TEORIA DE LA PROBABILIDAD

2.1. Un poco de historia de la teoría de la probabilidad.

Parece evidente que la idea de probabilidad debe ser tan antigua como el hombre. La idea "es muy probable que llueva mañana" la debía pensar y trasmitir el hombre prehistórico.

Pero es recién en 1654 que comienza a desarrollarse el cálculo de probabilidades, cuando Fermat (1601-1665) y Pascal (1623-1662) en 1654, en correspondencia no publicada, comienzan a aplicar métodos matemáticos para resolver problemas de juegos de azar con cartas y dados. Otros nombres destacados en el desarrollo del cálculo de probabilidades son Jakob Bernoulli (1654-1705) ("Ars Conjectandi", publicado póstumo en 1713, que contiene la hoy llamada "ley de los grandes números de Bernoulli") y Abraham de Moivre (1667-1754) ("Doctrina de las Chances, 1718). En el siglo siguiente se destaca Laplace (1749-1827) y su obra "Teoría analítica de la probabilidad" (1812).

Después de un lento progreso, se acelera el desarrollo de la teoría de probabilidades a mediados del siglo pasado. Tchebycheff (1821-1894) es el primero de la escuela rusa que contribuyó mucho al desarrollo de la teoría de probabilidades, con matemáticos como Markov (1856-1922) y Kolmogorov (1903-1987)). Problemas de genética que se plantearon a fines del siglo pasado (Galton) y el rápido desarrollo al comienzo de este siglo en Física de las teorías de movimiento browniano y mecánica estadística le dieron a la teoría de probabilidades fuentes de nuevos problemas.

La definición que se usa actualmente de Probabilidad fue dada recién en este siglo, en 1933, por Kolmogorov. Es una definición axiomática, similar a la definición de medida de la teoría de la medida (teoría desarrollada en 1898 por Borel (1871-1956) y que sirve de base a la teoría de integración de Lebesgue (1910)(1875-1941).

Además de sus aplicaciones a la genética, física, tecnología, etc., la teoría de probabilidades sirve de base a la teoría de inferencia estadística, ya que en inferencia estadística se mide la probabilidad de equivocarse al hacer una inferencia inductiva.

2.2. Introducción intuitiva: Experimentos aleatorios. Espacio muestral. Sucesos. Idea intuitiva de la probabilidad como límite de frecuencias relativas.

2.2.1. A qué tipo de problemas se aplica la teoría de la probabilidad?

Ejemplo 1) Un técnico del servicio meteorológico, basándose en los datos sobre las condiciones atmosféricas que recibió hasta las 20 horas de hoy dice "la probablidad de que llueva mañana es 70%".

Ej. 2) Un jugador de ruleta juega a que sale "1a docena", cuál es la probabilidad de que gane?

Ej. 3) En la planta industrial del ejemplo de la clase 1 se va a fabricar un nuevo lote del producto químico, cuál es la probabilidad de que el rendimiento sea ≥ 80%?

Ej. 4) Elijo un hombre adulto al azar en la ciudad de Buenos Aires, cuál es la probabilidad de que mida más de 1,75 metros?

Decimos que en estos cuatro problemas se realiza un experimento aleatorio (aleatorio = al azar).

Qué entendemos por experimento aleatorio? Que tienen en común estos cuatro experimentos?

1) Se conocen todos los resultados posibles del experimento.
2) Si realizo una sola vez el experimento no puedo predecir cuál resultado saldrá.
3) Luego lo veremos

2.2.2. Espacio de resultados o espacio muestral

Llamaremos espacio muestral (S) a un conjunto que contiene a todos los resultados posibles.

En el ejemplo 1) $S=\{L, N\}$,
En el ejemplo 2) $S = \{0,1,2,...,36\}$

En el ejemplo 4) puede tomarse

S= conjunto de los números reales >=0

Observese en este ejemplo que S puede contener elementos que no son resultados del experimento. Lo importante es que contenga a todos los resultados.

2.2.3. Sucesos o eventos:

A veces no interesa exactamente cuál resultado ocurrió, sino si salió o no cierto subconjunto de resultados.
En el ejemplo 2 interesa si salió o no el subconjunto

$$A=\{1,2,...,12\} \subseteq S$$

En la teoría de probabilidad se llama suceso o evento a cualquier subconjunto de S . En la teoría más avanzada los "eventos" a los que se les puede asignar una probabilidad pueden no ser "todos los subconjuntos de S" sino sólo algunos de ellos, pero no nos ocuparemos de esta dificultad.

A los subconjuntos de un solo elemento se los suele llamar "sucesos elementales".

Habíamos dicho que intuitivamente los experimentos aleatorios tenía dos propiedades 1) y 2). Veremos ahora una tercera:

3) Sea A un suceso. Si repito n veces el experimento en las mismas condiciones y cuento

n_A = nro de veces que ocurre un resultado de A

y luego calculo

$$n_A / n = \text{proporción de veces que ocurre A}$$

la experiencia muestra que cuando n es grande esta proporción se parece a un número. A ese número lo llamamos **probabilidad de A = P(A)**

Esta es entonces la idea intuitiva de probabilidad:

$$\boxed{\lim_{n \to \infty} \frac{n_A}{n} = P(A)} \qquad (1)$$

Hubo a principios del siglo XX un intento de usar (1) como definición de probabilidad (von Mises) pero la teoría matemática era muy complicada. Es más simple desarrollar la teoría a partir de la definición axiomática de Kolmogorov.

2.3. Definición axiomática de Probabilidad. Deducción de otras propiedades de las probabilidades a partir de los axiomas.

La definición axiomática de Probabilidad (Kolmogorov, 1933) es similar a la definición de medida de la teoría de la medida y de la integración de Lebesgue (principio siglo XX).

Definición axiomática de Probabilidad

Sea S un conjunto no vacío al que llamaremos espacio muestral. Una probabilidad es una función que a cada evento (subconjunto de S) le asigna un número real y que satisface las siguientes tres propiedades:

Axioma 1: **$P(A) \geq 0$** para todo evento A

Axioma 2: **$P(S) = 1$**

Axioma 3:
a) (aditividad) Si $A_1, A_2, \ldots, A_n$ son eventos disjuntos (o sea $A_i \cap A_j = 0$, para $i \neq j$) entonces

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

b((σ-aditividad) Si $A_1, A_2, \ldots, A_n, \ldots$ es una sucesión de eventos disjuntos (o sea $A_i \cap A_j = 0$, para $i \neq j$) entonces

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

Comentarios:

- Obviamente estos axiomas fueron pensados teniendo en cuenta la interpretación intuitiva de probabilidad, pero luego nos podemos olvidar de la idea intuitiva y, a partir de estos tres axiomas, se puede construir toda la teoría de la probabilidad.

- P es una función que toma valores reales, cuál es el dominio de esta función?

- Veamos ahora algunas propiedades de la probabilidad (algunas tan intuitivas como los axiomas) que pueden demostrarse facilmente a partir de los tres axiomas.

Algunas propiedades de las probabilidades

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

b) $P(\phi) = 0$

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

c*) Si $A \subseteq B$ entonces $P(B-A) = P(B) - P(A)$

d) Si $A \subseteq B$ entonces $P(A) \leq P(B)$

e) $P(A) <= 1$ para todo suceso A

f) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

g) $P(A \cup B \cup C) = P(A)+P(B)+P(C)-P(A \cap B)-P(A \cap C)- P(B \cap C)+P(A \cap B \cap C)$

Todas las propiedades anteriores (y toda la teoria de probabilidad) puede demostrase partir de los tres axiomas. Demostrar las propiedades a) hasta g)

2.4. Caso particular de espacio muestral es finito o infinito numerable.
Caso particular de espacio muestral finito, con resultados equiprobables.

2.4.1.Caso particular en que el espacio muestral es finito o infinito numerable

En este caso particular basta dar las probabilidades de los sucesos elementales. Con ellas se puede calcular la probabilidad de cualquier subconjunto de S. Esto no es cierto para conjuntos S no numerables.

Denotemos con s_i a los elemtos de S, o sea:

$$S = \{s_1, s_2, ..., s_n, ...\}$$

Supongamos que conocemos las $P(\{s_i\})$ para todo $s_i \in S$ (o sea conocemos la probabilidad de todos los sucesos elementales).

Entonces es fácil demostrar que puede calcularse P(A) para cualquier subconjunto de S. (Demostrarlo).

2.4.2. Caso particular en que el espacio muestral es finito y todos los resultados son igualmente probables.

En este caso particular, es fácil demostrar que:

$$P(A) = \frac{\# A}{\# S} \qquad (2)$$

para todo subconjunto $A \subseteq S$.
Demostrarlo.

El resultado (2) suele enunciarse tambien como

$$P(A) = \frac{\text{Nro. de resultados favorables}}{\text{Nro. de resultados posibles}}$$

lo que muestra que fue usada primero para los juegos de azar, verdad?.

Ejemplos: La expresión (2) puede usarse para calcular las probabilidades en los siguientes ejemplos:

- Si se juega a "1a docena" en la ruleta, cuál es la probabilidad de ganar?
- Un bolillero tiene 3 bolitas blancas y 7 rojas. Si extraigo una bolita al azar, cuál es la probabilidad de que sea blanca?

Puede usarse en los siguientes?

- Tiro una moneda que no sé si es equilibrada o no, cuál es la probabilidad de que salga cara?
- Cuál es la probabilidad de que llueva mañana?
- Cuál es la probabilidad de un hombre adulto elegido al azar mida más de 1,75 metros?

- Cuál es la probabilidad de que el rendimiento del próximo lote (en la planta industrial) sea ≥ 80%?

- Bolillero con 3 blancas y 7 rojas. Extraigo 2 bolitas al azar, cuál es la probabilidad de que ambas sean blancas, si las extracciones se hacen:
  a) las dos bolitas juntas;
  b) 1 a 1 sin reposición; c) 1 a 1 con reposición

2.5. Probabilidad condicional. Independencia. Regla de la multiplicación.

2.5.1. Probabilidad condicional

-Idea intuitiva: es la probabilidad de que salga un suceso, sabiendo que ocurrió otro.

-Ejemplo 1

Supongamos que un pueblo tiene 1000 habitantes de 18 años o más. Supongamos también que se administra una encuesta a todos los habitantes. En la práctia las encuestas no se administran a toda la población (no se hace un censo) sino a una muestra, pero por ahora vamos a pensar que sí.

Consideremos dos de las preguntas de la encuesta: nivel de instrucción y si apoya cierto proyecto del gobierno. Clasificando al nivel de instrucción en tres categorías (alto, medio y bajo) las respuestas obtenidas se muestran en la siguiente tabla:

Nivel de instrucción	**Apoya el proyecto**	**No lo apoya**	**Total**
Alto	80	120	200
Medio	150	350	500
Bajo	30	270	300
Total	260	740	1000

Si se elige un habitante al azar de este pueblo. Cuál es la probabilidad de que:

a) apoye el proyecto?
b) tenga nivel de instrucción “Alto”?
c) apoye el proyecto si se sabe que tiene nivel de instrucción alto?

Observemos que la respuesta al inciso c) puede escribirse como un cociente de dos probabilidades, ya que si llamamos

A = apoya el proyecto
B = tiene nivel de instrucción alto

Prob. de que apoye el proyecto si tiene nivel de inst. alto=

= "prob. condicional de A dado B" = (notación) = P(A|B) =

$$= \frac{80}{200} = \frac{80/1000}{200/1000} = \frac{P(A \cap B)}{P(B)}$$

Entonces es intuitivamente razonable la siguiente definición de probabilidad condicional.

<u>Definición de probabilidad condicional</u>

Sea S un espacio muestral, A y B dos sucesos con P(B) >0.
Se define

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

(P(A|B) se lee probabilidad condicional de A dado B).

- Ejemplo 2:
  Se tira un dado equilibrado.
  Sea A="sale impar" , B=" sal enro. <= 3".

  Como modelo probabilistico para este experimento, consideremos el espacio muestral S={1,2,3,4,5,6} con resultados equiprobables, y los suceso A={1,3,5}, B={1,2,3}

  Cuánto vale:
  i) P(A) ?
  ii) P(A|B) ?

2.5.2. Independencia entre dos sucesos.

Continuemos con el ejemplo 2, en el que se tira un dado equilibrado y se considera A= "sale impar", B="sale <=3"

$P(A) = 3/6 = 1/2$
$P(A|B) = P(A \cap B)/P(B) = (2/6)/(3/6) = 2/3$

Luego $P(A|B) > P(A)$

Sea ahora C="sale ≤ 2" = {1,2}

$P(C) = 2/6 = 1/3$

$P(C/A) = P(A \cap C)/P(A) = (1/6) / (3/6) = 1/3$

En este ejemplo $P(C|A) = P(C)$

Intuitivamente, el hecho de que salga A no altera la probabilidad de que salga C. En este caso se dice que "A y C son independientes".

Dos sucesos A y B se dice que son independientes, cuando

$$P(A|B) = P(A), \tag{3}$$

Si $P(A|B) = P(A)$ entonces $P(A \cap B)/P(B) = P(A)$, luego

$$P(A \cap B) = P(A)\ P(B) \tag{4}$$

Esta última expresión la usaremos como definición.

<u>Definición de independencia entre dos sucesos</u>

Sea S un espacio muestral, A y B dos sucesos. Se dice que A y B son independientes cuando

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

Por qué hemos usado (4) en lugar de (3) (que es más intuitivo) como definición de independencia? Qué ventaja tiene?

Proposición:

a) Sea $P(B)>0$ entonces

$$A \text{ y } B \text{ son independientes} \Leftrightarrow P(A|B) = P(A)$$

b) Sea $P(A)>0$ entonces

$$A \text{ y } B \text{ son independientes} \Leftrightarrow P(B|A) = P(B)$$

(Demostrar).

Ejemplos:

1) Es independiente apoyar el proyecto del gobierno y tener nivel de instrucción alto?
2) Es independiente "estar de acuerdo con cierta medida de gobierno" y "haber votado al actual presidente en las últimas elecciones"?
3) Es independiente "haber sacado una nota ≥ 7 en Análisis" y "sacar una nota ≥ 7 en Estadistica"?
4) Dar algún ejemplo de dos sucesos que (al menos intuitivamente) sean independientes.

Comentario: Si dos sucesos A y B son disjuntos y P(A)>0, P(B)>0, entonces A y B no son independientes.
Demostrar.

Ejercicio:

$$A \text{ y } B \text{ indep} \Leftrightarrow A^c \text{ y } B \text{ indep} \Leftrightarrow A^c \text{ y } B^c \text{ indep}$$

Nota: basta demostrar que si A y B son independientes, entonces A y B^c también.
Dar un ejemplo para entender el significado de estas implicaciones.

2.5.3. Regla de la multiplicación.

Recordemos que la definición de probabilidad condicional es:

$$P(A|B) = P(A \cap B) / P(B) \qquad \text{(está definido si } P(B)>0).$$

A partir de esta definición se deduce que (si P(B)>0) es $P(A\cap B) = P(A|B)P(B)$

Análogamente de la definición de $P(B|A)= P(A\cap B) / P(A)$ resulta que

<u>Regla de la multiplicacion:</u>

Si P(A)>0 entonces $\mathbf{P(A\cap B) = P(A)\ P(B|A)}$

En qué casos es útil la regla de la multiplicación?
Es útil cuando es más facil darse cuenta de cuánto vale P(B|A) que calcular $P(A\cap B)$.

Ejemplo: bolillero con 3 bolitas blancas, 7 rojas. Se extraen dos bolitas cuál es la probabilidad de que las dos sean blancas?

a) si las extracciones se realizan sin reposición
b) con reposición.

Generalización de la regla de la multiplicación a n sucesos (n≥2):

$$P(A_1\cap A_2\cap ...\cap A_n) =$$

$$P(A_1)\ P(A_2|A_1)\ P(A_3|A_1\cap A_2)\ \ldots\ P(A_n|A_1\cap A_2\cap ...\cap A_{n-1})$$

Ejemplo: del bolillero del ej. anterior, saco tres bolitas

a) cuál es la probabilidad de que las tres sean blancas?
b) cuál es la probabilidad de que las dos primeras sean blancas y la tercera roja?

(considerar extracciones sin y con reposicion).

2.5.4. Generalización de la definicion de independencia para mas de dos sucesos.

Generalización de la definición de independencia

Sean $A_1, A_2, ..., A_n$ sucesos. Diremos que son independientes si para todo subconjunto de índices $\{i_1, i_2, …, i_k\} \subseteq \{1, 2, …, n\}$ se cumple

$$P(A_{i1} \cap A_{i2} \cap ... \cap A_{ik}) = P(A_{i1})\ P(A_{i2})\ ...\ P(A_{ik})$$

Para el caso particulas de tres sucesos la definicion anterior dice que A_1, A_2 y A_3 son independientes si se cumple

$$P(A_1 \cap A_2) = P(A_1)P(A_2)$$
$$P(A_1 \cap A_3) = P(A_1)P(A_3)$$
$$P(A_2 \cap A_3) = P(A_2)P(A_3)$$
$$P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2)P(A_3)$$

Ejemplo: el ejemplo de las extraccion de tres bolitas de un bolillero sirve para dar ejemplos de 3 sucesos independientes?