_i \right) \mathrm{P}(A_i)$

de ambas expresiones se sigue:

Esta última expresión se conoce como **fórmula de Bayes**.

►***EJEMPLO 5.3***

En el bar de la Facultad de CC. EE. y EE. el 90% de los clientes son estudiantes. Se sabe que el 45% de los clientes toman café y que dado un estudiante, la probabilidad de que éste tome café es del 30%.

a) ¿Cuál es la probabilidad de que un cliente elegido al azar sea estudiante y tome café?
b) Se elige al azar un cliente que toma café, ¿cuál es la probabilidad de que sea estudiante?
c) ¿Son independientes los sucesos *tomar café* y *ser estudiante*?

Solución:

◄

►***EJEMPLO 5.4***

En una ciudad se publican tres periódicos (*A*, *B* y *C*). Se sabe que un 60% de la población está suscrito al periódico *A*, un 40% al *B*, un 30% al *C*, un 20% a *A* y *B*, un 15% a *A* y *C*, un 25% a *B* y *C*, y un 10% a los tres periódicos. ¿Qué parte de la población está suscrita al menos a un periódico?

Solución:

En este ejemplo se hace una interpretación de la probabilidad según la concepción frecuentista.

◄

►***EJEMPLO 5.5***

La producción de una factoría se realiza en cuatro máquinas, $M_1, M_2, M_3 \; y \; M_4$. Diariamente la producción de cada una de las máquinas es la siguiente:

M_1	M_2	M_3	M_4	TOTAL
600	500	350	250	1700

Además sabemos que los porcentajes de piezas defectuosas producidas por cada una de las máquinas son:

M_1	M_2	M_3	M_4
4%	3,5%	4,6%	2%

a) Si las piezas se almacenan conjuntamente, ¿cuál es la probabilidad de que al seleccionar una pieza al azar ésta sea defectuosa?

b) Se ha seleccionado una pieza defectuosa, ¿cuál es la probabilidad de que haya sido producida en la máquina M_2?

►***EJEMPLO 5.6***

En una empresa el 8% de los hombres y el 4,3% de las mujeres ganan más de 25000€ al año. Se sabe que el porcentaje de mujeres en la empresa es del 47%. Se selecciona al azar un empleado que gana menos de 25000€, ¿cuál es la probabilidad de que sea mujer?

Resumen del Tema 5: Probabilidad

Probabilidad condicionada	
Fórmula de la probabilidad compuesta (fórmula del producto de probabilidades)	
Sucesos independientes	
Fórmula de probabilidad Total	
Fórmula de Bayes	

Preguntas de teoría del Tema 5

1. Definición de probabilidad. Propiedades

2. Definición de probabilidad condicionada. Fórmula de la probabilidad compuesta. Sucesos dependientes e independientes.

3. Fórmula de la probabilidad total. Fórmula de Bayes.

Tema 6. Variables aleatorias y distribuciones de probabilidad

Algunas medidas de variables estadísticas (Temas 1 y 2) y variables aleatorias (Tema 6)

	Temas 1 Y 2	**Tema 6**	
		Discreta	**Continua**
Media aritmética (Esperanza Matemática)			
Varianza			
Desviación típica			
Momentos no centrados			
Momentos centrados respecto la media aritmética			
Moda			
Mediana			
Cuantiles			
Coeficiente de Variación			
Coeficiente de asimetría			
Coeficiente de apuntamiento o curtosis			

6.1. Concepto de variable aleatoria y distribución de probabilidad.

Sea Ω el conjunto de sucesos elementales, ω_i, de un fenómeno aleatorio. Se define una **variable aleatoria *X*** como una función sobre Ω en R

$$\Omega \xrightarrow{X} \mathrm{R}$$

$$\omega_i \rightarrow X(\omega_i) = x_i \in \mathrm{R}$$

Es decir, mediante una variable aleatoria transformamos los resultados de un fenómeno aleatorio en valores numéricos que miden aquella característica del fenómeno que nos interesa destacar (*en el siguiente ejemplo la característica es el número de cruces en tres lanzamientos de una moneda*).

Las probabilidades de que la variable aleatoria X tome los distintos valores dependen de las probabilidades, $\mathrm{P}(\omega_i)$, asociadas al fenómeno aleatorio sobre el que se ha definido.

Se denomina **distribución de probabilidad** de la variable aleatoria *X* a las probabilidades asociadas a los distintos valores que puede tomar la variable aleatoria, $p_i = \mathrm{P}[X = x_i]$.

Toda distribución de probabilidad ha de cumplir:

- $p_i \geq 0 \quad \forall i$
- $\sum_{\forall i} p_i = 1$

(*Esta definición es válida para variables que toman un conjunto discreto de valores, como el siguiente ejemplo, sin embargo cuando la variable aleatoria puede tomar un conjunto continuo de valores la distribución de probabilidad no está caracterizada por probabilidades de valores aislados de la variable como veremos en el epígrafe 6.2*).

►EJEMPLO 6.1

Represente la variable aleatoria *X=número de cruces en tres lanzamientos de una moneda* y obtenga la distribución de probabilidad asociada.

Solución:

C=cara, +=cruz

$$\Omega \xrightarrow{X} \mathrm{R}$$

$$\begin{array}{lcl} \mathrm{CCC} & \rightarrow & X(CCC)=0 \\ +\mathrm{CC} & \rightarrow & X(+CC)=1 \\ \mathrm{C+C} & \rightarrow & 1 \\ \mathrm{CC+} & \rightarrow & 1 \\ \mathrm{++C} & \rightarrow & 2 \\ \mathrm{+C+} & \rightarrow & 2 \\ \mathrm{C++} & \rightarrow & 2 \\ \mathrm{+++} & \rightarrow & 3 \end{array}$$

Suponemos que la probabilidad de cara y cruz es la misma, $\frac{1}{2}$. Por tanto la probabilidad de cada uno de los sucesos elementales de Ω es $\left(\frac{1}{2}\right)^3=\frac{1}{8}$ dado que son independientes los resultados de los tres lanzamientos de la moneda

$$\mathrm{P}(CCC)=\mathrm{P}(C\cap C\cap C)=\mathrm{P}(C)\mathrm{P}(C)\mathrm{P}(C)=\frac{1}{2}\frac{1}{2}\frac{1}{2}=\frac{1}{8} \ \ldots \ \mathrm{P}(+++)=\mathrm{P}(+)\mathrm{P}(+)\mathrm{P}(+)=\left(\frac{1}{2}\right)^3=\frac{1}{8}$$

$$p_0=\mathrm{P}[X=0]=\mathrm{P}[(CCC)]=\frac{1}{8}$$

$$p_1=\mathrm{P}[X=1]=\mathrm{P}[(+CC)\cup(C+C)\cup(CC+)]=\frac{1}{8}+\frac{1}{8}+\frac{1}{8}=\frac{3}{8}$$

$$p_2=\mathrm{P}[X=2]=\mathrm{P}[(++C)\cup(+C+)\cup(C++)]=\frac{1}{8}+\frac{1}{8}+\frac{1}{8}=\frac{3}{8}$$

$$p_3=\mathrm{P}[X=3]=\mathrm{P}[(+++)]=\frac{1}{8}$$

La **distribución de probabilidad** se representa en una tabla como la que sigue

x_i	p_i
0	1/8
1	3/8
2	3/8
3	1/8

Como puede comprobarse $\sum_{i=0}^{3} p_i=1$. ◄

Si la variable aleatoria toma un conjunto de valores aislados (como el ejemplo 6.1 donde la variable aleatoria sólo toma los valores 0, 1, 2 y 3) diremos que es una **variable aleatoria discreta.**

Si la variable aleatoria toma un conjunto continuo de valores (por ejemplo la duración de una bombilla, que puede tomar cualquier valor a partir de 0) diremos que es una **variable aleatoria continua**.

6.2. Variables aleatorias discretas

6.2.1. Función de distribución

Se define la **función de distribución** de una variable aleatoria *X* como:

$$F(x) = \mathrm{P}\left[X \le x\right] = \mathrm{P}\left[\omega \in \Omega / X(\omega) \le x\right] \qquad \forall x \in \mathrm{R}$$

Propiedades:

- La función de distribución es una función no decreciente, es decir:

  Si $x_i < x_j \quad \Rightarrow \quad F(x_i) \le F(x_j)$

- $F(-\infty) = 0 \qquad F(+\infty) = 1$
- $\mathrm{P}\left[X > x\right] = 1 - F(x)$
- $\mathrm{P}\left[x_1 < X \le x_2\right] = F(x_2) - F(x_1)$

►EJEMPLO 6.2

Calcule la función de distribución asociada a la variable aleatoria *X* del ejemplo 6.1. Represente su gráfica.

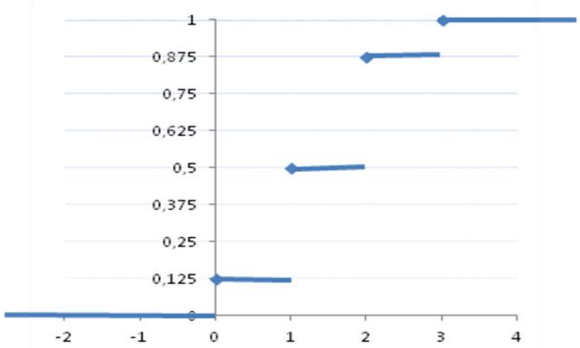
Solución:

La distribución de probabilidad del ejemplo 6.1 está dada por:

x_i	p_i
0	1/8
1	3/8
2	3/8
3	1/8

Su función de distribución, por tanto, será:

y su representación gráfica:



6.2.2. Valor esperado de una variable aleatoria discreta. Momentos

Se define el **valor esperado, esperanza matemática o media** de una **variable aleatoria discreta** X como:

$$E[X] = \sum_{i=1}^{\infty} x_i p_i$$

Esta definición puede generalizarse para cualquier potencia, de orden *r*, de la variable aleatoria. Son los **momentos no centrados o respecto al origen**:

$$\alpha_r = E\left[X^r\right] = \sum_{i=1}^{\infty} x_i^r p_i$$

La **media** es el momento no centrado de orden 1, adopta las siguientes notaciones en distintos contextos:

$$E[X] = \alpha_1 = \alpha = \mu$$

Se definen los **momentos centrados, respecto a la esperanza o respecto a la media** como:

$$\mu_r = E\left[\left(X - E[X]\right)^r\right] = \sum_{i=1}^{\infty}\left(x_i - E[X]\right)^r p_i$$

La **varianza** es el momento centrado de orden 2, μ_2, se suele notar como $V[X]$ o σ^2. Su raíz cuadrada es la **desviación típica,** y la denotaremos como $D[X]$ o σ.

Las mismas relaciones que se encontraron entre los momentos en Estadística Descriptiva siguen siendo válidas, destacando por su interés las siguientes:

$$\sigma^2 = \mu_2 = \alpha_2 - \alpha_1^2$$

$$\mu_3 = \alpha_3 - 3\alpha_1\alpha_2 + 2\alpha_1^3$$

$$\mu_4 = \alpha_4 - 4\alpha_1\alpha_3 + 6\alpha_1^2\alpha_2 - 3\alpha_1^4$$

Propiedades de la media y de la varianza.

- $E[aX + b] = aE[X] + b$
- $V[aX + b] = a^2 V[X]$

Las propiedades anteriores muestran como le afecta al valor de la media y de la varianza un cambio de origen y escala sobre la variable aleatoria.

Ejemplos

►***EJEMPLO 6.3***

Calcule la media y la varianza de la distribución de probabilidad del ejemplo 6.2.

Solución:

x_i	p_i	$x_i p_i$	$x_i^2 p_i$
0	1/8	0	0
1	3/8	0,375	0,375
2	3/8	0,750	1,500
3	1/8	0,375	1,125
total		1,500	3

6.2.3. Otras medidas de posición, dispersión y forma

- Moda

Se define la moda, ***Mo***, como aquel valor de la variable aleatoria donde la distribución de probabilidad alcanza su máximo.

- Mediana y cuantiles

- Coeficiente de variación

Es la medida de dispersión relativa más utilizada. Nos permite comparar la dispersión de distintas variables aleatorias.

- Coeficiente de asimetría

- Coeficiente de apuntamiento

►***EJEMPLO 6.4***

Calcule la moda, mediana, percentil 75 y coeficiente de variación de la distribución de probabilidad del ejemplo 6.2.

Solución:

6.3. Variables aleatorias continuas

6.3.1. Función de distribución y función de densidad

La **función de distribución** de una **variable aleatoria continua** es una **función continua.** La derivada de la función de distribución de una variable aleatoria continua se denomina **función de densidad**

$$F'(x) = f(x)$$

Como consecuencia de las propiedades de la función de distribución, toda función de densidad cumple que:

- $f(x) \geq 0 \qquad \forall x \in \mathrm{R}$
- $\int_{-\infty}^{+\infty} f(x)\, dx = 1$

La probabilidad de que la variable aleatoria X tome valores en un determinado conjunto es igual a la integral definida de la función de densidad sobre dicho conjunto.

En particular:

$$\mathrm{P}[X = a] = \int_{a}^{a} f(x)\, dx = 0$$

$$\mathrm{P}[a < X < b] = \int_{a}^{b} f(x)\, dx$$

$$\mathrm{P}[a < X] = \int_{a}^{+\infty} f(x)\, dx$$

$$\mathrm{P}[X < a] = \int_{-\infty}^{a} f(x)\, dx$$

y por tanto la función de distribución se puede obtener a partir de la función de densidad:

$$F(x) = \mathrm{P}[X \leq x] = \int_{-\infty}^{x} f(t)\, dt$$

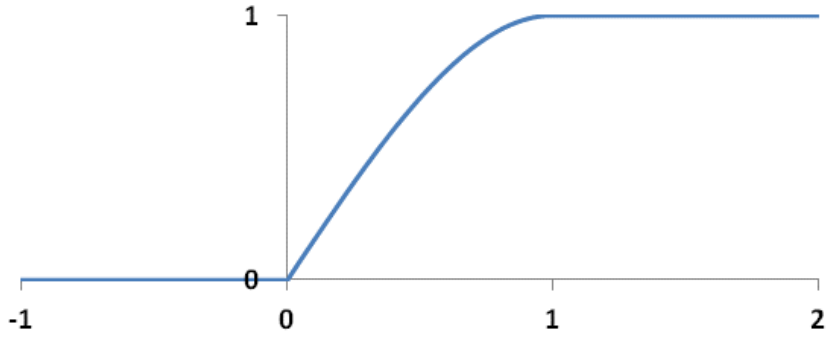
Como puede verse, en variables aleatorias continuas la probabilidad de un valor aislado es cero por lo que no tiene sentido definir para ellas el concepto de distribución de probabilidad tal y como apareció en el epígrafe 6.1. En ***variables aleatorias continuas*** *su* ***distribución de probabilidad*** *está definida por su* ***función de densidad****.*

►***EJEMPLO 6.5***

Sea la función de distribución de la variable aleatoria *X*:

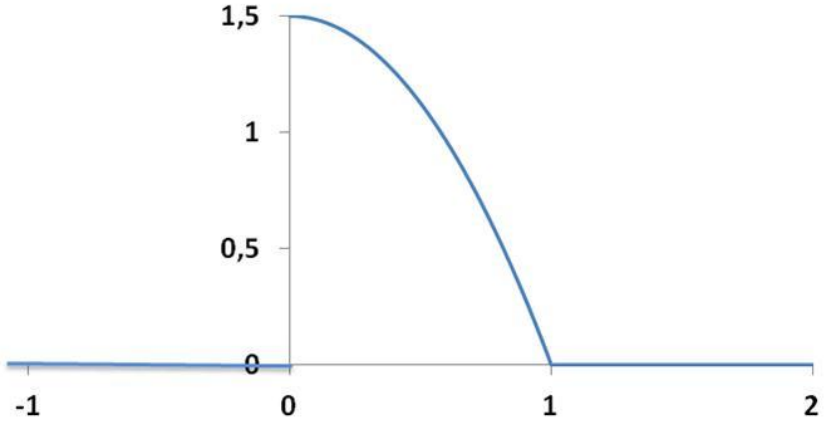
$$F(x)=\begin{cases} 0 & ,x<0 \\ \dfrac{3x-x^3}{2} & ,0\le x\le 1 \\ 1 & ,1<x \end{cases}$$

Como puede comprobarse es una función continua cuya gráfica es:



a) Obtener la correspondiente función de densidad

Derivando la función de distribución obtenemos la función de densidad de la variable aleatoria *X*



b) Comprobar que $f(x)$ cumple las dos propiedades que caracterizan a toda función de densidad.

c) Calcular la función de distribución a partir de la función de densidad

d) Calculemos las probabilidades de que $\frac{1}{3} < x < \frac{1}{2}$ y de que $\frac{1}{2} < x$.

Lo haremos primero con la función de densidad y posteriormente con la función de distribución.

6.3.2. Valor esperado de una variable aleatoria continua. Momentos

- Valor esperado de una variable aleatoria continua *X*

- Momento no centrado para una variable aleatoria continua X

- Momento centrado para una variable aleatoria continua X

- Propiedades de la media y la varianza

►***EJEMPLO 6.6***

Calcule la media y la varianza de la distribución de probabilidad del ejemplo 6.5

$$f(x)=\begin{cases} 0 & , x<0 \\ \dfrac{3-3x^2}{2}=\dfrac{3}{2}(1-x^2) & , 0\le x\le 1 \\ 0 & , 1<x \end{cases}$$

6.3.3. Otras medidas de posición, dispersión y forma

Moda.

Se define la moda, ***Mo***, como aquel valor de la variable aleatoria donde la función de densidad alcanza su máximo

$$f(Mo) \geq f(x) \qquad \forall x \in \mathrm{R}$$

La variable aleatoria puede tener una o varias modas según se alcance el máximo en uno o más puntos.

Mediana, cuartiles y percentiles.

Se define la mediana, ***Me,*** como aquel valor de la variable aleatoria tal que

$$F(Me) = \mathrm{P}\left[X \leq Me\right] = \frac{1}{2}$$

De forma análoga, se definen los cuartiles Q_i y los percentiles P_i como aquellos valores de la variable aleatoria que verifican:

$$F(Q_i) = \mathrm{P}\left[X \leq Q_i\right] = \frac{i}{4}$$

$$F(P_i) = \mathrm{P}\left[X \leq P_i\right] = \frac{i}{100}$$

Coeficiente de variación.

Coeficiente de asimetría.

Coeficiente de apuntamiento o curtosis.

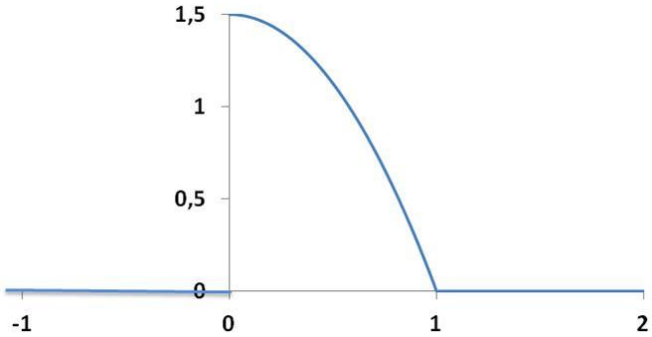
►***EJEMPLO 6.7***

Calcule la moda, mediana, percentil 75 y coeficiente de variación de la distribución de probabilidad del ejemplo 6.5

Solución:

Para hallar la **moda** en distribuciones de probabilidad continuas hay que calcular el punto o puntos donde la función de densidad alcanza su máximo.

$$f(x)=\begin{cases} 0 & ,x<0 \\ \dfrac{3-3x^2}{2}=\dfrac{3}{2}(1-x^2) & ,0\le x\le 1 \\ 0 & ,1<x \end{cases}$$

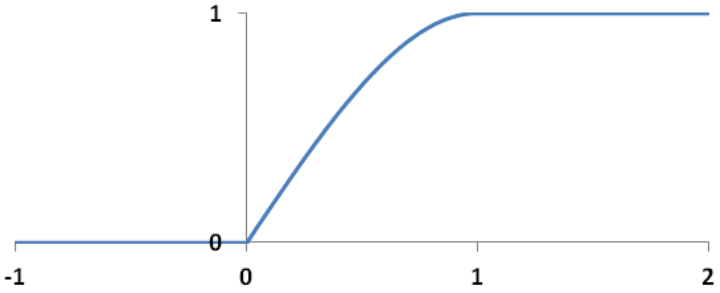


Como puede observarse en la gráfica de la función de densidad, ésta alcanza su máximo en Mo=0.

Para obtener la **mediana y el percentil 75** utilizaremos la función de distribución

$$F(Me)=\mathrm{P}\left[X\le Me\right]=\frac{1}{2} \qquad F(P_{75})=\mathrm{P}\left[X\le P_{75}\right]=0,75$$

$$F(x)=\begin{cases} 0 & ,x<0 \\ \dfrac{3x-x^3}{2} & ,0\le x\le 1 \\ 1 & ,1<x \end{cases}$$



$$\frac{3x-x^3}{2}=\frac{1}{2} \quad\Rightarrow\quad 0=x^3-3x+1 \quad\Rightarrow\quad x\cong 0,3473$$

$$\frac{3x-x^3}{2}=0,75=\frac{3}{4} \quad\Rightarrow\quad 0=2x^3-6x+3 \quad\Rightarrow\quad x\cong 0,557875$$

(Las anteriores ecuaciones no son fáciles de resolver a mano, se han resuelto con la ayuda del ordenador).

Para calcular el **coeficiente de variación** utilizamos los valores de la media y varianza calculados en el ejemplo 6.4

$$CV=\frac{\sigma}{E\left[X\right]}=\frac{\sqrt{0,059375}}{0,375}=\frac{0,24367}{0,375}=0,6498$$

◄

6.4. Variables aleatorias bidimensionales. Independencia de variables aleatorias.

Sea Ω el conjunto de sucesos elementales, ω_i, de un fenómeno aleatorio. Se define una **variable aleatoria bidimensional** (X,Y) como una función sobre Ω en $\mathrm{R}^2 = \mathrm{R}\times\mathrm{R}$.

$$\begin{array}{ccc} & (X,Y) & \\ \Omega & \rightarrow & \mathrm{R}\times\mathrm{R} \\ \omega_i & \rightarrow & (X(\omega_i), Y(\omega_i)) \end{array}$$

Es decir, mediante una variable aleatoria bidimensional transformamos los resultados de un fenómeno aleatorio en dos valores numéricos que miden aquellas dos características del fenómeno que nos interesan destacar. Una variable aleatoria bidimensional es un par de variables aleatorias unidimensionales definidas sobre el mismo fenómeno aleatorio que determina a Ω.

Las probabilidades de que la variable aleatoria bidimensional (X,Y) tome los distintos valores dependen de las probabilidades, $\mathrm{P}(\omega_i)$, asociadas al fenómeno aleatorio sobre el que se han definido.

Se denomina **distribución de probabilidad bidimensional** de las variables aleatorias (X,Y) a las probabilidades asociadas a los distintos valores que pueden tomar conjuntamente dichas variables aleatorias, $p_{ij} = \mathrm{P}\left[(X,Y) = (x_i, y_j)\right] = \mathrm{P}\left[X = x_i, Y = y_j\right]$.

Toda distribución de probabilidad bidimensional cumple (*como en el caso unidimensional*):

- $p_{ij} \geq 0 \quad \forall i, j$
- $\sum_{\forall i}\sum_{\forall j} p_{ij} = 1$

(*Esta definición es válida cuando ambas variables son discretas, sin embargo cuando las variables aleatorias son continuas la distribución de probabilidad bidimensional está caracterizada por la función de densidad bidimensional como se verá más adelante*).

►***EJEMPLO 6.8***

Veamos un sencillo ejemplo de variable aleatoria bidimensional discreta.

X=número de cruces en tres lanzamientos de una moneda.

Y=número de caras en los dos primeros lanzamientos de la moneda.

C=cara, +=cruz. La variable aleatoria bidimensional (X,Y) está definida por la siguiente aplicación:

$$\Omega \xrightarrow{(X,Y)} \mathrm{R}\times\mathrm{R}$$

$$\begin{array}{lcl}
\mathrm{CCC} & \rightarrow & (X(CCC),Y(CCC))=(0,2) \\
+\mathrm{CC} & \rightarrow & (X(+CC),Y(+CC))=(1,1) \\
\mathrm{C+C} & \rightarrow & (1,1) \\
\mathrm{CC+} & \rightarrow & (1,2) \\
\mathrm{++C} & \rightarrow & (2,0) \\
\mathrm{+C+} & \rightarrow & (2,1) \\
\mathrm{C++} & \rightarrow & (2,1) \\
\mathrm{+++} & \rightarrow & (3,0)
\end{array}$$

Suponemos que la probabilidad de cara y cruz es la misma, $\frac{1}{2}$. Por tanto la probabilidad de cada uno de los sucesos elementales de Ω es $\left(\frac{1}{2}\right)^3=\frac{1}{8}$ dado que son independientes los resultados de los tres lanzamientos de la moneda

$$\mathrm{P}(CCC)=\mathrm{P}(C\cap C\cap C)=\mathrm{P}(C)\mathrm{P}(C)\mathrm{P}(C)=\frac{1}{2}\frac{1}{2}\frac{1}{2}=\frac{1}{8}\ \ldots\ \ \mathrm{P}(+++)=\mathrm{P}(+)\mathrm{P}(+)\mathrm{P}(+)=\left(\frac{1}{2}\right)^3=\frac{1}{8}$$

$$p_{02}=\mathrm{P}[X=0,Y=2]=\mathrm{P}[(CCC)]=\frac{1}{8}$$

$$p_{11}=\mathrm{P}[X=1,Y=1]=\mathrm{P}[(+CC)\cup(C+C)]=\frac{1}{8}+\frac{1}{8}=\frac{2}{8}=\frac{1}{4}$$

$$p_{12}=\mathrm{P}[X=1,Y=2]=\mathrm{P}[(CC+)]=\frac{1}{8}$$

$$p_{20}=\mathrm{P}[X=2,Y=0]=\mathrm{P}[(++C)]=\frac{1}{8}$$

$$p_{21}=\mathrm{P}[X=2,Y=1]=\mathrm{P}[(C++)\cup(+C+)]=\frac{1}{8}+\frac{1}{8}=\frac{2}{8}=\frac{1}{4}$$

$$p_{30}=\mathrm{P}[X=3,Y=0]=\mathrm{P}[(+++)]=\frac{1}{8}$$

Las anteriores probabilidades se suelen recoger en una tabla de doble entrada (como l*as variables estadísticas bidimensionales en Estadística Descriptiva*)

$X \backslash Y$	0	1	2
0	0	0	$1/8$
1	0	$2/8$	$1/8$
2	$1/8$	$2/8$	0
3	$1/8$	0	0

Como puede comprobarse $\sum_{i=0}^{3}\sum_{j=0}^{2} p_{ij} = 1$. ◄

Función de distribución bidimensional. Función de densidad bidimensional.

Dada una variable aleatoria bidimensional (X,Y), se define la **función de distribución** (conjunta) de (X,Y) como:

$$F(x,y) = \mathrm{P}[X \le x, Y \le y] = \mathrm{P}\left[\omega \in \Omega / X(\omega) \le x, Y(\omega) \le y\right] \qquad \forall (x,y) \in \mathrm{R}^2$$

Si las variables aleatorias son discretas la función de distribución es escalonada (*discontinua*). Si las variables aleatorias son continuas la siguiente derivada de la función de distribución conjunta es la **función de densidad** (conjunta) de (X,Y):

$$\frac{\partial^2 F(x,y)}{\partial x\, \partial y} = f(x,y)$$

Toda función de densidad bidimensional cumple las siguientes propiedades que la caracterizan:

- $f(x,y) \ge 0 \qquad \forall (x,y) \in \mathrm{R}^2$
- $\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)\, dx\, dy = 1$

La probabilidad de que la variable aleatoria bidimensional (X,Y) tome valores en un determinado conjunto es igual a la integral definida de la función de densidad bidimensional sobre dicho conjunto. En particular:

$$F(x,y) = \mathrm{P}[X \le x, Y \le y] = \int_{-\infty}^{x}\int_{-\infty}^{y} f(u,v)\, du\, dv$$

►***EJEMPLO 6.9***

Sea la variable aleatoria bidimensional de tipo continuo con función de distribución conjunta

$$F(x,y)=\begin{cases}\dfrac{3xy^2-x^3y^2}{2} & ,0\le x\le 1 \quad 0\le y\le 1\\ 0 & , \quad en\ otro\ caso\end{cases}$$

Calcule la función de densidad conjunta y compruebe que cumple las dos propiedades que caracterizan a toda función de densidad.

Solución:

Derivamos primero respecto de *x*

$$\frac{\partial F(x,y)}{\partial x}=\begin{cases}\dfrac{3y^2-3x^2y^2}{2} & ,0\le x\le 1 \quad 0\le y\le 1\\ 0 & , \quad en\ otro\ caso\end{cases}$$

y a continuación derivamos de nuevo respecto de *y*

$$f(x,y)=\frac{\partial^2 F(x,y)}{\partial x\,\partial y}=\begin{cases}\dfrac{6y-6x^2y}{2}=3(1-x^2)y & ,0\le x\le 1 \quad 0\le y\le 1\\ 0 & , \quad en\ otro\ caso\end{cases}$$

- $f(x,y)\ge 0 \qquad \forall(x,y)\in \mathrm{R}^2$

$$\left.\begin{array}{l}0\le x\le 1 \ \Rightarrow\ 0\le\left(1-x^2\right)\\ 0\le y\le 1 \ \Rightarrow\ 0\le y\end{array}\right\} \ \Rightarrow\ 0\le\left(1-x^2\right)y$$

- $\displaystyle\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)\,dx\,dy=1$

$$\int_0^1\int_0^1 3y\left(1-x^2\right)dx\,dy=\int_0^1 3y\left[x-\frac{x^3}{3}\right]_0^1 dy=\int_0^1 3y\left[1-\frac{1}{3}\right]dy=2\int_0^1 y\,dy=2\left[\frac{y^2}{2}\right]_0^1=2\left[\frac{1}{2}\right]=1$$

◄

Distribuciones de probabilidad marginales.

Para variables aleatorias discretas, se define la **distribución de probabilidad marginal** de *X* como

$$p_{i\bullet}=\mathrm{P}\left[X=x_i\right]=\mathrm{P}\left[X=x_i,\quad Y=cualquier\ valor\right]=\sum_{\forall j}p_{ij}$$

Análogamente para la variable aleatoria *Y*

$$p_{\bullet j}=\mathrm{P}\left[Y=y_j\right]=\mathrm{P}\left[X=cualquier\ valor,\quad Y=y_j\right]=\sum_{\forall i}p_{ij}$$

Para variables aleatorias continuas, se define la **función de densidad marginal** de X como

$$f_1(x) = f_X(x) = \int_{-\infty}^{+\infty} f(x,y)\,dy$$

Análogamente para la variable aleatoria Y

$$f_2(y) = f_Y(y) = \int_{-\infty}^{+\infty} f(x,y)\,dx$$

►***EJEMPLO 6.10***

Calcule las distribuciones de probabilidad marginales para las distribuciones de probabilidad bidimensionales de los ejemplos 6.6 y 6.7.

Solución:

Distribución de probabilidad bidimensional del ejemplo 6.6:

Sumando en los márgenes de la tabla de la distribución de probabilidad bidimensional (*de forma análoga a como se hizo en Estadística Descriptiva*) se obtienen la distribución de probabilidad marginal de X ($p_{i\bullet}$) y la distribución de probabilidad marginal de Y ($p_{\bullet j}$).

Como puede comprobarse, la distribución de probabilidad marginal de X coincide con la distribución de probabilidad unidimensional de X=*número de cruces en tres lanzamientos de una moneda* que se obtuvo en el ejemplo 6.1

$X \setminus Y$	0	1	2	$p_{i\bullet}$
0	0	0	$1/8$	$1/8$
1	0	$2/8$	$1/8$	$3/8$
2	$1/8$	$2/8$	0	$3/8$
3	$1/8$	0	0	$1/8$
$p_{\bullet j}$	$2/8$	$4/8$	$2/8$	1

Distribución de probabilidad bidimensional del ejemplo 6.7:

En variables aleatorias continuas la distribución de probabilidad está definida por su función de densidad. En este ejemplo por

$$f(x,y) = \begin{cases} \dfrac{6y - 6x^2 y}{2} = 3(1-x^2)y & , 0 \le x \le 1 \quad 0 \le y \le 1 \\ 0 & , \quad \textit{en otro caso} \end{cases}$$

La distribución de probabilidad marginal de X está dada por su función de densidad marginal

$$f_1(x) = \int_{-\infty}^{+\infty} f(x,y)\,dy = \int_0^1 3\left(1-x^2\right) y\,dy = 3\left(1-x^2\right)\left[\frac{y^2}{2}\right]_0^1 = \frac{3}{2}\left(1-x^2\right) \qquad si \quad 0 \le x \le 1$$

$$f_1(x) = 0 \qquad en \quad otro \quad caso$$

Análogamente para la variable aleatoria Y

$$f_2(y) = \int_{-\infty}^{+\infty} f(x,y)\,dx = \int_0^1 3\left(1-x^2\right) y\,dx = 3y\left[x - \frac{x^3}{3}\right]_0^1 = 3y\left[1-\frac{1}{3}\right] = 2y \qquad si \quad 0 \le y \le 1$$

$$f_2(y) = 0 \qquad en \quad otro \quad caso$$

◄

Independencia de variables aleatorias.

Recordemos que dos sucesos A y B son independientes si

$$\mathrm{P}\left(A \cap B\right) = \mathrm{P}\left(A\right)\mathrm{P}\left(B\right)$$

En el caso discreto, diremos que las variables X e Y son **independientes** si

$$\mathrm{P}\left[X = x_i, Y = y_j\right] = \mathrm{P}\left[X = x_i\right]\mathrm{P}\left[Y = y_j\right] \qquad \forall x_i, y_j$$

o lo que es lo mismo

$$p_{ij} = p_{i\bullet} p_{\bullet j} \qquad \forall i, j$$

(*obsérvese la analogía con la independencia en Estadística Descriptiva donde* $f_{ij} = f_{i\bullet} f_{\bullet j} \quad \forall i, j$)

En el caso continuo, diremos que las variables X e Y son **independientes** si

$$f\left(x,y\right) = f_1\left(x\right) f_2\left(y\right) \qquad \forall x, y$$

►***EJEMPLO 6.11***

Estudiar la independencia de las variables aleatorias de los ejemplos 6.6 y 6.7

Solución:

Distribución de probabilidad bidimensional del ejemplo 6.6:

$X \setminus Y$	0	1	2	$p_{i\bullet}$
0	0	0	$1/8$	$1/8$
1	0	$2/8$	$1/8$	$3/8$
2	$1/8$	$2/8$	0	$3/8$
3	$1/8$	0	0	$1/8$
$p_{\bullet j}$	$2/8$	$4/8$	$2/8$	1

Como puede comprobarse fácilmente, $p_{00} = 0 \neq p_{0\bullet} p_{\bullet 0} = \frac{1}{8}\frac{2}{8} = \frac{1}{32}$, luego no se cumple la condición de independencia, $p_{ij} = p_{i\bullet} p_{\bullet j} \quad \forall i, j$. En ese caso diremos que las variables aleatorias X e Y son **dependientes**.

Distribución de probabilidad bidimensional del ejemplo 6.7:

$$f(x,y) = \begin{cases} \dfrac{6y - 6x^2 y}{2} = 3(1-x^2)y & , 0 \le x \le 1 \quad 0 \le y \le 1 \\ 0 & , \textit{en otro caso} \end{cases}$$

$$f_1(x) = \begin{cases} \dfrac{3}{2}(1-x^2) & , 0 \le x \le 1 \\ 0 & , \textit{en otro caso} \end{cases} \qquad f_2(y) = \begin{cases} 2y & , 0 \le y \le 1 \\ 0 & , \textit{en otro caso} \end{cases}$$

Es inmediato comprobar que $f(x,y) = f_1(x)\, f_2(y) \quad \forall x, y$. Luego X e Y son **independientes.** ◄

Principales momentos en variables aleatorias bidimensionales. Definición y propiedades.

Media (marginal) de *X:*

Variables discretas: $E[X] = \sum_{i=1}^{\infty} x_i p_{i\bullet}$

Variables continuas: $E[X] = \int_{-\infty}^{+\infty} x\, f_1(x)\, dx$

Media (marginal) de *Y:*

Variables discretas: $E[Y]=\sum_{j=1}^{\infty} y_j p_{\bullet j}$

Variables continuas: $E[Y]=\int_{-\infty}^{+\infty} y\, f_2(y)\, dy$

Propiedades:

$E[X+Y]=E[X]+E[Y]$

Esta propiedad es igualmente válida si la suma se refiere a más de dos variables aleatorias.

Varianza (marginal) de *X:*

Variables discretas: $Var[X]=\sigma^2[X]=\sigma_x^2=\sum_{i=1}^{\infty}\left(x_i-E[X]\right)^2 p_{i\bullet}=E\left[X^2\right]-E[X]^2$

Variables continuas: $Var[X]=\sigma^2[X]=\sigma_x^2=\int_{-\infty}^{+\infty}\left(x-E[X]\right)^2 f_1(x)\, dx=E\left[X^2\right]-E[X]^2$

Varianza (marginal) de *Y:*

Variables discretas: $Var[Y]=\sigma^2[Y]=\sigma_y^2=\sum_{j=1}^{\infty}\left(y_j-E[Y]\right)^2 p_{\bullet j}=E\left[Y^2\right]-E[Y]^2$

Variables continuas: $Var[Y]=\sigma^2[Y]=\sigma_y^2=\int_{-\infty}^{+\infty}\left(y-E[Y]\right)^2 f_2(y)\, dy=E\left[Y^2\right]-E[Y]^2$

Propiedades:

$Var[X+Y]=Var[X]+Var[Y]+2Cov[X,Y]$

$Var[X-Y]=Var[X]+Var[Y]-2Cov[X,Y]$

Covarianza (o momento centrado de ordenes 1,1):

Variables discretas:

$$Cov[X,Y]=\sigma[X,Y]=\sigma_{xy}=E\left[\left(X-E[X]\right)\left(Y-E[Y]\right)\right]=\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\left(x_i-E[X]\right)\left(y_j-E[Y]\right)p_{ij}$$

Variables continuas:

$$Cov[X,Y]=\sigma[X,Y]=\sigma_{xy}=E\left[\left(X-E[X]\right)\left(Y-E[Y]\right)\right]=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(x-E[X]\right)\left(y-E[Y]\right) f(x,y)\, dx\, dy$$

Propiedades:

$Cov[X,Y]=E[XY]-E[X]E[Y]$.

Si X e Y son **independientes** $E[XY]=E[X]E[Y]$

Demostración: (*se hace para variables continuas, de forma análoga se haría para variables discretas*)

$$E[XY]=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\ f(x,y)\,dx\,dy=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\ f_1(x)\ f_2(y)\,dx\,dy=\int_{-\infty}^{\infty} y\ f_2(y)\int_{-\infty}^{\infty} x\ f_1(x)\,dx\,dy=$$

$$=\int_{-\infty}^{\infty} y\ f_2(y)\,E[X]dy=E[X]\int_{-\infty}^{\infty} y\ f_2(y)\,dy=E[X]E[Y]$$

Por tanto en caso de **independencia**: $Cov[X,Y]=0$ y $Var[X\pm Y]=Var[X]+Var[Y]$.

Preguntas de teoría del Tema 6

1. Definición de función de distribución de una variable aleatoria unidimensional. Escriba sus propiedades, estableciendo las diferencias entre los casos discreto y continuo.

Tema 7. Algunas distribuciones discretas de probabilidad

7.1. Distribución Uniforme discreta.

Una variable aleatoria discreta X sigue una **distribución Uniforme** en n puntos $x_1, ..., x_n$, si su distribución de probabilidad es:

$$p_k = \mathrm{P}\left[X = x_k\right] = \frac{1}{n} \qquad k = 1, ..., n$$

(Esta es la distribución que se asume en la concepción de la probabilidad clásica o de Laplace)

La variable aleatoria que representa el valor obtenido al lanzar un dado no cargado sigue una distribución Uniforme discreta con probabilidad $\frac{1}{6}$ para cada uno de los seis posibles valores.

La media y varianza para estas variables aleatorias se calculan como sigue:

Media:

$$E\left[X\right] = \sum_{k=1}^{n} x_k p_k = \sum_{k=1}^{n} x_k \frac{1}{n} = \frac{\sum_{k=1}^{n} x_k}{n}$$

Varianza:

$$E\left[X^2\right] = \sum_{k=1}^{n} x_k^2 p_k = \sum_{k=1}^{n} x_k^2 \frac{1}{n} = \frac{\sum_{k=1}^{n} x_k^2}{n} \qquad Var\left[X\right] = \frac{\sum_{k=1}^{n} x_k^2}{n} - \left(\frac{\sum_{k=1}^{n} x_k}{n}\right)^2$$

7.2. Distribución Binomial.

Distribución Binomial de parámetros 1 y *p* (distribución de Bernouilli).

Consideramos un experimento que puede presentar dos resultados: *A=éxito* y $\overline{A} = fracaso$ con probabilidades $\mathrm{P}(A) = p$, $\mathrm{P}(\overline{A}) = q$, donde evidentemente $p + q = 1$.

A dicho experimento le asociamos la variable aleatoria:

X = Número de éxitos en UNA realización del experimento aleatorio,

es decir la variable aleatoria X toma los valores:

Su distribución de probabilidad será:

La anterior distribución de probabilidad se conoce como **distribución Binomial de parámetros 1 y *p***, y esto de denota como:

donde 1 es el número de veces que se observa o realiza el experimento y p es la probabilidad de obtener *éxito.*

PROPIEDADES

1) **Media:**

2) **Varianza:**

Distribución Binomial de parámetros *n* y *p*.

Sea la variable aleatoria X la suma de n variables aleatorias independientes e idénticamente distribuidas (i.i.d.) según una distribución de probabilidad $B(1, p)$, es decir

$$X = X_1 + X_2 + ... + X_n$$

La variable aleatoria X se define como:

$X =$ *Número de éxitos que se obtienen en n realizaciones independientes del experimento aleatorio,*

y puede tomar todos los valores enteros comprendidos entre ____ y _____, ambos inclusive.

A la distribución de esta variable aleatoria se le llama **distribución Binomial de parámetros *n* y *p***, y esto se denota como:

donde n es el número de pruebas idénticas e independientes que se realizan u observan y p la probabilidad de obtener *éxito* en cada prueba.

La distribución de probabilidad asociada a esta variable aleatoria X es:

$$p_x = \mathrm{P}\left[X = x\right] = \binom{n}{x} p^x q^{n-x} \qquad x = 0, \ldots, n$$

donde

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}$$

Se puede comprobar que los valores p_x verifican las dos condiciones que toda distribución de probabilidad ha de cumplir:

- $p_x \geq 0 \quad \forall x$
- $\sum_{x=0}^{n} p_x = 1$

PROPIEDADES

1) **Media:**

2) **Varianza:**

3) **Propiedad de aditividad o reproductividad.**

Sean X e Y dos variables aleatorias independientes con distribuciones de probabilidad $X \sim \mathrm{B}(n, p)$ e $Y \sim \mathrm{B}(m, p)$. Entonces $X + Y \sim \mathrm{B}(n+m, p)$.

►***EJEMPLO 7.1***

En un almacén hay 6000 piezas de las cuales 900 son defectuosas. Se seleccionan aleatoriamente (sin reemplazamiento) 10 piezas. Calcular la probabilidad de que todas estén en buen estado.

►***EJEMPLO 7.2***

Un proceso de fabricación produce un 1% de piezas defectuosas. Las piezas se empaquetan en cajas de 20 unidades. ¿Cuál es la probabilidad de que si se seleccionan 5 cajas al azar, ninguna de ellas tenga piezas defectuosas?

Solución:

X=nº de piezas defectuosas en una caja de 20 unidades $\sim \mathrm{B}(20, \frac{1}{100})$

$$\mathrm{P}[X=0]=\binom{20}{0}0,01^0\times 0,99^{20}=\frac{20!}{0!20!}1\times 0,8179=0,8179$$

Y=nº de cajas sin piezas defectuosas $\sim \mathrm{B}(5,\ 0,8179)$

$$\mathrm{P}[Y=5]=\binom{5}{5}0,8179^5\times 0,1821^0=\frac{5!}{5!0!}0,366\times 1=0,366$$ ◄

7.3. Distribución de Poisson.

Una variable aleatoria X sigue una **distribución de Poisson de parámetro** λ si puede tomar todos los valores enteros no negativos (0, 1, 2, …) con probabilidades:

$$p_x=\mathrm{P}[X=x]=\frac{e^{-\lambda}\lambda^x}{x!} \qquad x=0,1,2,...$$

Se nota abreviadamente como $X \sim \mathrm{P}(\lambda)$.

Una de las aplicaciones comunes de la distribución de Poisson es calcular la probabilidad del número de eventos en un determinado período de tiempo (por ejemplo, el número de automóviles que se presenta a una estación de peaje de una autopista en el intervalo de un minuto).

Media:

$E[X]=\lambda$

Varianza:

$\sigma^2[X]=\lambda$

Propiedad de aditividad o reproductividad.

Sean X e Y dos variables aleatorias independientes con distribuciones de probabilidad $X \sim \mathrm{P}(\lambda_1)$ e $Y \sim \mathrm{P}(\lambda_2)$. Entonces $X+Y \sim \mathrm{P}(\lambda_1+\lambda_2)$.

►***EJEMPLO 7.3***

El número de fallos por hora en un determinado mecanismo sigue una distribución de Poisson de media 1,5.

a) ¿Cuál es la probabilidad de que haya algún fallo durante una hora?

b) ¿Cuál es la probabilidad de que haya algún fallo durante tres horas?

Solución:

a) $X_1 = n^o \ de \ fallos \ durante \ una \ hora \ \square \ \mathrm{P}\left(1,5\right)$

$$\mathrm{P}\left[X_1 \geq 1\right] = 1 - \mathrm{P}\left[X_1 < 1\right] = 1 - \mathrm{P}\left[X_1 = 0\right] = 1 - \frac{e^{-1,5}1,5^0}{0!} = 1 - 0,2231 = 0,7769$$

b) X = número de fallos durante tres horas= $X_1 + X_2 + X_3$

$X_1 = n^o \ de \ fallos \ durante \ la \ primera \ hora \ \square \ \mathrm{P}\left(1,5\right)$

$X_2 = n^o \ de \ fallos \ durante \ la \ segunda \ hora \ \square \ \mathrm{P}\left(1,5\right)$

$X_3 = n^o \ de \ fallos \ durante \ la \ tercera \ hora \ \square \ \mathrm{P}\left(1,5\right)$

$X \ \square \ \mathrm{P}\left(1,5+1,5+1,5\right) = \mathrm{P}\left(4,5\right)$

$$\mathrm{P}\left[X \geq 1\right] = 1 - \mathrm{P}\left[X < 1\right] = 1 - \mathrm{P}\left[X = 0\right] = 1 - \frac{e^{-4,5}4,5^0}{0!} = 1 - 0,0111 = 0,9889$$ ◄

Distribución de Poisson como límite de la distribución Binomial.

Cuando en la distribución $\mathrm{B}(n, p)$ p es muy pequeño y n suficientemente grande, se puede aproximar por la distribución de Poisson $\mathrm{P}(np)$ para el cálculo de las probabilidades asociadas. Esta propiedad sirve para caracterizar a una gran cantidad de fenómenos que siguen una distribución de Poisson, son fenómenos con una pequeña probabilidad de éxito que son observados en un número elevado de ocasiones.

►***EJEMPLO 7.4***

La probabilidad de que aparezca una pieza defectuosa en un proceso es 0,0001. La producción de un año es de 36000 piezas. Calcular la probabilidad de que en la producción anual el número de piezas defectuosas sea por lo menos dos.

Solución:

X = número de piezas defectuosas en la producción anual

$X \ \square \ \mathrm{B}(36000\,,\,0,0001) \rightarrow \mathrm{P}\left(36000 \times 0,0001\right) = \mathrm{P}\left(3,6\right)$

$$\mathrm{P}\left[X \geq 2\right] = 1 - \mathrm{P}\left[X < 2\right] = 1 - \mathrm{P}\left[X = 0\right] - \mathrm{P}\left[X = 1\right] = 1 - \frac{e^{-3,6}3,6^0}{0!} - \frac{e^{-3,6}3,6^1}{1!} = 1 - 0,0273 - 0,0984 = 0,8743$$

Se podría haber resuelto también usando la distribución Binomial:

$$\mathrm{P}\left[X \geq 2\right] = 1 - \mathrm{P}\left[X < 2\right] = 1 - \mathrm{P}\left[X = 0\right] - \mathrm{P}\left[X = 1\right] =$$

$$= 1 - \binom{36000}{0} 0,0001^{0}\ 0,9999^{36000} - \binom{36000}{1} 0,0001\ \ 0,9999^{35999} = 0,874323663$$

Pero es más complicado e incluso a veces imposible realizar los cálculos debido a la magnitud de *n* y *p*. ◄

7.4. Distribución Hipergeométrica.

Se selecciona una muestra (subconjunto) de una población finita con *N* elementos donde los individuos están clasificados en dos categorías (*éxito* y *fracaso*). Se puede hacer básicamente de dos formas:

- *Muestreo con reemplazamiento*: los individuos se extraen, observan y devuelven a la población, de forma que la composición de la población es constante en cada extracción. Para calcular la probabilidad de que cierto número de individuos de la muestra presentan la característica *éxito* se utiliza la distribución *Binomial*.
- *Muestreo sin reemplazamiento*: los individuos extraídos se dejan fuera de la población con lo que la composición de la población va cambiando con cada extracción (es equivalente extraer de uno en uno o en conjunto). Para calcular la probabilidad de que cierto número de individuos de la muestra presentan la característica *éxito* se utiliza la distribución *Hipergeométrica* que vamos a estudiar a continuación.

Sea una población finita con *N* individuos de los cuales *Np* individuos presentan la característica *éxito* y *N(1-p)=Nq* individuos no la presentan. La probabilidad de obtener *éxito* en la primera extracción es $\frac{Np}{N} = p$ y la de no obtenerlo (u obtener *fracaso*) es *q*. En las sucesivas extracciones estas probabilidades de obtener *éxito* o *fracaso* van cambiando al modificarse la composición de la población.

La variable aleatoria *X=número de éxitos en una muestra sin reemplazamiento de tamaño n* sigue una **distribución de probabilidad Hipergeométrica**, abreviadamente $X \sim \mathrm{H}\left(N, n, p\right)$, cuyas probabilidades asociadas son:

$$p_x = \mathrm{P}\left[X = x\right] = \frac{\binom{Np}{x}\binom{Nq}{n-x}}{\binom{N}{n}}$$

La variable aleatoria X puede tomar todos los valores enteros comprendidos entre $max\{0,\ n\text{-}Nq\}$ y $min\{n,\ Np\}$.

Cuando N es infinito (en la práctica un valor muy elevado) las distribuciones *Hipergeométrica* y *Binomial* coinciden.

Media:

$$E[X] = np$$

Varianza:

$$\sigma^2[X] = \frac{N-n}{N-1}npq$$

►***EJEMPLO 7.5***

Una urna contiene 15 bolas blancas y 5 bolas negras. Se extraen 5 bolas, ¿cuál es la probabilidad de que 2 sean negras?

a) Se extraen las bolas una a una, se observa el color y se devuelven a la urna antes de la siguiente extracción.
b) Se extraen las 5 bolas de una vez.

Solución:

X = número de bolas negras entre las 5 extraídas

a) $n = 5 \qquad p = \frac{5}{20} \qquad X \sim \mathrm{B}(5, \frac{5}{20}) = \mathrm{B}(5,\ 0,25)$

$$\mathrm{P}[X=2] = \binom{5}{2} 0,25^2 (1-0,25)^3 = 0,26367$$

b) $N = 20 \qquad n = 5 \qquad p = \frac{5}{20} = 0,25 \qquad X \sim \mathrm{H}(20, 5,\ 0,25)$

$Np = 20 \times 0,25 = 5 \qquad Nq = 20 \times (1-0,25) = 15$

$$\mathrm{P}[X=2] = \frac{\binom{5}{2}\binom{15}{3}}{\binom{20}{5}} = 0,29347$$

◄

►***EJEMPLO 7.6***

Si jugamos una combinación de la lotería primitiva. ¿Cuál es la probabilidad de que acertemos 4 números o más?

Solución:

En el sorteo de la lotería primitiva las extracciones se hacen sin reemplazamiento.

X = número de bolas extraídas iguales a alguno de nuestros 6 números.

$$N = 49 \qquad n = 6 \qquad p = \frac{6}{49} \qquad X \sim \mathrm{H}(49, 6, \frac{6}{49})$$

$$Np = 49\frac{6}{49} = 6 \qquad Nq = N(1-p) = N - Np = 49 - 6 = 43$$

$$\mathrm{P}[X \geq 4] = \mathrm{P}[X = 4] + \mathrm{P}[X = 5] + \mathrm{P}[X = 6] = \frac{\binom{6}{4}\binom{43}{2}}{\binom{49}{6}} + \frac{\binom{6}{5}\binom{43}{1}}{\binom{49}{6}} + \frac{\binom{6}{6}\binom{43}{0}}{\binom{49}{6}} =$$

$$= 0{,}00096862 + 0{,}00001845 + 0{,}000000071 = 0{,}000987141$$

◄

7.5. Distribución Geométrica.

Sea X la variable aleatoria que indica el *número de fracasos* antes de obtener el *primer éxito* en repeticiones idénticas e independientes de un experimento. La variable aleatoria X sigue una **distribución Geométrica de parámetro *p***, abreviadamente $X \sim \mathrm{G}(p)$, donde p es la probabilidad de éxito. X puede tomar todos los valores enteros no negativos (0, 1, 2, …) con probabilidades:

$$p_x = \mathrm{P}[X = x] = (1-p)^x p = q^x p \qquad x = 0, 1, 2, \ldots$$

Si llamamos A al suceso *éxito*:

$$p_x = \mathrm{P}[X = x] = \mathrm{P}\left[\underbrace{\overline{A}\ldots\overline{A}}_{x} A\right] = \underbrace{\mathrm{P}\left[\overline{A}\right]\ldots\mathrm{P}\left[\overline{A}\right]}_{x}\mathrm{P}[A] = \underbrace{q\ldots q}_{x}\, p = q^x p$$

Dado que p y q son positivos lo serán las probabilidades $p_x = \mathrm{P}[X = x]$. Y la suma de todas ellas es 1:

$$\mathrm{P}[X = 0] + \mathrm{P}[X = 1] + \mathrm{P}[X = 2] + \ldots + \mathrm{P}[X = x] + \ldots =$$

$$= p + pq + pq^2 + \ldots + pq^x + \ldots = p(1 + q + q^2 + \ldots) = p\frac{1}{1-q} = p\frac{1}{p} = 1$$

Donde se ha utilizado que la suma de una progresión geométrica de razón $r<1$ es igual a $\frac{a_1}{1-r}$ (a_1=*primer término de la progresión geométrica*).

Media:

$$E[X]=\frac{1-p}{p}=\frac{q}{p}$$

Varianza:

$$\sigma^2[X]=\frac{1-p}{p^2}=\frac{q}{p^2}$$

►***EJEMPLO 7.7***

En una Facultad suelen aprobar el 70% de los alumnos. Hallar la probabilidad de que un alumno apruebe en alguna de las 6 convocatorias de examen que dispone.

Solución:

X=número de suspensos antes de aprobar sigue una distribución Geométrica de parámetro p=0,70.

p=0,70 q=1-0,70=0,30

$$\mathrm{P}[X\le 5]=\mathrm{P}[X=0]+\mathrm{P}[X=1]+\mathrm{P}[X=2]+\mathrm{P}[X=3]+\mathrm{P}[X=4]+\mathrm{P}[X=5]=$$

$$=(0,3^0\times 0,7)+(0,3\times 0,7)+(0,3^2\times 0,7)+(0,3^3\times 0,7)+(0,3^4\times 0,7)+(0,3^5\times 0,7)=$$

$$=0,7+0,21+0,063+0,0189+0,00567+0,001701=0,999271$$

◄

Preguntas de teoría del Tema 7

- *De este tema se puede preguntar cualquier de las distribuciones discretas de probabilidad estudiadas.*