Apuntes de Probabilidad y Estadística para Ingeniería y Administración

Ignacio Vélez Pareja
Decano
Facultad de Ingeniería Industrial
Politécnico Grancolombiano
Bogotá, Colombia
Octubre, 2002

Conceptos básicos de probabilidad

"La necesidad de jugar es tan apremiante y su práctica tan placentera, que supongo que debe ser pecado".
Heywood Broun (Citado por Thomas y Ronald Wonnacott)
"Debemos creer en la suerte. Porque, ¿de qué otra manera se explica el éxito de las personas que no nos gustan?"
Jean Cocteau (Citado por Thomas y Ronald Wonnacott)

Abstract

This is a course material for an introductory course in Probability and Statistics for Engineering and Management. It is part of some course notes for my courses in Spanish on that subject. The draft of the book is Apuntes de Probabilidad y Estadística para Ingeniería y Administración (Notes for Probability and Statistic for Engineering and Management) and this part is Conceptos básicos de probabilidad (Basic Concepts in Probability). When we make decisions and we are confronted with future and known events the only reason a decision maker makes a mistake is when she has made mistake in her analysis. This is known as decision making under certainty. Unfortunately this is not the most common situation.

Reality is unpredictable and we make decisions under risk and uncertainty and the decision maker has to assume risks. For these typical and common situations we need to study the probability associated to each event and we need to know how to measure and apply it in the decision process.

In this chapter we study the basic concepts of events and probability, including the idea of inductive and deductive processes. We study the notion of statistical independence and Bayes' Theorem. All these methods are studied using intensively the spreadsheet.

Resumen

Cuando se toman decisiones sobre resultados futuros que se conocen, la única razón para que se cometa un error es que exista un error en el análisis por parte del decisor. Esta situación se conoce como certidumbre completa.

Pero la realidad casi nunca es totalmente predecible. Por lo tanto, aunque el decisor haya hecho el análisis correcto, siempre hay factores que no puede controlar y que influyen para que los resultados sean imprevistos. Cuando prevalecen estas condiciones se dice que se trabaja "bajo incertidumbre" y, por lo tanto, el decisor se ve obligado a asumir riesgos. Por ejemplo, que los resultados de sus decisiones no sean favorables. Una forma de hacerlo es medir el riesgo asociado a cada predicción; riesgo que significa qué tantas posibilidades hay de que la decisión adoptada sea errónea. Con esa información el decisor tomará la "mejor" determinación y sólo queda esperar para saber si el resultado es o no favorable. Por más cuidado que se tenga en el análisis, siempre existirá la posibilidad de que el resultado sea desfavorable.

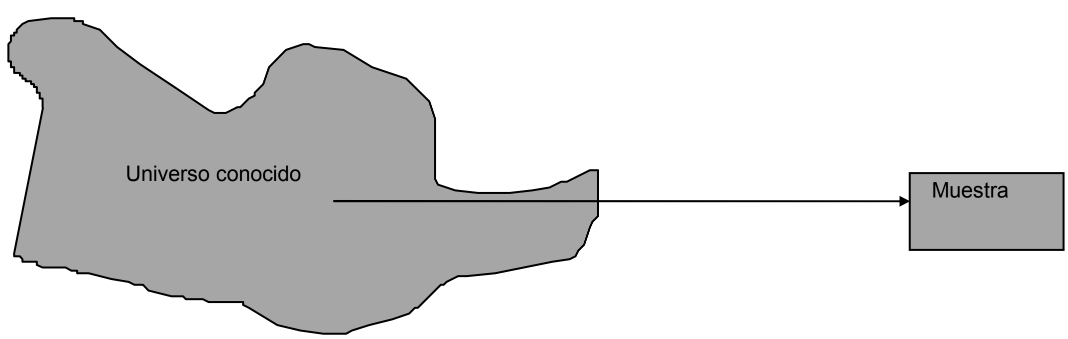
Hay que aceptar que este tipo de análisis es anticipado —*"a priori"*— y que se sabe que el resultado será único. Sin embargo, al hacer el análisis *"a priori"*, hay que identificar el mayor número de resultados *posibles* y medir, para cada resultado, la probabilidad de que ocurra. Cuando se trabaja con decisiones bajo riesgo es necesario entonces introducir el concepto de probabilidad. Esta idea se utiliza en forma intuitiva y en el léxico corriente. Así, por ejemplo, se habla de la probabilidad de que llueva o de que un candidato gane cierta elección. Estas son probabilidades subjetivas que, ante escasez de información, son válidas.

Deducción e inducción

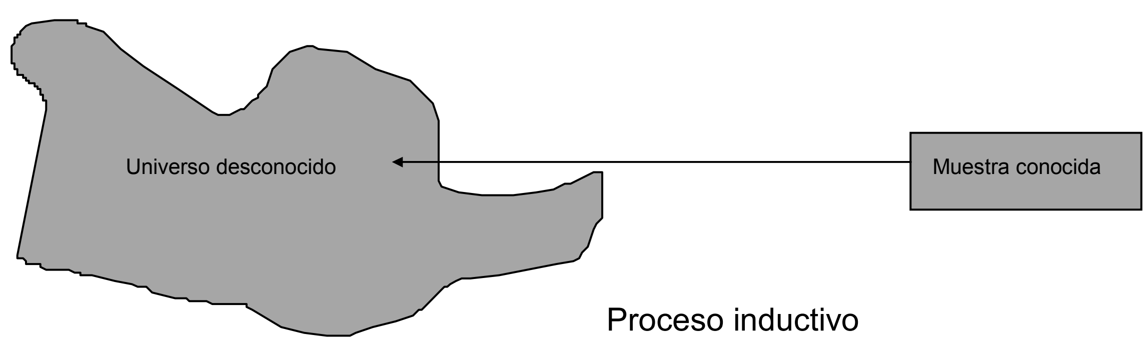
Al abordar el problema de la probabilidad y hacer análisis de tipo probabilístico, conviene distinguir entre un proceso deductivo y uno inductivo. Al estudiar la Teoría de la probabilidad y los rudimentos de la estadística, se está utilizando un proceso deductivo, esto es, que se parte de lo general para decir algo de lo particular. Por ejemplo, se conoce un universo de elementos y se desea saber cuál es el comportamiento de un grupo reducido de observaciones tomadas de ese universo.

En los estudios más avanzados de la Estadística se utiliza un método inductivo, esto es, que a partir de la información obtenida de unos pocos elementos de un universo —una muestra— se trata de encontrar las características de todo el universo. Una manera de recordar estas ideas es tener presente que la partícula *de* en *de*ducción indica hacia afuera y la partícula *in* que se encuentra en *in*ducción indica hacia adentro. Esto es que la deducción va del universo conocido (general) a la muestra desconocida (particular) —hacia afuera del universo— y la inducción va de la muestra conocida (particular) al universo desconocido (general) —hacia adentro del universo—.

Gráficamente:



Proceso deductivo



Proceso inductivo

Cuando se utiliza la inferencia estadística se está trabajando el concepto de inducción, o sea que de unas pocas observaciones se obtienen conclusiones sobre la totalidad del universo. El mejor ejemplo de este trabajo es el que hacen las firmas que realizan encuestas de opinión y las que miden el "rating" o la sintonía de los programas de televisión: A partir de 500 o 1.000 observaciones, hacen afirmaciones sobre el comportamiento de la totalidad de la población.

Probabilidad

Supóngase un experimento cualquiera, por ejemplo, el número dos en el lanzamiento de un dado. El conjunto de todos los resultados posibles se llama universo o espacio de la muestra, en este caso, los números de 1 a 6 en el lanzamiento del dado en cuestión.

Usualmente se utiliza el concepto de frecuencia para ilustrar el concepto de probabilidad. Supóngase que se estudian n resultados de un experimento, de los cuales m se consideran ocurrencias exitosas de un resultado deseado, E y $P(E)$ denota la probabilidad de ocurrencia de dicho resultado; la relación entre el número de resultados exitosos m y el número de resultados posibles n, es una medida aproximada de la probabilidad de ese resultado, es decir:

$$P(E) = \frac{m}{n} \qquad (1)$$

Esto es rigurosamente cierto cuando n es muy grande. Más formalmente, se deberá escribir así:

$$P(E) = \lim_{n\to\infty} \frac{m}{n} \qquad (2)$$

Donde:

$P(E)$: Probabilidad que el resultado E ocurra.

E: Resultado que interesa analizar.

M: Número de veces que ocurre E.

n: Número de veces que se ejecuta el experimento.

Por ejemplo, si se desea saber cuál es la probabilidad de ocurrencia de que aparezca el número 2 en la cara superior cuando se lanza un dado, se podrían hacer lanzamientos seguidos y anotar cuántas veces aparece cada número, en particular el 2. Si esto se repite varias veces, entonces la relación entre el número de veces que apareció el 2 y el número de lanzamientos será un estimativo de la probabilidad. Esta frecuencia relativa tiende a un número; en el caso de un dado que no esté cargado, esta frecuencia tiende a 1/6.

Una variable aleatoria está definida por una función que asigna un valor de dicha variable aleatoria a cada punto del universo. Por ejemplo, la variable aleatoria puede ser el valor que aparezca en la cara superior del dado, o el cuadrado de este valor, etc. En este ejemplo, $E=2$, m es el número de veces que aparece el número 2 y n es el número de lanzamientos.

Propiedades básicas de la probabilidad

A continuación se presentan algunas propiedades básicas de la probabilidad.

1) La probabilidad de un resultado del universo es una cantidad menor o igual que uno y mayor o igual que cero. Esto se explica porque la probabilidad está definida por la proporción entre un número de casos "exitosos" y el número total de casos. El número de casos "exitosos" es menor que el número total de casos.

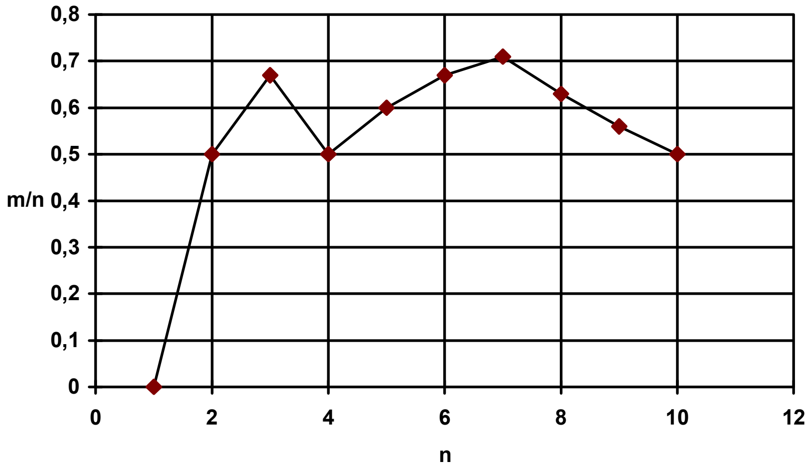
Ejercicio

Lanzar una moneda 50 veces. Construir y completar en la hoja de cálculo la siguiente tabla de ejemplo:

	A	B	C	D
1	Lanzamiento número (*n*)	Cara = 1 Sello = 0	Frecuencia acumulada de caras (*m*)	Frecuencia relativa de caras (*m/n*)
2	1	0	0	0,00
3	2	1	1	0,50
4	3	1	2	0,67
5	4	0	2	0,50
6	5	1	3	0,60
7	6	1	4	0,67
8	7	1	5	0,71
9	8	0	5	0,63
10	9	0	5	0,56
11	10	0	5	0,50

Construir una gráfica de los resultados con *n* en las abcisas y *m/n* en las ordenadas, como se ilustra a continuación.

FRECUENCIA RELATIVA ACUMULADA DEL LANZAMIENTO DE UNA MONEDA



2) La probabilidad de un resultado que no puede ocurrir, o sea que no pertenece al universo, es cero.

3) La probabilidad del universo es uno. Es decir, la probabilidad de que ocurra alguno de los resultados de todo el conjunto posible de ellos es $P(E_1+E_2+...+E_m)$ y es igual a 1, donde ($E_1,E_2,...,E_m$), son todos los resultados posibles, mutuamente excluyentes y exhaustivos del universo.

Se dice que unos resultados son mutuamente excluyentes cuando la ocurrencia de cualquiera de ellos elimina la ocurrencia de cualquier otro.

Todos los resultados posibles m_i suman n, o sea:

$$m_1+m_2+m_3+...m_k = n \tag{3}$$

Si esta ecuación se divide por n, entonces la suma de las frecuencias relativas es igual a 1.

$$m_1/n+m_2/n+m_3/n+...m_k/n = n/n = 1 \tag{4}$$

Así pues en el límite:

$$P(E_1)+P(E_2)+P(E_3)+...P(E_k) = 1 \tag{5}$$

4) Si E y F son resultados mutuamente excluyentes, o sea que sólo uno de ellos puede ocurrir, entonces la probabilidad de que ocurra E o F es $P(E+F) = P(E) + P(F)$. Nuevamente, en el lanzamiento de un dado de seis caras numeradas de 1 a 6, sólo un número aparecerá en la cara superior, por lo tanto, los resultados (E_2) y (E_6), o sea que aparezca 2 en un caso o que aparezca 6 en el otro, son resultados mutuamente excluyentes. La probabilidad de que ocurra E_2 o E_6 es de 1/6+1/6 o sea, 1/3.

5) Si E y F son resultados independientes, esto es, que la ocurrencia de uno no afecta la ocurrencia del otro, la probabilidad de que ocurran simultáneamente $P(EF)$, es $P(E)$ x $P(F)$.

Tomando como ejemplo el dado de seis caras, el hecho que en el primer lanzamiento del dado aparezca un 2, no influye para que en el segundo lanzamiento aparezca cierto número; los lanzamientos son resultados independientes. Entonces, la probabilidad de que en el primer lanzamiento aparezca un 2 y en el segundo aparezca un 6 será 1/6x1/6, o sea, 1/36.

Obsérvese que cuando se trata de resultados mutuamente excluyentes y se desea saber la *probabilidad de que uno de los dos* ocurra, se expresa con frases ligadas por *o*; en el caso de resultados independientes y si se desea calcular la *probabilidad de que ambos* ocurran, las frases se ligan con *y*.

Estas propiedades son formales pero coinciden con las nociones intuitivas de probabilidad.

Eventos y sus probabilidades

En la realidad los hechos no son tan simples como en el ejemplo del dado. Ocurren combinaciones que complican un poco la situación. El cálculo de sus probabilidades es más complejo.

Eventos y sus probabilidades

La realidad es compleja y ocurren combinaciones de resultados; la combinación de varios resultados origina un evento. A través de un ejemplo se ilustrará esta idea.

Ejemplo

Supóngase que se desea analizar los resultados de una inversión $1.000 a tres años. El resultado de cada año es la ocurrencia de un ingreso por valor de $600 o $0. Los resultados posibles son:

$$(NNN) = m_1$$
$$(NNS) = m_2$$
$$(NSN) = m_3$$
$$(NSS) = m_4$$
$$(SNN) = m_5$$
$$(SSN) = m_6$$
$$(SNS) = m_7$$
$$(SSS) = m_8$$

El orden de las letras se refiere año 1, 2 ó 3 y *S* indica si hay ingreso y *N* si no lo hay.

(NSS) significa un flujo de caja como este:

Año	Flujo de caja
0	-1.000
1	0
2	600
3	600

Y así para los demás casos. Las probabilidades de que el resultado sea cero son:

$$P(N)_1 = ,3$$
$$P(N)_2 = ,3$$
$$P(N)_3 = ,3$$

NOTA

Se supone que los eventos son independientes entre sí. Esto significa que el resultado positivo de un año no influye en la probabilidad de que, en los años siguientes, el resultado sea también positivo. Esto en la realidad puede que no ocurra. Sin embargo, para efectos del análisis, se hará caso omiso de esta consideración.

Entonces, las probabilidades asociadas a cada resultado combinado son:

	A	B
1	Evento combinado	Probabilidad total
2	$P(N)_1 =$	30%
3	$P(N)_2 =$	30%
4	$P(N)_3 =$	30%
5	$(NNN) = m_1$	=B1*B2*B3 [0,027]
6	$(NNS) = m_2$	=B1*B2*(1-B3) [0,063]
7	$(NSN) = m_3$	=B1*(1-B2)*B3 [0,063]
8	$(NSS) = m_4$	=B1*(1-B2)*(1-B3) [0,147]
9	$(SNN) = m_5$	=(1-B1)*B2*B3 [0,063]
10	$(SSN) = m_6$	=(1-B1)*(1-B2)*B3 [0,147]
11	$(SNS) = m_7$	=(1-B1)*B2*(1-B3) [0,147]
12	$(SSS) = m_8$	=(1-B1)*(1-B2)*(1-B3) [0,343]

Estos resultados se denominarán puntos. Los eventos serán una combinación cualquiera de puntos. Así se puede pensar en el evento, "por lo menos un año con ingreso", el cual incluiría los puntos $m_2, m_3, m_4, m_5, m_6, m_7$ y m_8, o en el evento "a lo sumo un año con ingreso cero", el cual incluiría los puntos m_4, m_6, m_7 y m_8.

NOTA

Si la probabilidad de que ocurra el ingreso es diferente a 70%, hay que introducir los valores adecuados en los cálculos.

La probabilidad de estos eventos será la suma de la probabilidad de los puntos. En el primer evento, la probabilidad será de:

$0,063+0,063+0,147+0,063+0,147+0,147+0,343 = 0,973$

en el segundo caso de:

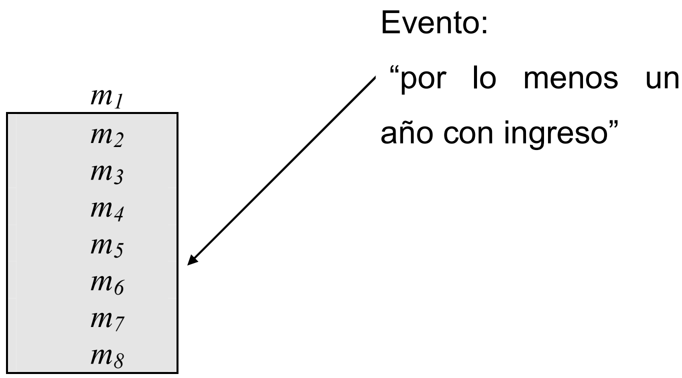
0,147+0,147+0,147+0,343 = 0,784

Diagramas de Venn

Los resultados y sus combinaciones en eventos se pueden visualizar en forma gráfica; estas gráficas se conocen como *diagramas de Venn* y se introducirán mediante la continuación del ejemplo anterior.

Ejemplo

Si se tienen los resultados *m*, entonces los eventos "por lo menos un año con ingreso", o "a lo sumo un año sin ingreso", se pueden ilustrar con una gráfica o *diagrama de Venn*.



Combinación de eventos

Puede ser que se deseen evaluar combinaciones de eventos en términos probabilísticos. En ese caso se debe tener cuidado que no haya "duplicaciones" de puntos en los eventos combinados, puesto que al tener puntos "duplicados" las probabilidades se podrían doblar. Por ejemplo, que ocurrieran los eventos: "menos de dos años con ingreso" o "el primero sin ingreso seguido por dos años con ingresos".

Para facilitar el análisis se puede construir una tabla con algunos eventos posibles, así:

Evento	Descripción	Resultados que incluye	Probabilidad
A	Todos sin ingreso.	m_1	0,027
B	Todos con ingreso.	m_8	0,343
C	Primero sin ingreso, el resto con ingreso.	m_4	0,147
D	Menos de dos años con ingreso.	m_1,m_2,m_3,m_5	0,216
E	Un sólo año con ingreso.	m_2,m_3,m_5	0,189
F	Un sólo año sin ingreso.	m_6,m_7	0,294
G	Por lo menos dos años con ingreso.	m_4,m_6,m_7,m_8	0,784
H	Por lo menos un año con ingreso.	$m_2,m_3,m_4,m_5,m_6,m_7,m_8$	0,973

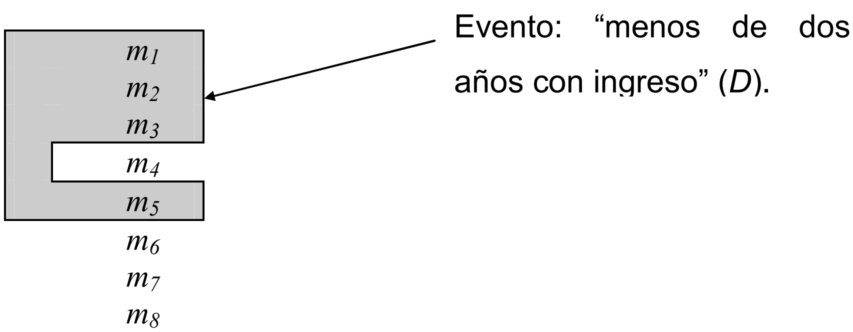
Completar esta tabla con diez eventos adicionales.

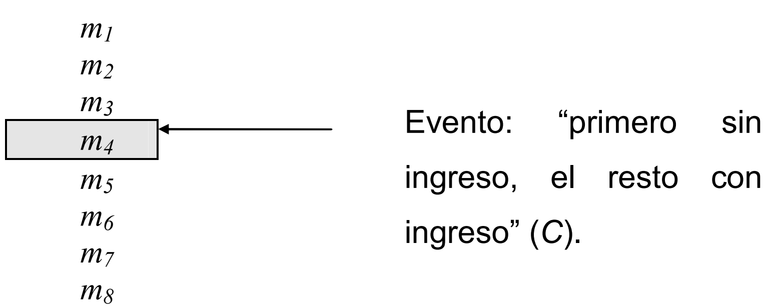
Los eventos mencionados “menos de dos años con ingreso” —evento D— y “el primero sin ingreso el resto con ingreso” —evento C— , contienen los puntos m_1, m_2, m_3, m_5 y m_4 respectivamente. Si se desea encontrar la probabilidad que de ocurra uno de los dos, entonces se dice que se halla la probabilidad de los eventos combinados D o C, en notación matemática, $P(D \cup C)$. Esta combinación de eventos contiene los puntos m_1, m_2, m_3, m_4 y m_5; su probabilidad es 0,363.

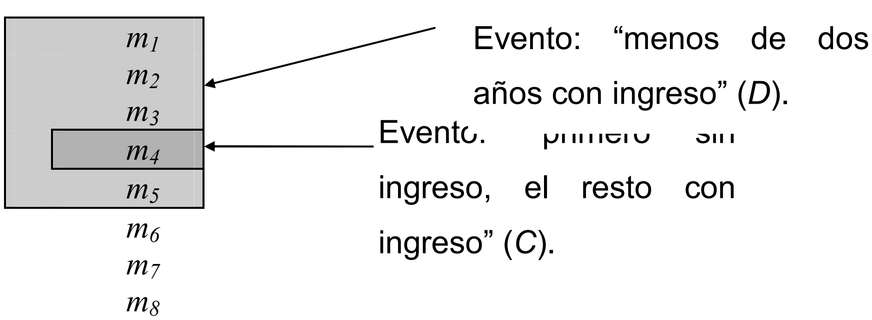
NOTA

La notación $(D \cup C)$ quiere decir que el nuevo evento contiene los puntos que se encuentran en D, en C o en ambos. El nuevo evento se lee D “unión” C.

En diagrama de Venn:







Combinación de eventos C y D en $(D \cup C)$

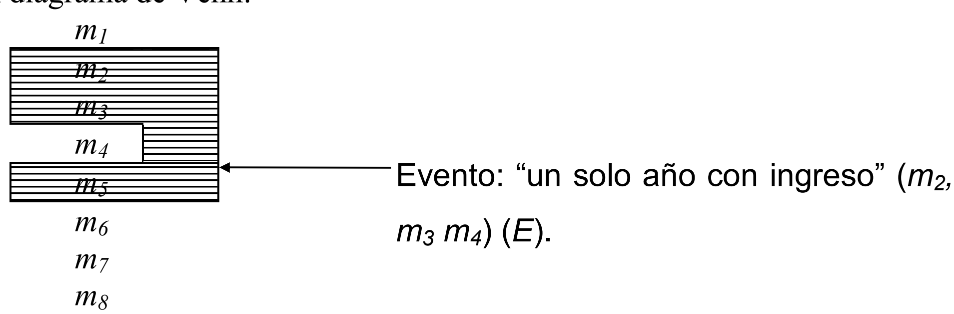
Si se estipulara que deben ocurrir ambos eventos, o sea que ocurriera el evento "menos de dos años con ingreso" —evento D— **y** además ocurriera el evento "el primero sin ingreso seguido por dos años con ingreso" —evento C—, se escribiría como $(D \cap C)$. La notación $(D \cap C)$ indica que el nuevo evento D y C, contiene sólo los elementos comunes a ambos. En este caso se ve claramente que no existen esos elementos comunes y se dice que es un conjunto *vacío*. Esto se puede verificar examinando los resultados m_i y el *diagrama de Venn*.

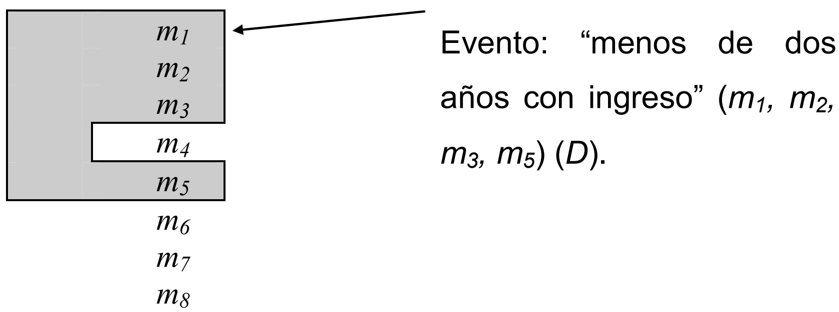
Por no existir ningún resultado posible común, su probabilidad es cero. El nuevo evento se lee D "intersección" C.

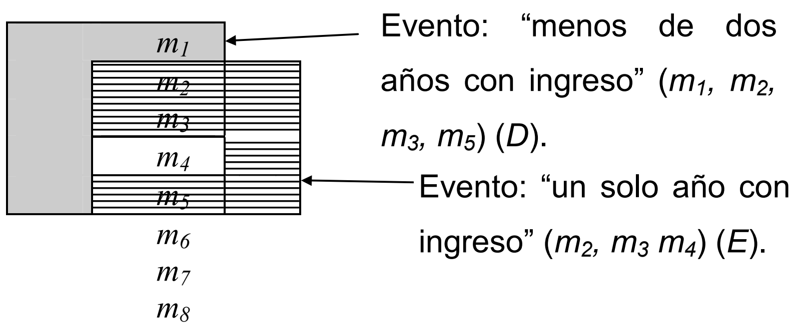
Si se consideraran los eventos D y E, "menos de dos años con ingreso" (m_1, m_2, m_3, m_5) y "un sólo año con ingreso" (m_2, m_3, m_5), se tendría lo siguiente:

$$P(D \cup E) = P(m_1, m_2, m_3, m_5) = 0{,}027+0{,}063+0{,}063+0{,}063 = 0{,}216$$

En diagrama de Venn:

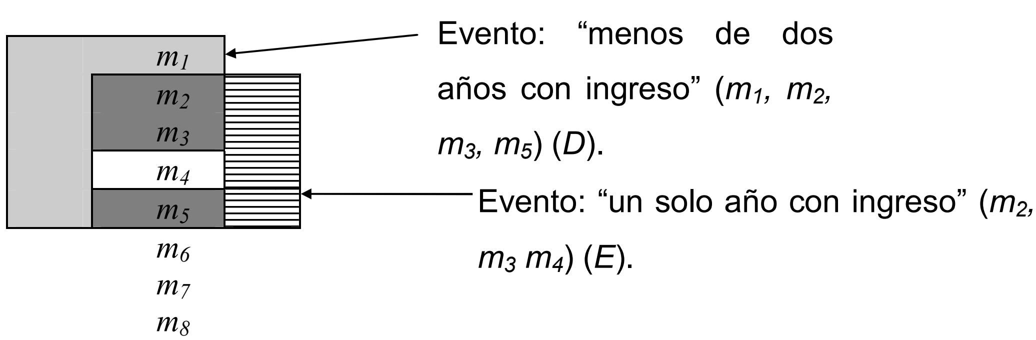






Combinación de eventos D o E en $(D \cup E)$

$P(D \cap E) = P(m_2, m_3, m_5)$ = 0,063+0,063+0,063 = 0,189



Combinación de eventos D y E en $(D \cap E)$ (sombreado fuerte).

Obsérvese que $P(D \cup E) = P(m_1, m_2, m_3, m_5)$ no es igual a 0,500+0,375, o sea a la suma de las probabilidades de los dos eventos, D y E. Si se sumaran las dos probabilidades, se estaría contando dos veces la probabilidad asociada a los resultados comunes a los dos eventos. En forma matemática esta precaución se escribe así:

$$P(D \cup E) = P(D) + P(E) - P(D \cap E) \qquad (6)$$

Lo que en la práctica dice esta ecuación, es que se le elimina una vez la probabilidad asociada a los resultados que están en ambos eventos.

Cuando los eventos -*D* y *E*, por ejemplo- son mutuamente excluyentes, entonces sí se puede calcular la probabilidad de la unión de ellos como la suma:

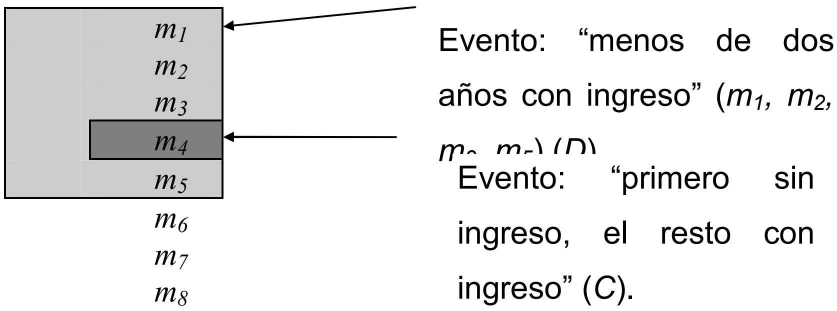
$$P(D \cup E) = P(D) + P(E) \qquad (7)$$

Esto es fácil de deducir, ya que cuando dos eventos son excluyentes, la unión de ellos es cero, porque no existen resultados comunes a ambos.

Particiones y complementos

Cuando un grupo de eventos (colección de un cierto número de puntos o resultados posibles) es mutuamente excluyente, se dice que no existen resultados posibles comunes entre ellos, es decir que ningún resultado o punto pertenece a más de un evento. En el ejemplo de la inversión, se tiene que los eventos *C,* "menos de dos años con ingreso", y *D* "primero sin ingreso y el resto con ingreso", son mutuamente excluyentes; es imposible que esos dos eventos puedan darse al mismo tiempo.

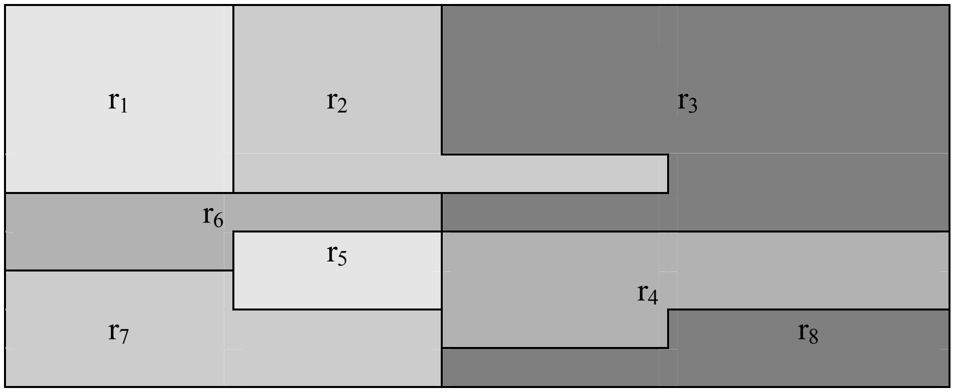
En diagrama de Venn:



Combinación de eventos mutuamente excluyentes, *C* y *D*

A partir de las posibilidades del ejemplo de la inversión a tres años, identificar tres eventos mutuamente excluyentes y construir los *diagramas de Venn* correspondientes, para verificar su resultado.

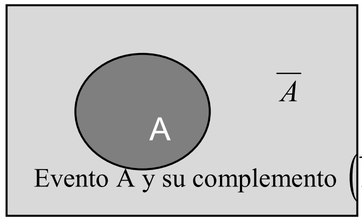
Cuando el conjunto de eventos mutuamente excluyentes es tal que cubre todos los resultados posibles del universo, entonces se les llama particiones. Esto significa que la unión de una colección de eventos mutuamente excluyentes y que conforman una partición es la totalidad del universo.



Partición de un universo

Por otro lado, cuando un evento contiene exactamente los puntos o resultados que no están en otro, y entre los dos se forma una partición, esto es, que entre los dos contienen todos los resultados posibles del universo, se dice que uno es el complemento del otro. Por ejemplo, el evento *A* "todos sin ingreso" y el evento *H* "por lo menos un año con ingreso", son complementarios entre sí. Se dice que el evento *A* es el complemento del evento *H*. La notación es la siguiente: $\bar{A} = H$; lo cual quiere decir que *H* es el complemento de *A*. Es fácil concluir lo siguiente:

$$P(A) + P(\overline{A}) = 1$$



Evento A y su complemento $(\overline{A})$

Probabilidad Condicional y Probabilidad Conjunta

Cuando se conoce que un evento ha ocurrido y se desea saber la probabilidad de que ocurra otro, "dado" que ocurrió el primero, se dice que se calcula la *probabilidad condicional*. O sea que se puede hablar de calcular la probabilidad condicional que ningún año tenga ingreso, dado que se sabe que en el primero no hubo.

Al observar la tabla de los ocho casos posibles, se concluye que al condicionar los eventos al hecho, que el primero no tenga ingreso, en la realidad se "recorta" el universo y por eso las probabilidades cambian. El nuevo universo, cuando se sabe que en el primer año no hubo ingresos queda así:

	A
13	Eventos condicionados a que en el año no hubo ingresos.
14	(*NNN*) = m_1
15	(*NNS*) = *m2*
16	(*NSN*) = *m3*
17	(*NSS*) = *m4*

Si alguien quisiera apostar al hecho de que en ningún año hubiera ingresos, tendría dos estimativos de la probabilidad diferentes: Antes de saber sobre el resultado de alguno de los años y después de tener alguna información sobre el primer año. Una persona razonable, por ejemplo, le asignará probabilidad cero a los resultados m_4, m_5, m_6 y m_7, después de saber que el primer año no tuvo ingresos, ya que esos resultados son imposibles, puesto que en ellos el primer año sí tiene ingresos; así las cosas, los eventos posibles son m_1, m_2, m_3 y m_4 y de estos cuatro eventos, sólo m_1 interesa.

	A	B
5	$(NNN) = m_1$	=B1*B2*B3 [0,027]
6	$(NNS) = m_2$	=B1*B2*(1-B3) [0,063]
7	$(NSN) = m_3$	=B1*(1-B2)*B3 [0,063]
8	$(NSS) = m_4$	=B1*(1-B2)*(1-B3) [0,147]
9	Total	0,30

Como los cuatro resultados tienen una probabilidad total de 0,30 y son los únicos posibles, se concluye que el evento m_1 tiene probabilidad 0,09, o sea 0,027/0,3. Se deduce fácilmente que la probabilidad ha sido "aumentada" y que este aumento debe ser proporcional en cada caso; esto es, que la probabilidad inicial de cada resultado debe ser dividida por la probabilidad total de los eventos que quedan al condicionar los eventos.

	A	B
13	Eventos condicionados a que el primero no tenga ingresos.	Probabilidad condicional que ninguno tenga ingreso, dado que el primero no lo tiene.
14	$(NNN) = m_1$	=B5/B18 [0,09]
15	$(NNS) = m_2$	0,21
16	$(NSN) = m_3$	0,21
17	$(NSS) = m_4$	0,49
18	Probabilidad que el primero no tenga ingresos.	=B5+B6+B7+B8 [0,300]

La notación para la probabilidad condicional de un evento A, dado que ocurrió otro B, se indica así:

$$P(A/B) = \frac{P(A \cap B)}{P(B)} \qquad (8)$$

Aquí se observa lo siguiente: El valor 0,027 de la celda $B5$ es el valor de la probabilidad P("primer año cero"∩ "todos cero") y el denominador es la probabilidad P("primer año cero"), o sea que la probabilidad condicional es:

$$P(\text{"}todoscero\text{"}/\text{"}primerañocero\text{"}) = \frac{P(\text{"}primeraño\ cero\text{"} \cap \text{"}todos\ cero\text{"})}{P(\text{"}primer\ año\ cero\text{"})}$$

$$P(\text{"}todos\ cero\text{"}/\text{"}primer\ año\ cero\text{"}) = \frac{0{,}027}{0{,}3} = 0{,}09$$

La expresión de la probabilidad condicional puede ser escrita de otra forma -en “reversa”- y se conoce como la *regla multiplicativa*, así:

$$P(A \cap B) = P(B) \times P(A / B) \tag{9}$$

Análisis Bayesiano

La fortaleza de la teoría de la probabilidad es la de reforzar el buen juicio o criterio y la experiencia, nunca los reemplaza. De hecho, ningún método para tomar decisiones puede reemplazar ese criterio y experiencia. Los métodos para la toma de decisiones se enriquecen con la experiencia del decisor y lo ayudan a tomar mejores decisiones.

En los estudios de mercado se puede aplicar muy bien la idea del análisis bayesiano. Esto consiste en hacer inferencias sobre unas *causas*, a partir de los *efectos conocidos*; aquí se aplica la idea presentada arriba de la probabilidad condicional. Se calcula la *probabilidad a posteriori*, después de que se ha observado un efecto determinado. En otras palabras, el análisis bayesiano o estadística bayesiana se caracteriza por el hecho de *revisar* o *ajustar* una *probabilidad a priori* acerca de un determinado parámetro a una *probabilidad* ***a*** *posteriori* más confiable, basado en los resultados que ofrece la evidencia de una muestra o estudio adicional.

A través de dos ejemplos se ilustrará la idea del Análisis Bayesiano, para después introducir la presentación de tipo matemático y formal.

Ejemplo

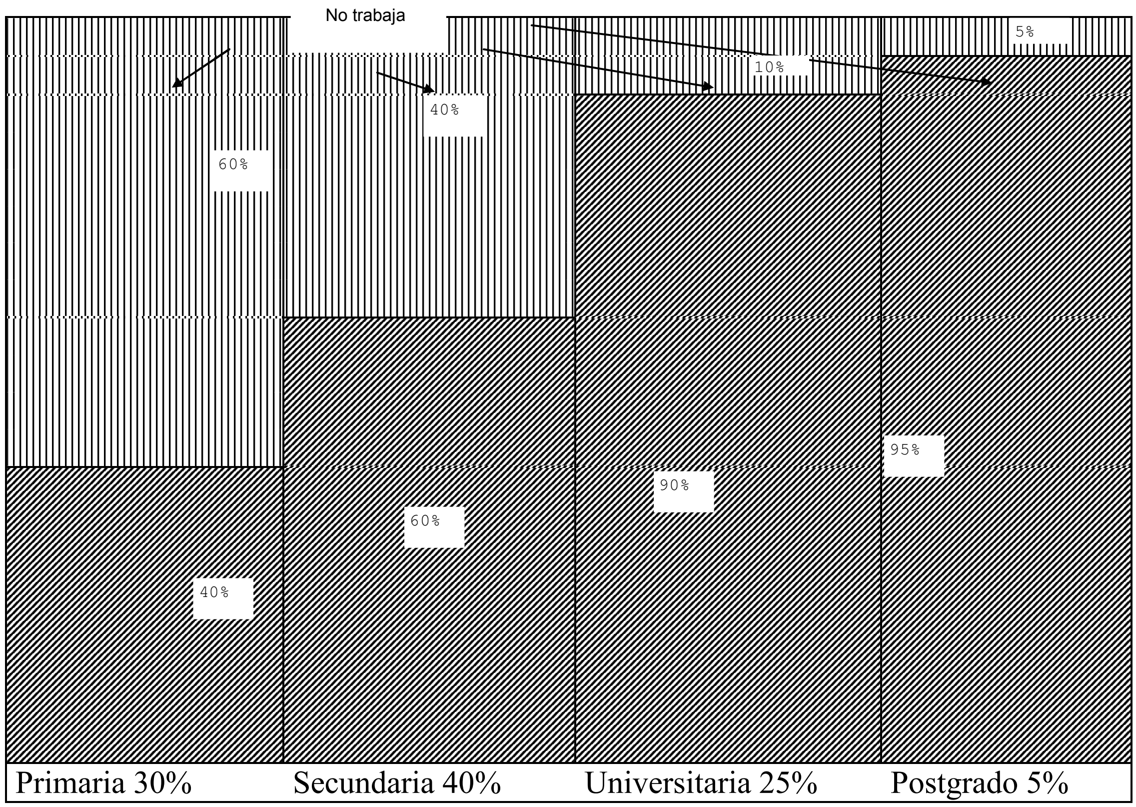
En un estudio de mercado se tienen los siguientes cálculos del mercado potencial para cierto producto:

Mercado objetivo: Amas de casa.

	A	**B**	**C**
1		Educación	Trabaja
2	Primaria	30%	40%
3	Secundaria	40%	60%
4	Universitara	25%	90%
5	Postgrado	5%	95%

Si se hiciera un muestreo de esta población y se escogiera una persona en forma aleatoria y se encuentra que está trabajando, ¿cuál es la probabilidad de que esta persona tuviera título de postgrado?

Una forma intuitiva de analizar esta situación es la de estimar qué porcentajes de esas amas de casa trabajan (Efecto (E)) en cada categoría de educación (Causa (C)). En forma gráfica se tiene:



Se podría esperar que de acuerdo con cada nivel de educación y desempleo, hubiera entonces el siguiente porcentaje total de gente trabajando:

	A	B	C	D
1		Educación	Trabaja	% Total trabajando
2	Primaria.	30%	40%	=B2*C2 [12%]
3	Secundaria.	40%	60%	=B3*C3 [24%]
4	Universitaria.	25%	90%	=B4*C4 [22,5%]
5	Postgrado.	5%	95%	=B5*C5 [4,75%]
6	Total.			=D2+D3+D4+D5 [63,25]
7	Probabilidad que entre las que estén trabajando, esa persona tenga título de Postgrado.			=D5/D6 [7,51%]

Ejemplo

Supóngase ahora que se tiene una inversión que tiene un beneficio neto presente (valor presente) de $7 millones si ocurre el evento *A1,* y de -$3 millones si ocurre el evento *A2*. Las probabilidades *a priori* son $P(A1) = 0,3$ y $P(A2) = 0,7$.

Esto significa que el valor esperado de la inversión es:

$$E(A) = \$7x0,3 - \$3x0,7 = 0 \text{ millones}$$

Si no se invierte el valor esperado también sería cero. Si se contrata un estudio adicional sobre este proyecto, los resultados sería que si el resultado del estudio fuera *B1*, entonces esto significaría que habría una ganancia *A1* de $7 millones, y si el resultado del estudio fuera *B2*, significaría que hay una pérdida *A2* de $3 millones. Si llegara a resultar *A1* —$7 millones— el estudio resultaría en *B1* con una probabilidad de 0,8, y si llegara a resultar *A2* - —$3 millones— el estudio resultaría en *B2* con probabilidad 0,6. En resumen se tiene:

Con esta información se puede mejorar el estimativo inicial que se tenía para la probabilidad de *A1* y de *A2.* El análisis sería el siguiente: Si se espera que *A1* ocurra el 30% de los casos y suponiendo que ello ocurriera, sólo el 65% de las veces, el estudio indicará que sí ocurrirá *B1*; así mismo, si se espera que *A2* ocurra el 70% de los casos, el estudio indicará que así ocurrirá -*B2*- el 60% de las veces. En términos de probabilidad condicional:

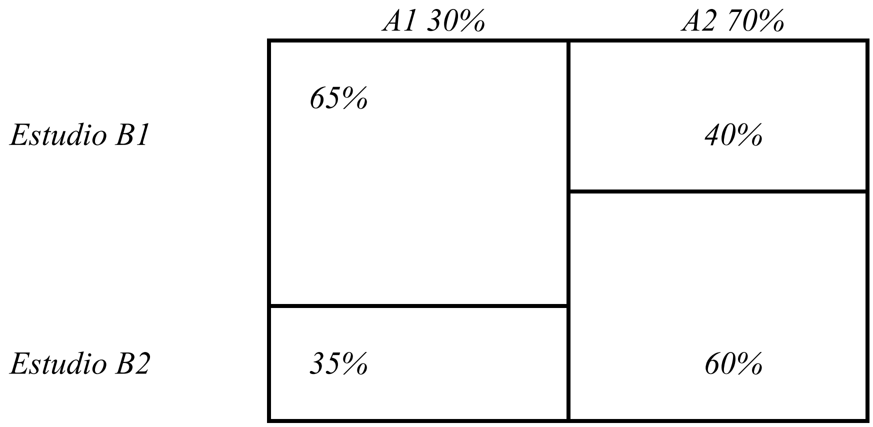
$P(A1) = 0,3$ $P(A2) = 0,7$
$P(B1|A1)= 0,65$
$P(B2|A1) = 1- P(B1|A1) = 1-0,65 = 0,35$
$P(B2|A2) = 0,6$
$P(B1|A2) = 1 - P(B2|A2) = 1 – 0,6 = 0,4$

En forma tabular se tiene:

	A	B	C
1		A1	A2
2	Estudio B1	65%	40%
3	Estudio B2	35%	60%
4	Por resultado, A1 o A2	30%	70%

Obsérvese que los porcentajes suman 100% en sentido vertical, esto es, que los porcentajes asignados a los resultados del estudio son las *probabilidades condicionales* de que sus resultados sean *B1* o *B2*, dado que la inversión producirá *A1* o *A2*. En cambio, los porcentajes asignados por resultado del estudio no son probabilidades condicionales, sino "**incondicionales**", esto es, que no dependen de si el resultado de la inversión es *A1* o *A2*. Si el estudio resulta en *B1,* ¿cuál es la probabilidad de que en la realidad ocurra *A1*?

Gráficamente se tiene:



	A	B	C
1		A1	A2
2	Estudio *B1*	65%	40%
3	Estudio *B2*	35%	60%
4	Por resultado, *A1* o *A2*	30%	70%
5	Proporción de veces en que si *A1* ocurre, el estudio diga que va a ocurrir *A1* (*B1*)	=B2*B4 [19,5%]	
6	Proporción de veces en que si *A2* ocurre, el estudio diga que va a ocurrir *A1* (*B1*)	=C2*C4 [28%]	
7	Proporción de veces en que el estudio acierta en su predicción. (Probabilidad de que el resultado sea *A1* dado que el resultado del estudio es B1)	=B5/(B5+B6) [41,05%]	

Obsérvese que antes de obtener la información adicional, la probabilidad que se le asignó al resultado *A1* era 40%, probabilidad a priori; después de contar con la información adicional que el estudio resultó en *B1*, esa probabilidad subió a 41,05%, probabilidad a posteriori o revisada. Ahora el valor esperado de esta inversión será:

$E(A)$ = \$7x0,4105-\$3x(1-0,4105) = 2,8735-1,7685 = 1,1050

Lo que estos ejemplos ilustran es conocido como el *Teorema de Bayes*.

En el primer ejemplo, se cuenta con la siguiente información:

$P(P)$ = 30%, $P(S)$ = 40%, $P(U)$= 25% y $P(Pos)$ = 5% proporciones (probabilidades) por tipo de educación.

$P(T|P)$= 40%, proporción de amas de casa con primaria que trabaja y $P(NT|P)$=1-40% = 60%, proporción de amas de casa con primaria que no trabaja.

$P(T|S)$= 60%, proporción de amas de casa con secundaria que trabaja y $P(NT|S)$=1-60% = 40%, proporción de amas de casa con secundaria que no trabaja.

$P(T|U)$= 25%, proporción de amas de casa con universitaria que trabaja y $P(NT|U)$=1-25% = 75%, proporción de amas de casa con universitaria que no trabaja.

$P(T|Pos)$= 90%, proporción de amas de casa con postgrado que trabaja y $P(NT|P)$=1-90% = 10%, proporción de amas de casa con postgrado que no trabaja.

La probabilidad de que tenga título de postgrado y de que esté trabajando, en términos de probabilidades condicionales se calcula a continuación.

La probabilidad de que tenga primaria **y** trabaje:

$P(P \cap T) = P(P) \times P(T/P)$=30%x40%= 12%

La probabilidad de que tenga secundaria **y** trabaje:

$P(S \cap T) = P(S) \times P(T/S)$= 40%x60% = 24%

La probabilidad de que tenga universitaria **y** trabaje:

$P(U \cap T) = P(U) \times P(T/U)$ = 25%x90% = 22.5%

La probabilidad de que tenga postgrado **y** trabaje:

$P(Pos \cap T) = P(Pos) \times P(T/Pos)$ = 5%x95% = 4.75%

Si la persona seleccionada trabaja, podrá haber cursado primaria, secundaria, universidad o postgrado y estas posibilidades son excluyentes, por lo tanto, esas probabilidades se pueden sumar y se obtiene la probabilidad, o proporción de personas trabajando, así:

$P(T) = P(P)xP(T/P) + P(S)xP(T/S) + P(U)xP(T/U) + P(Pos) \times P(T/Pos)$

A esta probabilidad se le conoce como *probabilidad marginal*.

Esta es la probabilidad de que esté trabajando. Pero de las que están trabajando sólo interesan las que tienen postgrado y, de acuerdo con la probabilidad condicional, se tiene:

$$P(Pos|T) = \frac{P(Pos \cap T)}{P(T)}$$

$$P(Pos|T) = \frac{P(Pos) \times P(T/Pos)}{P(P)xP(T/P) + P(S)xP(T/S) + P(U)xP(T/U) + P(Pos) \times P(T/Pos)}$$

En el ejemplo, usando directamente el *Teorema de Bayes*:

	A	**B**	**C**	**D**	**F**
8	Evento a analizar.	Probabilidades básicas.		Probabilidades calculadas	
9		Probabilidad del evento.	Probabilidad de que trabaje dado el evento.	Probabilidad conjunta $P(T∩P)$, $P(T∩S)$, $P(T∩U)$ $P(T∩Pos)$	Probabilidad del evento (*P, S, U, Pos*) dado que trabaja
10	Primaria	*=B2* [30%]	*=C2* [40%]	*=B10*C10* [12%]	*=E10/D14* [18,97%]
11	Secundaria	*=B3* [40%]	*=C3* [60%]	*=B11*C11* [24%]	*=D11/D14* [37,94%]
12	Universitaria	*=B4* [25%]	*=C4* [90%]	*=B12*C12* [22,5%]	*+D12/D14* [35,57%]
13	Postgrado	*=B5* [5%]	*=C5* [95%]	*=B13*C13* [4,75%]	*+D13/D14* [7,51%]
14	Total.			*=D10+D11+D12+D13* [63,25%]	

Obsérvese cómo las probabilidades o proporciones iniciales de los niveles educativos quedaron "revisadas", después de que se obtuvo la información de la persona seleccionada trabaja. Nótese que las probabilidades de la columnas *B* y *C* son datos conocidos o estimados y no dependen de la teoría de la probabilidad; las probabilidades de las columnas *D* y *E* son calculadas, de acuerdo con la teoría de la probabilidad. El análisis bayesiano conduce a revisar las probabilidades asignadas tipo de educación si se sabe el resultado de la muestra.

En el segundo ejemplo, se cuenta con la siguiente información:
P(A1) = 30% y *P(A2)* = 70% proporciones (probabilidades) por resultados.
P(B1|A1)= 65%, proporción de veces que el estudio dice que va a resultar el evento *A1*, cuando en realidad va a ser *A1*.
P(B2|A1)=1-65% = 35% proporción de veces que el estudio se equivoca cuando en realidad el resultado va a ser *A1*.
P(B2|A2)= 60%, proporción de veces que el estudio dice que va a resultar el evento *A2*, cuando en realidad va a ser *A2*.
P(B1|A2)=1-60% = 40% proporción de veces que el estudio se equivoca cuando en realidad el resultado va a ser *A2*.

La probabilidad que la inversión resulte en *A1* si el estudio resulta en *B1*, en términos de probabilidades condicionales se calcula a continuación.

La probabilidad de que el estudio resulte en *B1* y ocurra *A1*:

$P(A1 \cap B1) = P(A1)xP(B1/A1)$=30% 65%= 19,5%

La probabilidad de que el estudio resulte en *B2* y ocurra *A1*:

$P(A2 \cap B1) = P(A2)xP(B1/A2)$= 70%x40% = 28%

Si el estudio indica *B1*, esto puede ocurrir siendo cierto que ocurra *A1* o siendo falso que ocurra *A1*, sino que ocurra *A2* y estas posibilidades son excluyentes, por lo tanto, esas probabilidades se pueden sumar y se obtiene la probabilidad, o proporción de veces que el estudio dice *B1*, así:

$P(B1) = P(A1)xP(B1/A1) + P(A2)xP(B1/A2)$

A esta probabilidad se le conoce como *probabilidad marginal*.

Esta es la probabilidad de que este estudio resulte en *B1*. Pero de los resultados *B1* sólo interesa el que dice la verdad, y de acuerdo con la probabilidad condicional, se tiene:

$$P(A1/B1) = \frac{P(B1 \cap A1)}{P(B1)}$$

$$P(A1/B1) = \frac{P(A1)xP(B1/A1)}{P(A1)xP(B1/A1) + P(A2)xP(B1/A2)}$$

En el ejemplo, usando directamente el *Teorema de Bayes*:

	A	**B**	**C**	**D**	**F**
15	Evento a analizar.	Probabilidades básicas.		Probabilidades calculadas	
16		Probabilidad del evento.	Probabilidad de que el estudio diga lo correcto dado el evento.	Probabilidad conjunta P(A1∪B1) y P(A2∪B1)	Probabilidad del evento (A1 o A2) dado que el estudio dice B1
17	A1	=B4 [30%]	=B2 [65%]	=B17*C17 [19,5%]	=D17/D19 [41,05%]
18	A2	=C4 [70%]	=C3 [60%]	=B18*(1-C18) [28%]	=D18/D19 [58,95%]
19	Total.			=D17+D18 [47,5%]	

Obsérvese cómo las probabilidades o proporciones iniciales de los resultados de la inversión quedaron "revisadas", después de que se obtuvo la información de que el estudio dice *B1*. Nótese que las probabilidades de la columnas *B* y *C* son datos conocidos o estimados y no dependen de la teoría de la probabilidad; las probabilidades de las columnas *D* y *E* son calculadas, de acuerdo con la teoría de la probabilidad. El análisis bayesiano conduce a revisar las probabilidades asignadas a los resultados de la inversión, dependiendo de los resultados del estudio.

Ahora el valor esperado de esta inversión será:

$E(A)$ = \$7x0,4105-\$3x(1-0,4105) = 2,8735-1,7685 = 1,1050

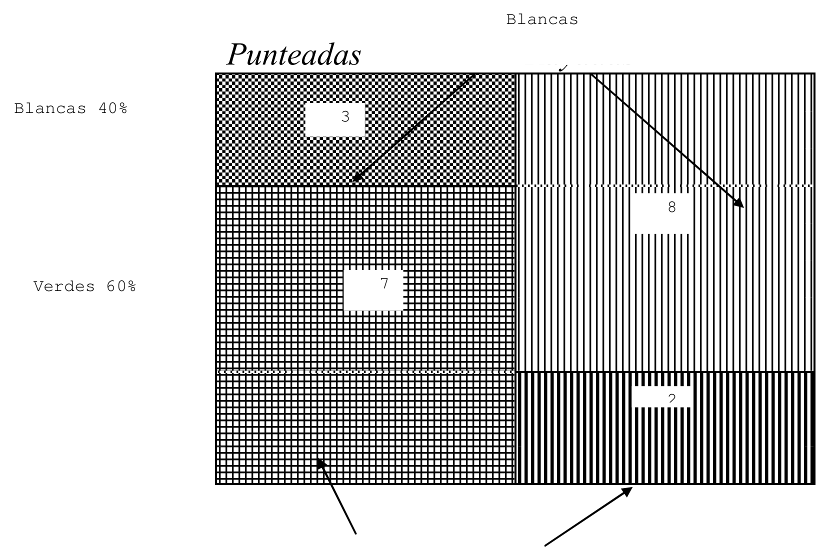
Ejemplo

Supóngase ahora que se tiene una urna con balotas blancas y verdes, y que, a su vez, estas tienen rayas y puntos. La distribución es la siguiente:

	A	**B**	**C**
1		Blancas	Verdes
2	Punteadas.	30%	70%
3	Rayadas.	80%	20%
4	% por color.	40%	60%

Obsérvese que los porcentajes suman 100% en sentido horizontal, esto es, que los porcentajes asignados a las balotas con puntos son las *probabilidades condicionales* que sean blancas o verdes, dado que la balota tenga puntos y lo mismo para las balotas con rayas. En cambio, los porcentajes asignados por color no son probabilidades condicionales, sino “*incondicionales*”, esto es, que no dependen de si son o no balotas con puntos o con rayas. Si se extrae una balota de esta urna, ¿cuál es la probabilidad de que sea blanca, si la balota que ha salido es rayada?

Gráficamente se tiene:



	A Verdes	B	C
1		Blancas	Verdes
2	Puntos.	30%	70%
3	Rayas.	80%	20%
4	% por color.	40%	60%
5	Probabilidad que la rayada que ha salido sea blanca.	=B3*B4 [32%]	
6	Probabilidad que la rayada que ha salido sea verde.	=C3*C4 [12%]	
7	Probabilidad que la rayada que ha salido sea blanca.	=B5/(B5+B6) [72,73%]	

Obsérvese que antes de obtener la información adicional, la probabilidad (la proporción que había en la urna) que se le asignó a una balota blanca era de 40%; después de contar con la información adicional de que la balota seleccionada era rayada, esa probabilidad ascendió a 72,73%.

En el tercer ejemplo, las probabilidad de que sea rayada, en términos de probabilidades condicionales, es:

La probabilidad de que sea rayada y blanca:

$$P(B \cap R) = P(B) \times P(R / B)$$

La probabilidad de que sea rayada y verde:

$$P(V \cap R) = P(V) \times P(R / V)$$

La balota será rayada y blanca o rayada y verde; por lo tanto, esas probabilidades se pueden sumar y se obtiene la probabilidad de que sea rayada, así:

$$P(R) = P(B) \times P(R / B) + P(V) \times P(R / V)$$

NOTA

A esta probabilidad se le conoce como *probabilidad marginal*.

Esta es la probabilidad de que sea rayada. De acuerdo con la probabilidad condicional, se tiene:

$$P(B|R)=\frac{(P(B\cap R))}{P(R)}$$

$$P(B\mid R)=\frac{P(B\cap R)}{P(B)\times P(R\mid B)+P(V)\times P(R\mid V)}$$

$$P(B\mid R)=\frac{P(B)\times P(R/B)}{P(B)\times P(R\mid B)+P(V)\times P(R\mid V)}$$

En el ejemplo, usando directamente el *Teorema de Bayes*:

	A	B	C	D	E	F
8	Evento a analizar.	Probabilidades básicas.				
9		Probabilidad del evento.	Probabilidad de rayada dado el evento blanca.	Probabilidad de rayada dado el evento verde.	Probabilidad conjunta P(B∩R) y P(V∩R).	Probabilidad del evento dado que es rayada.
10	Blanca.	=B4 [40%]	=B3 [80%]		=B10*C10 [32%]	=E10/E12 [72,73%]
11	Verde.	=C4 [60%]		=C3 [20%]	=B11*D11 [12%]	=E11/E12 [27,27%]
12	Total.	=B10+B11 [100%]			=D10+D11 [44%]	

NOTA

Nótese que las probabilidades de la columnas *B*, *C* y *D* son datos conocidos o estimados y no dependen de la teoría de probabilidad; las probabilidades de las columnas *E* y *F* son calculadas, de acuerdo con la teoría de la probabilidad.

Independencia Estadística

Se dice que dos eventos son *estadísticamente independientes* cuando la ocurrencia de uno de ellos no afecta la probabilidad de ocurrencia del otro. Esto ocurre cuando se tienen ciertas combinaciones de probabilidades. Técnica y formalmente se dice que un evento *A* es estadísticamente independiente de otro evento *B* si y sólo si, se cumple la siguiente condición:

$$P(A|B)=P(A) \qquad (10)$$

Esta expresión es muy importante tenerla presente. Cuando se dice de independencia estadística, se refiere en forma exclusiva a que se cumpla la anterior condición. No se hace referencia a ninguna otra clase de independencia, por ejemplo, independencia física al poder lanzar un dado o una moneda, o independencia lógica de poder hacerlo en “forma independiente”, etc.

Si esta condición se cumple, entonces cuando *dos eventos son estadísticamente independientes,* la relación conocida como *regla multiplicativa* queda así:

$$P(A \cap B) = P(B) \times P(A) \qquad (11)$$

Se comete un error con frecuencia, al aplicar la regla multiplicativa de eventos estadísticamente independientes, a eventos que no lo son. Hay que ser cuidadoso en su uso. Sólo existe independencia estadística cuando la ocurrencia de un evento, no afecta la probabilidad de ocurrencia del otro.

Ejemplo

Si en el ejemplo anterior se tuvieran los siguientes datos:

	A	**B**	**C**
1		Blancas	Verdes
2	Puntos.	12%	18%
3	Rayas.	28%	42%
4	% por color.	40%	60%
5	Probabilidad que sea rayada.	=B3+C3 [70%]	
6	Probabilidad que sea rayada y blanca.	=B3 [28%]	
7	Probabilidad que la rayada que ha salido, sea blanca.	=B6/B5 [40%]	

Es decir, que la probabilidad de que sea blanca –“*a priori*”-, es 40%, y la probabilidad de que sea blanca, dado que salió rayada es también 40%, lo cual indica que con las probabilidades que se asignaron al ejemplo, en este último caso, hay independencia estadística entre el evento balotas rayada o punteada y balota verde o blanca.

Otra vez, no importa que conceptualmente “el ser balota rayada no tiene nada que ver con ser balota blanca”, sino que lo importante es que las probabilidades guarden ciertas proporciones entre sí, para que se de la independencia estadística. Se puede pensar en si un evento aporta o no, más información de la que se tenía, para considerar la independencia estadística.

Referencias bibliográficas

Drake, A.W. Fundamentals of Applied Probability Theory, McGraw-Hill, 1967.