Capítulo VII	PROBABILIDAD

1. Introducción

- Se indicaba en el capítulo anterior que cuando un experimento aleatorio se repite un gran número de veces, los posibles resultados tienden a presentarse un número muy parecido de veces, lo cual indica que la frecuencia de aparición de cada resultado tiende a estabilizarse.
- El concepto o idea que generalmente se tiene del término probabilidad es adquirido de forma intuitiva, siendo suficiente para manejarlo en la vida corriente.

Nos interesa ahora la medida numérica de la posibilidad de que ocurra un suceso ***A*** cuando se realiza el experimento aleatorio. A esta medida la llamaremos **probabilidad del suceso *A*** y la representaremos por ***p(A)***.

La probabilidad es una medida sobre la **escala 0 a 1** de tal forma que:
- Al suceso imposible le corresponde el valor 0
- Al suceso seguro le corresponde el valor 1
- El resto de sucesos tendrán una probabilidad comprendida entre 0 y 1

El concepto de probabilidad no es único, pues se puede considerar desde distintos puntos de vista:
- El punto de vista objetivo
    - Definición clásica o a priori
    - Definición frecuentista o a posteriori
- El punto de vista subjetivo

2. Definición Clásica de la Probabilidad

Sea un experimento aleatorio cuyo correspondiente espacio muestral ***E*** está formado por un número ***n*** finito de posibles resultados distintos y con la misma probabilidad de ocurrir $\{e_1, e_2, \ldots, e_n\}$.
Si n_1 resultados constituyen el subconjunto o suceso A_1, n_2 resultados constituyen el subconjunto o suceso A_2 y, en general, n_k resultados constituyen el subconjunto o suceso A_k de tal forma que:

$$n_1 + n_2 + \ldots + n_k = n$$

Las probabilidades de los sucesos A_1, A_1,..., A_n son:

$$p(A_1) = \frac{n_1}{n} \quad p(A_2) = \frac{n_2}{n} \quad \ldots \quad p(A_k) = \frac{n_k}{n}$$

es decir, que la probabilidad de cualquier suceso ***A*** es igual al cociente entre el número de casos favorables que integran el suceso ***A*** y el número

Regla de Laplace para ***E*** finitos

$$p(A) = \frac{N^o \text{ de casos favorables de } A}{N^o \text{ de casos posibles de } E}$$

de casos posibles del espacio muestral ***E***.

- Para que se pueda aplicar la regla de Laplace es necesario que todos los sucesos elementales sean equiprobables, es decir: $p(e_1) = p(e_2) = \dots = p(e_n)$ y por tanto $p(e_i)=1/n \quad \forall i=1,2,\dots,n$

- Siendo $A=\{e_1, e_2, \dots , e_k\}$ el suceso formado por k sucesos elementales siendo $k \le n$ tendremos: $$p(A) = \sum_{j=1}^{n} p(e_j) = \frac{k}{n} = \frac{N^o\ casos\ favorables}{N^o\ casos\ posibles}$$

La probabilidad verifica las siguientes condiciones:

- La probabilidad de cualquier suceso es siempre un número no negativo entre 0 y 1 $\quad p(A) = \frac{n_i}{n} \quad n_i \le n \quad n_i, n \ge 0$

- La probabilidad del suceso seguro ***E*** vale 1
- La probabilidad del suceso imposible es 0 $\quad p(E) = \frac{n}{n} = 1 \qquad p(\emptyset) = \frac{0}{n} = 0$

- La probabilidad de la unión de varios sucesos incompatibles o excluyentes A_1, A_1,..., A_r es igual a la suma de probabilidades de cada uno de ellos $\quad p(A_1 \cup \dots \cup A_r) = p(A_1) + p(A_2) + \dots + p(A_r)$

Esta definición clásica de probabilidad fue una de las primeras que se dieron (1900) y se atribuye a Laplace; también se conoce con el nombre de **probabilidad a priori** pues, para calcularla, es necesario conocer, antes de realizar el experimento aleatorio, el espacio muestral y el número de resultados o sucesos elementales que entran a formar parte del suceso.

La aplicación de la definición clásica de probabilidad puede presentar dificultades de aplicación cuando el espacio muestral es infinito o cuando los posibles resultados de un experimento no son equiprobables. Ej: En un proceso de fabricación de piezas puede haber algunas defectuosas y si queremos determinar la probabilidad de que una pieza sea defectuosa no podemos utilizar la definición clásica pues necesitaríamos conocer previamente el resultado del proceso de fabricación.
Para resolver estos casos, se hace una extensión de la definición de probabilidad, de manera que se pueda aplicar con menos restricciones, llegando así a la definición frecuentista de probabilidad.

3. Definición Frecuentista de la Probabilidad

La definición frecuentista consiste en definir la probabilidad como el límite cuando n tiende a infinito de la proporción o frecuencia relativa del suceso.

Sea un experimento aleatorio cuyo espacio muestral es ***E***
Sea ***A*** cualquier suceso perteneciente a ***E***
Si repetimos ***n*** veces el experimento en las mismas $\quad \frac{n(A)}{n}$

Condiciones, la frecuencia relativa del suceso ***A*** será:

Cuando el número ***n*** de repeticiones se hace muy grande
la frecuencia relativa converge hacia un valor que llamaremos **probabilidad del suceso *A***.

$$p(A) = \lim_{n \to \infty} \frac{n(A)}{n}$$

Es imposible llegar a este límite, ya que no podemos repetir el experimento un número infinito de veces, pero si podemos repetirlo muchas veces y observar como las frecuencias relativas tienden a estabilizarse.

Esta definición frecuentista de la probabilidad se llama también **probabilidad a posteriori** ya que sólo podemos dar la probabilidad de un suceso después de repetir y observar un gran número de veces el experimento aleatorio correspondiente. Algunos autores las llaman **probabilidades teóricas**.

4. Definición Subjetiva de la Probabilidad

Tanto la definición clásica como la frecuentista se basan en las repeticiones del experimento aleatorio; pero existen muchos experimentos que no se pueden repetir bajo las mismas condiciones y por tanto no puede aplicarse la interpretación objetiva de la probabilidad.
En esos casos es necesario acudir a un punto de vista alternativo, que no dependa de las repeticiones, sino que considere la probabilidad como un concepto **subjetivo** que exprese el grado de creencia o confianza individual sobre la posibilidad de que el suceso ocurra.
Se trata por tanto de un juicio personal o individual y es posible por tanto que, diferentes observadores tengan distintos grados de creencia sobre los posibles resultados, igualmente válidos.

5. Definición Axiomática de la Probabilidad

La definición axiomática de la probabilidad es quizás la más simple de todas las definiciones y la menos controvertida ya que está basada en un conjunto de axiomas que establecen los requisitos mínimos para dar una definición de probabilidad.

La ventaja de esta definición es que permite un desarrollo riguroso y matemático de la probabilidad. Fue introducida por A. N. Kolmogorov y aceptada por estadísticos y matemáticos en general.

Definición

Dado el espacio muestral ***E*** y la **α-Algebra *A=P(E)*** diremos que una función ***p: A*** $\rightarrow$ *[**0,1**]* es una probabilidad si satisface los siguientes **axiomas de Kolmogorov**:

- ***p(A)*** $\geq$ ***0*** para cualquier suceso ***A*** $\in$ ***A=P(A)***
- ***p(E) = 1***
- Dada una sucesión numerable de sucesos incompatibles $A_1, A_2, \ldots \in A$, se verifica que

$$p(A_1 \cup A_2 \cup \ldots) = p\left(\bigcup_{i=1}^{\infty} A_i\right) = p(A_1) + p(A_2) + \ldots$$

A la función ***p: A** $\rightarrow$ [**0,1**]*
A** $\rightarrow$ **p = p(A) se denomina **probabilidad del suceso *A***.

La terna (***E***, ***A***, ***p***) formada por el espacio muestral ***E***, la α**-Algebra *A=P(E)*** y la probabilidad ***p*** se denomina **espacio probabilístico**.

---

6. Teoremas Elementales o Consecuencias de los Axiomas

Los siguientes resultados se deducen directamente de los axiomas de probabilidad.

Teorema I

La probabilidad del suceso imposible es nula $p(\emptyset) = 0$

- Si para cualquier suceso ***A*** resulta que ***p(A)=0*** diremos que ***A*** es el suceso **nulo**, pero esto no implica que $A = \emptyset$
- Si para cualquier suceso ***A*** resulta que ***p(A)=1*** diremos que ***A*** es el suceso **casi seguro**, pero esto no implica que $A = E$

Teorema II

Para cualquier suceso ***A*** $\in$ ***A=P(A)*** se verifica que:
La probabilidad de su suceso complementario es $p(\overline{A}) = 1 - p(A)$

Teorema III

La probabilidad ***P*** es monótona no decreciente, es decir:
$\forall$***A,B*** $\in$ ***A=P(A)*** con ***A*** $\subset$ ***B*** $\Rightarrow$ ***p(A)*** $\leq$ ***p(B)*** y además ***p(B - A) = p(B) - p(A)***

Teorema IV

Para cualquier suceso ***A*** $\in$ ***A=P(A)*** se verifica que: ***p(A)*** $\leq$ ***1***

Teorema V

Para dos sucesos cualesquiera ***A,B*** $\in$ ***A=P(A)*** se verifica que: ***p(A*** $\cup$ ***B) = p(A) + p(B) - p(A*** $\cap$ ***B)***
Esta propiedad es generalizable a ***n*** sucesos:

$$p\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{n} p(A_i) - \sum_{i<j}^{n} p(A_i \cap A_j) + \sum_{i<j<k}^{n} p(A_i \cap A_j \cap A_k) + \ldots + (-1)^{n+1} p\left(\bigcap_{i=1}^{n} A_i\right)$$

Teorema VI

Para dos sucesos cualesquiera ***A,B*** $\in$ ***A=P(A)*** se verifica que: ***p(A***$\cup$***B)*** $\leq$ ***p(A) + p(B)***
Esta propiedad es generalizable a ***n*** sucesos:

$$p\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{n} p(A_i)$$

Teorema VII

Dada una sucesión creciente de sucesos A_1, A_2, ... , A_n (abreviadamente representado por { $A_n\uparrow$})
se verifica que:

$$\lim_{n\to\infty} p(A_n) = p(\lim_{n\to\infty} A_n) = p\left(\bigcup_{n=1}^{\infty} A_n\right)$$

Teorema VIII

Dada una sucesión decreciente de sucesos A_1, A_2, ... , A_n (abreviadamente representado por { $A_n\downarrow$})
se verifica que:

$$\lim_{n\to\infty} p(A_n) = p(\lim_{n\to\infty} A_n) = p\left(\bigcap_{n=1}^{\infty} A_n\right)$$

7. Probabilidad Condicionada

Hasta ahora hemos introducido el concepto de probabilidad considerando que la única información sobre el experimento era el espacio muestral. Sin embargo hay situaciones en las que se incorpora información suplementaria respecto de un suceso relacionado con el experimento aleatorio, cambiando su probabilidad de ocurrencia.

El hecho de introducir más información, como puede ser la ocurrencia de otro suceso, conduce a que determinados sucesos no pueden haber ocurrido, variando el espacio de resultados y cambiando sus probabilidades.

Definición

Dado un espacio probabilístico (***E***, ***A***, ***p***) asociado a un experimento aleatorio.
Sea ***A*** un suceso tal que ***A*** $\in$ ***A=p(A)*** y ***p(A)***$\geq$***0***
Sea ***B*** un suceso tal que ***B*** $\in$ ***A=p(A)***

Se define la **probabilidad condicionada de *B* dado *A*** o **probabilidad de *B* condicionada a *A*** como:

$$p(B / A) = \frac{p(A \cap B)}{p(A)}$$

La probabilidad condicionada cumple los **tres axiomas de Kolmogorov**:

- $\forall B \in A = P(E) \quad p(B / A) = \frac{p(A \cap B)}{p(A)} \geq 0$
- $p(E / A) = \frac{p(A \cap E)}{p(A)} = 1$

- Sea { A_i } una sucesión de sucesos disjuntos dos a dos, entonces: $p\left(\bigcup_{i=1}^{\infty} A_i / A\right) = \sum_{i=1}^{\infty} p(A_i / A)$

Regla de Multiplicación de Probabilidades o Probabilidad Compuesta

Partiendo de la definición de la probabilidad condicionada ***p(B/A)*** podemos escribir: $p(A \cap B) = p(A)\ p(B / A)$

8. Teorema de la Probabilidad Compuesta o Producto

Sean ***n*** sucesos $A_1, A_2, \ldots, A_n \in \mathcal{A} = \mathcal{P}(A)$ y tales que $p\left(\bigcap_{i=1}^{n-1} A_i\right) > 0$. Se verifica que:

$$p(A_1 \cap A_2 \cap \ldots \cap A_n) = p(A_1)\ p(A_2 / A_1)\ p(A_3 / A_1 \cap A_2) \ldots p(A_n / A_1 \cap \ldots \cap A_{n-1})$$

9. Teorema de la Probabilidad Total

Sean ***n*** sucesos disjuntos $A_1, A_2, \ldots, A_n \in \mathcal{A} = \mathcal{P}(A)$ tales que $p(A_i) > 0 \quad i = 1, 2, \ldots, n$ y tales que forman un sistema completo de sucesos. Para cualquier suceso $B \in \mathcal{A} = \mathcal{P}(A)$ cuyas probabilidades condicionadas son conocidas $p(B/A_i)$, se verifica que:

$$p(B) = \sum_{i=1}^{n} p(A_i)\ p(B / A_i)$$

10. Teorema de Bayes

Sean ***n*** sucesos disjuntos $A_1, A_2, \ldots, A_n \in \mathcal{A} = \mathcal{P}(A)$ tales que $p(A_i) > 0 \quad i = 1, 2, \ldots, n$ y tales que forman un sistema completo de sucesos. Para cualquier suceso $B \in \mathcal{A} = \mathcal{P}(A)$ se verifica que:

$$p(A_i / B) = \frac{p(A_i)\ p(B / A_i)}{\sum_{i=1}^{n} p(A_i)\ p(B / A_i)}$$

y aplicando el teorema de la probabilidad total:

$$p(A_i / B) = \frac{p(A_i)\ p(B / A_i)}{p(B)}$$

Sistema completo de sucesos $A_1, A_2, \ldots, A_n$	Se denominan **hipótesis**
Las probabilidades $p(A_i) > 0 \quad i = 1, 2, \ldots, n$	Se denominan **probabilidades a priori** ya que son las que se asignan inicialmente al los sucesos A_i
Las probabilidades $p(B/A_i) > 0 \quad i = 1, 2, \ldots, n$	Se denominan **verosimilitudes** del suceso ***B*** admitiendo la hipótesis A_i

Las verosimilitudes $p(B/A_i)$ nos permiten modificar nuestro grado de creencia original $p(A_i)$ obteniendo la probabilidad a posteriori $p(A_i / B)$.

El teorema de Bayes, además de ser una aplicación de las probabilidades condicionadas, es fundamental para el desarrollo de la estadística bayesiana, la cual utiliza la interpretación subjetiva de la probabilidad.

11. Independencia de Sucesos

Teniendo en cuenta la definición de la probabilidad del suceso ***B*** condicionada a ***A*** se puede decir:

- Cuando ***p(B/A) > P(B)*** entonces el suceso ***A* favorece al *B***
- Cuando ***p(B/A) < P(B)*** entonces el suceso ***A* desfavorece al *B***
- Cuando ***p(B/A) = P(B)*** entonces la ocurrencia de ***A*** no tiene ningún efecto sobre la de ***B***

Diremos que dos sucesos ***A*** y ***B*** **son independientes** si se verifica una cualquiera de las siguientes condiciones equivalentes:

- $p(B / A) = p(B)$ si ***P(A)>0***
- $p(A / B) = p(B)$ si ***P(B)>0***
- $p(A \cap B) = p(A)\ p(B)$

Podemos decir por tanto que si el suceso ***B*** es independiente del suceso ***A***, entonces el suceso ***A*** también es independiente del suceso ***B***, lo que equivale a decir que ambos sucesos son **mutuamente independientes**.

La independencia de sucesos puede extenderse a más de dos sucesos: $p(A \cap B \cap C) = p(A)\ p(B)\ p(C)$
Además, se cumple el siguiente teorema: Si ***A*** y ***B*** son dos sucesos independientes, entonces también lo son los sucesos $\overline{A}$ y B A y $\overline{B}$ $\overline{A}$ y $\overline{B}$