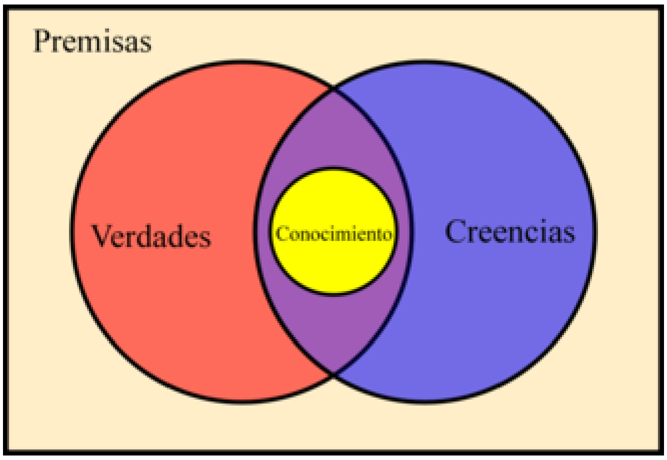
Tratamiento de la Incertidumbre

Incertidumbre

El conocimiento humano está lleno de incertidumbre e imprecisión. La incertidumbre implica no tener un conocimiento seguro y claro de algo, ya que así es como se define la certidumbre. Luego, significa que nuestro conocimiento no es muy seguro o confiable o que no lo tenemos claramente expresado o asimilado.

La incertidumbre se refiere a situaciones que implican información imperfecta, incompleta o desconocida. La incertidumbre se aplica a las predicciones de eventos futuros, a las mediciones físicas que ya se han realizado o a lo desconocido.



La incertidumbre se presenta en entornos parcialmente observables y / o estocásticos, así como debido a la ignorancia, la indolencia o ambas. En campos como los seguros, filosofía, física, estadística, economía, finanzas, psicología, sociología, ingeniería, metrología, meteorología, ecología y ciencias de la información resulta un aspecto importante a considerar.

La imprecisión se refiere a la calidad de aquello que medimos, al grado de exactitud o fidelidad con que estamos comparando dos cosas para saber si son iguales o no, en qué medida se asemejan o no. Depende de la exactitud conque puedo medir algo, por ejemplo, en centímetros o milímetros con una cinta métrica.

Otros aspectos a considerar son:

- la granularidad conque observo un fenómeno determinado por ejemplo el clima: una hora determinada, un día, un año, etc.
- contexto: no es lo mismo la velocidad máxima dentro de la ciudad que en la autopista. No es lo mismo la cantidad de ruido permitido durante el día que durante la noche.
- percepción: cómo se percibe el tamaño del coche según el contexto (país, zona), cómo se percibe la ropa que vestimos según el lugar (nuestra casa, la casa del vecino, el trabajo, el supermecado, un centro comercial, una fiesta)

Los esquemas habituales de representación del conocimiento no contemplan la incertidumbre inherente a la experiencia humana. Estos esquemas han de ser complementados con sistemas de representación de la incertidumbre. Debido a su enfoque principal en la representación del conocimiento y el razonamiento, la inteligencia artificial se ha visto obligada a lidiar con varios marcos para el manejo de la incertidumbre: la teoría de la probabilidad, la teoría de la posibilidad, la teoría de la evidencia y las probabilidades imprecisas.

El conocimiento queda representado por:

- un esquema de representación
- un método de representación de la incertidumbre

MYCIN, uno de los primeros y más conocidos sistemas expertos basados en reglas, ya proponía una técnica ad hoc para la propagación de la incertidumbre basada en grados de creencia e incredulidad.

La imprecisión se define como la falta de exactitud o precisión y se asocia con ambigüedad, generalidad, indeterminación, vaguedad, inexactitud, equivocación, desconocimiento, siendo lo contrario al rigor y la claridad.

Ejemplos cotidianos de incertidumbre e imprecisión:

- No es raro que un médico ponga un tratamiento a un paciente a partir de unos síntomas ambiguos.
- Entendemos frases del lenguaje “entre líneas”.
- Toma de decisiones en base a información imprecisa y/o incierta.

Hay incertidumbre debido a muchas causas:

- Insuficiente experiencia.
- Inadecuada representación del conocimiento.
- Información poco fiable.
- No completitud.
- Inexactitud inherente al lenguaje.

En teoría de la decisión, estadística y otros campos cuantitativos han definido la incertidumbre, el riesgo y su medición como:

- Incertidumbre: La falta de certeza, un estado de conocimiento limitado donde es imposible describir exactamente el estado existente, un resultado futuro o más de un resultado posible.

- Medida de incertidumbre: Un conjunto de posibles estados o resultados donde se asignan probabilidades a cada posible estado o resultado; esto también incluye la aplicación de una función de densidad de probabilidad a variables continuas.
- Incertidumbre de segundo orden: En estadística y economía, la incertidumbre de segundo orden se representa en las funciones de densidad de probabilidad sobre las probabilidades (de primer orden). Las opiniones en lógica subjetiva conllevan este tipo de incertidumbre.
- Riesgo: Un estado de incertidumbre donde algunos resultados posibles tienen un efecto no deseado o una pérdida significativa.
- Medida de riesgo: Un conjunto de incertidumbres medidas donde algunos posibles resultados son pérdidas y las magnitudes de esas pérdidas; esto también incluye funciones de pérdida sobre variables

Inferencia

La inferencia es el proceso por el cual se derivan conclusiones a partir de premisas o hipótesis iniciales. Cuando una conclusión se sigue de sus premisas o hipótesis de partida, por medio de deducciones lógicas válidas, se dice que las premisas implican la conclusión.

La inferencia es el objeto de estudio tradicional de la lógica, así como la vida es el objeto de estudio de la biología. La lógica investiga los fundamentos por los cuales algunas inferencias son aceptables, y otras no. Cuando una inferencia es aceptable, lo es por su estructura lógica y no por el contenido específico del argumento o el lenguaje utilizado (retórica). Por esto se construyen sistemas lógicos que capturan los factores relevantes de las deducciones que aparecen en el lenguaje natural.

Tradicionalmente, se distinguen tres clases de inferencias:

1. las deducciones,
2. las inducciones y
3. las abducciones, aunque a veces se cuenta a la abducción como un caso especial de inducción.

En las investigaciones sobre la inteligencia artificial, la inferencia es la operación lógica utilizada en los motores de inferencia de los sistemas expertos.

Razonamiento Deductivo

El razonamiento deductivo o deducción es el proceso de sacar inferencias deductivas. Una inferencia es deductivamente válida si su conclusión se sigue lógicamente de sus premisas, es decir, si es imposible que las premisas sean verdaderas y la conclusión falsa. Por ejemplo, la inferencia de las premisas "todos los hombres son mortales" y "Sócrates es hombre" a la conclusión "Sócrates es mortal" es deductivamente válida. Un argumento es sólido (sound) si es válido y todas sus premisas son verdaderas. Algunos teóricos definen la deducción en términos de las intenciones del autor para facilitar la distinción entre el razonamiento deductivo válido y el inválido.

La lógica se centra en la relación deductiva de consecuencia lógica entre las premisas y la conclusión o en cómo la gente debe sacar inferencias. Algunos entienden esta relación en términos de los posibles valores de verdad de interpretaciones. Otros, en cambio, se centran en las reglas de inferencia válidas. Una regla de inferencia es un esquema para sacar una conclusión de un conjunto de premisas basándose únicamente en su forma lógica. Hay varias reglas de inferencia, como el modus ponens y el modus tollens. Las deducciones son estudiadas por la mayor parte de la lógica contemporánea.

Razonamiento Inductivo

El razonamiento inductivo o inducción es una forma de razonamiento en que la conclusión se apoya en la verdad de las premisas, pero no la garantizan. Un ejemplo clásico de razonamiento inductivo es:

- Todos los cuervos observados hasta el momento han sido negros
- Por lo tanto, todos los cuervos son negros

En principio, podría ser que el próximo cuervo que se observe no sea negro.

En contraste a los razonamientos deductivos, los razonamientos inductivos tienen la ventaja de ser ampliativos, es decir que la conclusión contiene más información de la que hay contenida en las premisas. Dada su naturaleza ampliativa, los razonamientos inductivos son muy útiles y frecuentes en la ciencia y en la vida cotidiana.

Sin embargo, dada su naturaleza falible, su justificación resulta problemática. ¿Cuándo estamos justificados en realizar una inferencia inductiva, y concluir, por ejemplo, que todos los cuervos son negros a partir de una muestra limitada de ellos? ¿Qué distingue a un buen argumento inductivo de uno malo? Estos y otros problemas relacionados dan lugar al problema de la inducción, cuya vigencia e importancia continúa desde hace siglos.

La lógica inductiva estudia las maneras de medir la probabilidad de que una conclusión sea verdadera, así como las reglas para construir argumentos inductivos fuertes. A diferencia de los razonamientos deductivos, en los razonamientos inductivos no existe acuerdo sobre cuándo considerar un argumento como válido.

De este modo, se hace uso de la noción de «fuerza inductiva» que hace referencia al grado de probabilidad de que una conclusión sea verdadera cuando sus premisas son verdaderas. Así, un argumento inductivo es fuerte cuando es altamente improbable que su conclusión sea falsa si las premisas son verdaderas.

La inferencia estadística es el proceso de utilizar el análisis de datos para inferir propiedades de una distribución de probabilidad subyacente. El análisis estadístico inferencial infiere propiedades de una población, por ejemplo, al probar hipótesis y derivar estimaciones. Se supone que el conjunto de datos observados se muestrea de una población más grande.

Su objetivo es obtener conclusiones útiles para hacer razonamientos deductivos sobre una totalidad, basándose en la información dada por la muestra. Se dedica a la generación de los modelos y predicciones asociadas a los fenómenos en cuestión teniendo en cuenta la aleatoriedad de las observaciones. Se usa para modelar patrones en los datos y extraer inferencias acerca de la población bajo estudio.

Teorema de Bayes

La incertidumbre del conocimiento se puede modelar utilizando probabilidades. Las probabilidades se basan en mediciones que se asumen exactas. Se basa en la teoría formal de probabilidad del teorema de Bayes. Éste permite el cálculo de probabilidades complejas a partir de otras más simples basadas en observaciones reales.

Sabemos que:

- la probabilidad P de un suceso S cualquiera es $0 \leq P(S) \leq 1$
- Si $A \subset B$ entonces $P(A) \leq P(B)$
- $P(A - B) = P(A) - P(A \cap B)$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- Si los eventos A y B son independientes entonces $P(A \cup B) = P(A) + P(B)$ y $P(A \cap B) = P(A)P(B)$

El caso de $P(A \cap B)$ cuando los eventos no son independientes requiere otro concepto que veremos más adelante.

Dado un conjunto de sucesos S_i independientes (i.e., mutuamente excluyentes) entonces: $P(\cup S_i) = \sum P(S_i)$

La probabilidad de que ocurra un evento A, sabiendo que también sucede otro evento B (por lo cual B tiene probabilidad estrictamente positiva) se denomina **probabilidad condicional**, la cual se escribe $P(A|B)$ o $P(A/B)$, y se lee «la probabilidad de A dado B».

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

.

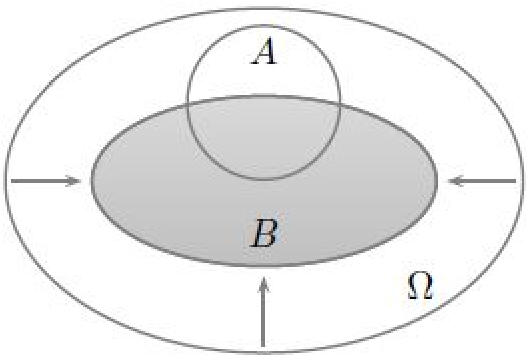
El evento B representa un evento que ya ha ocurrido, y la probabilidad condicional $P(A|B)$ es la probabilidad de A modificada con la información adicional de que B ha ocurrido.

De lo anterior se deduce que $P(A \cap B) = P(A|B)P(B)$ cuando los eventos no son excluyentes.

La probabilidad condicional $P(A|B)$ de los sucesos A y B se puede interpretar como la relación causa efecto entre A y B, siendo A la evidencia que soporta la hipótesis B. Por ejemplo, evidencia (A) tengo fiebre, hipótesis (B) tengo gripe, pero, en realidad, la gripe es la que provoca que tenga fiebre, por lo que la causa de que tengo fiebre es la gripe.

Uno puede imaginar que el espacio muestral Ω del experimento aleatorio se ha reducido al evento B de tal forma que todo lo que se encuentre fuera de este evento tiene probabilidad condicional cero.

La afirmación anterior es evidente a partir de observar que, si A y B son excluyentes, $\Rightarrow A \cap B = \varnothing \Rightarrow P(A \cap B) = 0 \Rightarrow P(A|B) = 0$, de donde la probabilidad condicional es mayor que 0 si y solo sí A y B están relacionados.



Ejemplo: considere el experimento de lanzar un dado equilibrado y defina los eventos:

$A = 2$ si se obtiene el número 2,

$B = 2, 4, 6$ si se obtiene un número par.

Está claro que $P(A) = 1/6$, sin embargo, sabiendo que B ha ocurrido, es decir, sabiendo que el resultado es un número par, la probabilidad del evento A es ahora:

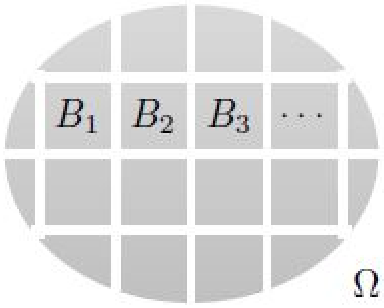
$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(\{2\})}{P(\{2,4,6\})} = \frac{1/6}{3/6} = \frac{1}{3}$$

Es decir, la información adicional de la ocurrencia del evento B ha hecho que la probabilidad de A se incremente de 1/6 a 1/3.

Sea Ω el espacio muestral de un experimento aleatorio. Decimos que la colección de eventos $\{B_1, \cdots, B_n\}$ es una partición finita de Ω si se cumplen las siguientes condiciones:

- $B_i \neq \emptyset, i = 1, \cdots, n.$
- $B_i \cap B_j = \emptyset, i \neq j.$
- $\cup_{i=1}^{n} B_i = \Omega$

Así, se requiere que cada uno de los elementos de una partición sea distinto del conjunto vacío, que sean excluyentes dos a dos y que la unión de todos ellos constituya la totalidad del espacio muestral. De manera grafica podemos representar una partición finita como se muestra a continuación.



Teorema de la Probabilidad Total: dada una partición $\{B_1, \cdots, B_n\}$ de Ω tal que $P(B_i) > 0, i = 1, \cdots, n$, para cualquier evento A se cumple que:

$$P(A) = \sum_{i=1}^{n} P(A|B_i)P(B_i)$$

NOTA: la hipótesis mutuamente excluyentes e independientes no es posible garantizarla en la vida real.

El **Teorema de Bayes** en la Teoría de la Probabilidad se debe al matemático inglés Thomas Bayes (siglo XVIII) que vincula la probabilidad del evento A dado el evento B con la probabilidad de B dado A.

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

donde A y B son eventos y $P(B) > 0$.

El teorema de Bayes se deduce a partir de la probabilidad condicional

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

y

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

entonces

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

de donde obtenemos el teorema.

No tiene por qué haber una relación causal o temporal entre A y B. A puede preceder en el tiempo a B, sucederlo o pueden ocurrir simultáneamente. A puede causar B, viceversa o pueden no tener relación causal. Las relaciones causales o temporales son nociones que no pertenecen al ámbito de la probabilidad. Pueden desempeñar un papel o no, dependiendo de la interpretación que se le dé a los eventos.

Un ejemplo clásico es el lanzamiento de una moneda para luego lanzar un dado. ¿Cuál es la probabilidad de que en el dado salga un 6 si ya ha salido una cara en la moneda? Esta probabilidad se denota de esta manera: $P(6/C)$.

Luego, sabiendo la probabilidad de tener un dolor de cabeza dado que se tiene gripe, se podría saber (si se tiene algún dato más), la probabilidad de tener gripe si se tiene un dolor de cabeza. Este sencillo ejemplo nos muestra la relevancia de este teorema para la ciencia en todas sus ramas, para la comprensión de la probabilidad de aspectos causales dados los efectos observados.

Ejemplo: Un médico sabe que la meningitis provoca una rigidez en el cuello del paciente, digamos, el 50% de las veces. Conoce también los siguientes hechos incondicionales:

- la probabilidad a priori de que un paciente sufra meningitis es de 1/50000, y
- la probabilidad a priori de que algún paciente padezca de rigidez del cuello es de 1/20.

Si S representa la proposición de que el paciente padezca de rigidez en el cuello y M la proposición de que el paciente tenga meningitis, tenemos que:

$$P(S/M) = 0.5; P(M) = \frac{1}{50000}; P(S) = \frac{1}{20}$$

$$P(M/S) = P(S/M) * \frac{P(M)}{P(S)} = \frac{0.5 * \frac{1}{50000}}{\frac{1}{20}} = 0.0002$$

In [1]:

```
# vamos a definir una función para calcular la probabilidad
# según el teorema de Bayes

def bayes(pBdadoA,pA,pB):
    """
    aplicar teorema de Bayes
    :param pBdadoA: Prob(B dado A)
    :param pA: Prob(A)
    :param pB: Prob(B)
    :return: Prob(A dado B)
    """
    return pBdadoA*pA/pB

pM = 1/50000
pS = 1/20
pSdadoM = 0.5
bayes(pSdadoM,pM,pS)
```

Out[1]:

```
0.0002
```

Una de las muchas aplicaciones del teorema de Bayes es la inferencia bayesiana, un enfoque particular de la inferencia estadística. La inferencia bayesiana es un tipo de inferencia estadística en la que las evidencias u observaciones se emplean para actualizar o inferir la probabilidad de que una hipótesis pueda ser cierta.

Cuando se aplican, las probabilidades implicadas en el teorema pueden tener diferentes interpretaciones de probabilidad. Con la interpretación probabilidad bayesiana, el teorema expresa cómo un grado de creencia, expresado como una probabilidad, debería cambiar racionalmente para tener en cuenta la disponibilidad de pruebas relacionadas. La inferencia bayesiana es fundamental para la estadística bayesiana, siendo considerada como; "a la teoría de la probabilidad lo que el teorema de Pitágoras es a la geometría”

Sea $\{A_1, A_2, \ldots, A_i, \ldots, A_n\}$ un conjunto de sucesos mutuamente excluyentes y exhaustivos tales que la probabilidad de cada uno de ellos es distinta de cero $P(A_i) \neq 0$ para $i = 1, 2, \ldots, n$. Si B es un suceso cualquiera del que se conocen las probabilidades condicionales $P(B|A_i)$ entonces la probabilidad $P(A_i|B)$ viene dada por la expresión:

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

donde:

- $P(A_i)$ son las probabilidades a priori,
- $P(B|A_i)$ es la probabilidad de B en la hipótesis A_i, y
- $P(A_i|B)$ son las probabilidades a posteriori.

En base a la definición de probabilidad condicionada se obtiene la Fórmula de Bayes, también conocida como Regla de Bayes:

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{k=1}^{n} P(B|A_k)P(A_k)}$$

la cual nos permite calcular la probabilidad condicional $P(A_i|B)$ de cualquiera de los eventos A_i dado B.

Supongamos que una prueba particular para determinar si alguien ha estado consumiendo cannabis conduce a un 90% de identificación correcta del consumo de cannabis (resultados positivos verdaderos, True Positive Rate, TPR = 0,90). En este caso, se dice que la prueba tiene una sensibilidad del 90%.

Por otra parte, supongamos que la prueba identifica correctamente el 80 % de las personas que no consumen cannabis, lo que significa una tasa de negativos verdaderos (True Negative Rate) TNR = 0,80. Pero, también genera un 20 % de falsos positivos, (False Positive Rate) FPR = 0,20, para los no consumidores. En este caso, se dice que la prueba tiene una especificidad del 80 %.

Suponiendo que el 5 % de las personas consumen cannabis (i.e., se tiene una prevalencia de 0,05), ¿cuál es la probabilidad de que una persona al azar que dé positivo sea realmente un consumidor de cannabis?

El valor predictivo positivo (VPP) de una prueba es la proporción de personas que realmente son positivas de todas las que dieron positivo, la cual se puede calcular a partir de una muestra como:

$$\text{VPP} = \frac{\text{Verdadero positivo}}{\text{Probado positivo}} = P(Usuario \mid Positivo)$$

teniendo en cuenta que $P(\text{Usuario} \mid \text{Positivo})$ significa "la probabilidad de que alguien sea consumidor de cannabis dado que da positivo". Entonces, podemos escribir:

$$P(\text{Usuario} \mid \text{Positivo})$$
$$= \frac{P(\text{Positivo} \mid \text{Usuario})P(\text{Usuario})}{P(\text{Positivo} \mid \text{Usuario})P(\text{Usuario}) + P(\text{Positivo} \mid \text{No-Usuario})P(\text{No-Usuario})}$$

In [2]:

```
# ¿qué tenemos?
# 90% identificación correcta de consumo (TPR True Positive Rate)
prob_pos_user = 0.90 # prob. dar positivo dado que es usuario
# 80 % identificación correcta no consumidores (TNR True Negative Rate)
prob_neg_no_user = 0.80
# False Positive Rate FPR = 0,20, para los no consumidores
prob_pos_no_user = 0.20 # prob. dar positivo dado que no es usuario
# Suponiendo el 5 % de las personas consumen cannabis
prob_user = 0.05 # =>
prob_no_user = 0.95
```

In [3]:

```
prob_user_pos = (prob_pos_user*prob_user)/(prob_pos_user*prob_user
                                            + prob_pos_no_user*prob_no_user)
```

$$= \frac{0.90 \times 0.05}{0.90 \times 0.05 + 0.20 \times 0.95} = \frac{0.045}{0.045 + 0.19} \approx 19\%$$

In [4]:

```
prob_user_pos
```

Out[4]:

```
0.19148936170212766
```

El hecho

$$P(\text{Positive}) = P(\text{Positive} \mid \text{User})P(\text{User}) + P(\text{Positive} \mid \text{Non-user})P(\text{Non-user})$$

es una aplicación directa de la Ley de Probabilidad Total. En la teoría de la probabilidad, esta ley expresa la probabilidad total de un resultado que se puede realizar a través de varios eventos distintos y relaciona las probabilidades marginales con las probabilidades condicionales.

La suma se puede interpretar como un promedio ponderado y, en consecuencia, la probabilidad marginal, $P(A)$, a veces se denomina "probabilidad promedio". También se utiliza denominarla "probabilidad general" en escritos menos formales.

Volviendo al caso del test de cannabis, la probabilidad de que alguien dé positivo es la probabilidad de que un usuario dé positivo, multiplicada por la probabilidad de ser usuario, más la probabilidad de que un no usuario dé positivo, multiplicada por la probabilidad de no ser usuario. Esto es cierto porque las clasificaciones de usuario y no usuario forman una partición de un conjunto, a saber, el conjunto de personas que se someten a la prueba de drogas.

In [5]:

```
# Esta función permite calcular la regla de Bayes
def bayesG(i,pBdadoAi,pAi):
    """
    calcula la probabilidad condicionada Prob(Ai dado B)
    :param i: indice del elemento a calcular Prob(Ai dado B)
    :param pBdadoAi: vector con las Prob(B dado Ak)
    :param pAi: vector con las Prob(Ak)
    :return: probabilidad condicionada Prob(Ai dado B)
    """

    s=0
    for k in range(len(pAi)):
        s = s + pBdadoAi[k]*pAi[k]
    return pBdadoAi[i]*pAi[i]/s
```

In [6]:

```
# pBdadoAi: vector con las Prob(B dado Ak)
# prob(positivo dado consumidor), prob(positivo dado no consumidor)

pBdadoAi = [prob_pos_user, prob_pos_no_user]

# pAi: vector con las Prob(Ak)
# prob(consumidor), prob(no consumidor)

pAi = [prob_user, prob_no_user]

bayesG(0,pBdadoAi,pAi)
```

Out[6]:

```
0.19148936170212766
```

En otras palabras, incluso si alguien da positivo, la probabilidad de que sea un consumidor de cannabis es solo del 19%, lo cual se debe a que, en este grupo, solo el 5% de las personas son usuarios, y la mayoría de los positivos son falsos positivos provenientes del 95% restante.

Si se hiciera la prueba a 1000 personas,

- 95% son no usuarios (=950) y 5% de ellos son usuarios (=50) y
- 20% de ellos dan falso positivo (0,20 × 950 = 190)
- 45 de ellos dan verdadero positivo (0,90 × 50)

Las 1.000 personas arrojan así 235 pruebas positivas, de las cuales sólo 45 son auténticos consumidores de drogas, alrededor del 19%.

```
prob_pos_user = 0.90 # identificación correcta usuario (TPR True Positive Rate)
# 80 % identificación correcta no consumidores (TNR True Negative Rate)
prob_pos_no_user = 0.20 # False Positive Rate FPR para no consumidores
prob_user = 0.05 # Suponiendo 5 % consumen cannabis, de donde
prob_no_user = 0.95 # el resto no consumen
```

In [7]:

```
non_user = 950
user = 50
total = non_user + user
total
```

Out[7]:

```
1000
```

Actual \ Test	Positive	Negative	Total
User	**45**	5	50
Non-user	190	760	950
Total	**235**	765	1000
90% sensitive, 80% specific, PPV=45/235 ≈ 19%			

La importancia de la especificidad se puede ver al mostrar que incluso si la sensibilidad aumenta al 100% y la especificidad permanece en el 80 %, la probabilidad de que alguien que dio positivo sea realmente un consumidor de cannabis solo aumenta del 19% al 21%

In [8]:

```
# sensibilidad = 100%
prob_pos_user = 1.00 # identificación correcta usuario (TPR True Positive Rate)
prob_user_pos = (prob_pos_user*prob_user)/(prob_pos_user*prob_user
                                           + prob_pos_no_user*prob_no_user)
prob_user_pos
```

Out[8]:

```
0.20833333333333334
```

Actual \ Test	Positive	Negative	Total
User	**50**	0	50
Non-user	190	760	950
Total	**240**	760	1000
100% sensitive, 80% specific, PPV=50/240 ≈ 21%			

Pero si la sensibilidad es mantenida en 90% y la especificidad aumenta a 95%, la probabilidad aumenta a 49%.

In [9]:

```
# sensibilidad = 90%
prob_pos_user = 0.90 # identificación correcta usuario (TPR True Positive Rate)
# especificidad = 95% (True Negative Rate) TNR
# antes especificidad = 80 % (TNR True Negative Rate)
# prob_pos_no_user = 0.20 # False Positive Rate FPR para no consumidores
prob_pos_no_user = 0.05

prob_user_pos = (prob_pos_user*prob_user)/(prob_pos_user*prob_user
                                           + prob_pos_no_user*prob_no_user)
prob_user_pos
```

Out[9]:

```
0.48648648648648657
```

Test / Actual	Positive	Negative	Total
User	**45**	5	50
Non-user	47	903	950
Total	**92**	908	1000

90% sensitive, 95% specific, PPV=45/92 ≈ 49%

La probabilidad condicional de una hipótesis h, dada la presencia de dos evidencias e_1 y e_2 está dada por :

$$P(h/(e_1 \wedge e_2)) = \frac{P(h) * P((e_1 \wedge e_2)/h)}{P(e_1 \wedge e_2)}$$

La probabilidad condicional de una hipótesis h, dada la presencia de n evidencias $e_1, \cdots, e_n$ será:

$$P(h/e_1 \wedge \cdots \wedge e_n) = P(h) * P(e_1 \wedge \cdots \wedge e_n/h)/P(e_1 \wedge \cdots \wedge e_n)$$

En caso de que tengamos k hipótesis mutuamente excluyentes y exhaustivas:

$$P(h_i \wedge h_j) = 0, \forall i, j \mid i \neq j$$

y

$$\sum_j P(h_j) = 1$$

y además hay n evidencias que las soportan $e_1, \cdots, e_n$ independientes, por lo cual $P(e_i/e_j) = P(e_i)$ entonces:

$$P(h_i/e_1 \wedge \cdots \wedge e_n) = \frac{[P(h_i) * P(e_1/h_i) * \cdots * P(e_n/h_i)]}{\sum_j [P(h_j) * P(e_1/h_j) * \cdots * P(e_n/h_j)]}$$

donde

$P(h_i/e_1 \wedge \cdots \wedge e_n)$: Probabilidad de la hipótesis h_i dadas las evidencias $e_1 \cdots e_n$.

$P(h_i)$: Probabilidad de la hipótesis h_i.

$P(e_i/h_j)$: Probabilidad de observar la evidencia e_i dado que la hipótesis h_j sea cierta.

Ejemplo: Sean las evidencias:

- $e1$: soltero
- $e2$: ingresos_altos
- $e3$: joven

que soportan las hipótesis:

- $h1$: inversor_de_alto_riesgo
- $h2$: inversor_de_bajo_riesgo

donde $h1$ y $h2$ son mutuamente exclusivas y exhaustivas

$$P(h1 \wedge h2) = 0,$$
$$P(h1) = 1 - P(h2)$$

Un experto financiero viendo sus registros de inversores estima las probabilidades “a posteriori” siguientes:

$$P(h1) = 0.3, P(h2) = 0.7,$$
$$P(e1/h1) = 0.6, P(e1/h2) = 0.4$$
$$P(e2/h1) = 0.2, P(e2/h2) = 0.8$$
$$P(e3/h1) = 0.5, P(e3/h2) = 0.5$$

Supongamos que se pretende predecir el perfil de los inversores de mayor y menor riesgo.

In [10]:

```
ph1 = 0.3
ph2 = 0.7
pe1h1 = 0.6
pe1h2 = 0.4
pe2h1 = 0.2
pe2h2 = 0.8
pe3h1 = 0.5
pe3h2 = 0.5
```

In [11]:

```
# Las probabilidades “a priori” serán las siguientes:
# P(h1/e1) = [ P(h1)P(e1/h1)]/[ P(h1)P(e1/h1) + P(h2)P(e1/h2)
ph1e1 = (ph1*pe1h1)/(ph1*pe1h1 + ph2*pe1h2)
ph1e1
```

Out[11]:

```
0.391304347826087
```

In [12]:

```
# vamos a hacer una función para realizar este cálculo, para ello
# nos conviene definir las probabilidades a posteriori mediante listas
ph = [0.3, 0.7]
peh = [
        [0.6, 0.4],
        [0.2, 0.8],
        [0.5, 0.5]
      ]
```

In [13]:

```
# P(h1/e1) = [ P(h1)P(e1/h1)]/[ P(h1)P(e1/h1) + P(h2)P(e1/h2)
def prob1(hi,ej):
    hi = hi - 1
    ej = ej - 1
    num = ph[hi]*peh[ej][hi]
    denom = ph[0]*peh[ej][0] + ph[1]*peh[ej][1]
    return num/denom

for i in [1,2]:
    for j in [1,2,3]:
        print('prob(h',i,'/e',j,')=',prob1(i,j))
```

```
prob(h 1 /e 1 )= 0.391304347826087
prob(h 1 /e 2 )= 0.09677419354838711
prob(h 1 /e 3 )= 0.3
prob(h 2 /e 1 )= 0.6086956521739131
prob(h 2 /e 2 )= 0.903225806451613
prob(h 2 /e 3 )= 0.7
```

In [14]:

```
# P(hi/e1 & .. & en )=[P(hi)*P(e1/hi)*....P(en/hi)]/
#                       sum[P(hj)*P(e1/hj)*....P(en/hj)]
def prob2(hi):
    hi = hi -1
    num = ph[hi]
    for j in [0,1,2]:
        num = num * peh[j][hi]
    denom = 0
    for j in [0,1]:
        prod = ph[j]
        for i in [0,1,2]:
            prod = prod * peh[i][j]
        denom = denom + prod
    return num/denom

for i in [1, 2]:
    print('prob(h', i, '/e1&e2&e3)=', prob2(i))
```

```
prob(h 1 /e1&e2&e3)= 0.13846153846153847
prob(h 2 /e1&e2&e3)= 0.8615384615384616
```

Resumiendo, hemos obtenido:

P.I.	e_1	e_2	e_3	t_1	t_2
h_1	0.3	0.39	0.097	0.3	0.4
h_2	0.7	0.61	0.903	0.7	0.6

donde:

$t_1 \, : e_1 \, \wedge e_3$

Vemos que

- h_2 es la situación más probable con la presencia conjunta de $e1 \wedge e2 \wedge e3$
- $h2$ es la situación más probable con la presencia conjunta de $e1 \wedge e3$

Defectos del Enfoque Probabilista

1. Se necesita saber un número importante de probabilidades $P(h_i)$ y $P(e_j/h_i)$. Estos valores no son siempre fáciles de conseguir o estimar.
2. Las probabilidades requieren muestras representativas.
3. Si se descubren nuevas evidencias que condicionan las hipótesis hay que reconstruir todas las probabilidades.
4. La independencia de las hipótesis es difícil que ocurra
5. Este enfoque asume que la presencia de una evidencia a para una hipótesis c también afecta a la negación de la misma. Así si $P(c/a) = 0.8$, entonces $P(\neg c/a) = 0.2$. Esto no es necesariamente cierto en todos los dominios.

En medicina la presencia de un síntoma no sería una evidencia que soportase a la vez la existencia y la no existencia de una enfermedad al mismo tiempo.

Inferencia bayesiana

La inferencia bayesiana es un tipo de inferencia estadística en la que las evidencias u observaciones se emplean para actualizar o inferir la probabilidad de que una hipótesis pueda ser cierta. El nombre proviene del uso frecuente que se hace del teorema de Bayes durante el proceso de inferencia.

Las técnicas de Inferencia Bayesiana muestran cómo debemos actualizar nuestras creencias a medida que observamos los datos. La actualización bayesiana es particularmente importante en el análisis dinámico de una secuencia de datos. La inferencia bayesiana ha encontrado aplicación en una amplia gama de actividades, incluidas la ciencia, la ingeniería, la filosofía, la medicina, el deporte y el derecho, en la teoría de la decisión, visión artificial y reconocimiento de patrones por ordenador.

La inferencia bayesiana es una forma de hacer inferencias estadísticas en las que el estadístico asigna probabilidades subjetivas a las distribuciones que podrían generar los datos. Estas probabilidades subjetivas forman la llamada distribución previa.

Una vez observados los datos, se utiliza la regla de Bayes para revisar las probabilidades asignadas a las posibles distribuciones generadoras de datos y ajustar la distribución previa. Estas probabilidades revisadas forman la llamada distribución posterior.

La incertidumbre y la imprecisión son connaturales en el proceso de razonamiento. Los métodos de razonamiento aproximado, entre los que se encuentran los métodos bayesianos, aportan modelos teóricos que simulan la capacidad de razonamiento en condiciones de incertidumbre e imprecisión, cuando no se conoce con absoluta certeza la verdad o falsedad de un enunciado o hipótesis, así como enunciados en los que se admite un rango de variación.

Entre los métodos de razonamiento aproximado se encuentran los métodos bayesianos, basados en el conocido teorema. Todos ellos tienen en común la asignación de una probabilidad como medida de credibilidad de las hipótesis. En este contexto, la inferencia se entiende como un proceso de actualización de las medidas de credibilidad al conocerse nuevas evidencias. Mediante la aplicación del Teorema de Bayes se busca obtener las probabilidades de las hipótesis condicionadas a las evidencias que se conocen.

La diferencia entre los distintos métodos bayesianos, modelos causales y redes bayesianas, estriba en las hipótesis de independencia condicional entre hipótesis y evidencias. Dichas relaciones se expresan comúnmente mediante un grafo acíclico dirigido.

La inferencia bayesiana utiliza aspectos del método científico, que implica recolectar evidencia que se considera consistente o inconsistente con una hipótesis dada. A medida que la evidencia se acumula, el grado de creencia en una hipótesis se va modificando. Con evidencia suficiente, a menudo dicho grado de creencia podrá hacerse muy alto o muy bajo.

Así, los que sostienen la inferencia bayesiana dicen que puede ser utilizada para discriminar entre hipótesis en conflicto: las hipótesis con un grado de creencia muy alto deben ser aceptadas como verdaderas y las que tienen un grado de creencia muy bajo deben ser rechazadas como falsas. Sin embargo, los detractores dicen que este método de inferencia puede estar afectado por un sesgo debido a las creencias iniciales que se deben sostener antes de comenzar a recolectar cualquier evidencia.

Supongamos que un proceso genera eventos independientes e idénticamente distribuidos $E_n, n = 1, 2, 3, \ldots$, pero se desconoce la distribución de probabilidad. Supongamos que, el espacio de eventos Ω representa el estado actual de creencia para este proceso.

Cada modelo está representado por el evento M_m. Las probabilidades condicionales $P(E_n \mid M_m)$ se especifican para definir los modelos. $P(M_m)$ es el grado de creencia en M_m. Antes del primer paso de inferencia, $\{P(M_m)\}$ es un conjunto de probabilidades iniciales previas. Estos deben sumar 1, pero por lo demás son arbitrarios.

Suponga que se observa que el proceso genera $E \in \{E_n\}$. Por cada $M \in \{M_m\}$, el $P(M)$ anterior se actualiza al $P(M \mid E)$ posterior. De acuerdo con el Teorema de Bayes:

$$P(M \mid E) = \frac{P(E \mid M)}{\sum_m P(E \mid M_m) P(M_m)} \cdot P(M)$$

Tras la observación de más pruebas, este procedimiento puede repetirse.

Dada una secuencia de observaciones independientes e idénticamente distribuidas $\mathbf{E} = (e_1, \ldots, e_n)$, se puede demostrar por inducción que la aplicación repetida de lo anterior es equivalente a

$$P(M \mid \mathbf{E}) = \frac{P(\mathbf{E} \mid M)}{\sum_m P(\mathbf{E} \mid M_m) P(M_m)} \cdot P(M)$$

donde

$$P(\mathbf{E} \mid M) = \prod P(e_k \mid M).$$

Ejemplo:

Supongamos que hay dos tazones llenos de galletas. El tazón 1 tiene 10 chispas de chocolate y 30 galletas simples, mientras que el tazón 2 tiene 20 de cada uno. Nuestro amigo Fred elige un tazón al azar y luego elige una galleta al azar. Podemos suponer que no hay razón para creer que Fred trata un tazón de manera diferente a otro, al igual que las galletas. La galleta resulta ser simple. ¿Qué tan probable es que Fred lo haya sacado del tazón 1?

Intuitivamente, la respuesta debería ser más de la mitad, ya que hay más galletas simples en el recipiente 1. La respuesta precisa la da el teorema de Bayes. Sea H_1 el tazón 1 y H_{2} el tazón 2. Dado que los tazones son idénticos desde el punto de vista de Fred, entonces $P(H_1) = P(H_2)$, y los dos deben sumar 1, por lo que ambos son iguales a 0.5. El evento E es la observación de una galleta simple. Por el contenido de los tazones, sabemos que $P(E \mid H_1) = \frac{30}{40} = 0.75$ y $P(E \mid H_2) = \frac{20}{40} = 0.5$. La fórmula de Bayes entonces produce

$$P(H_1 \mid E) = \frac{P(E \mid H_1)\, P(H_1)}{P(E \mid H_1)\, P(H_1) \; + \; P(E \mid H_2)\, P(H_2)}$$

$$= \frac{0.75 \times 0.5}{0.75 \times 0.5 + 0.5 \times 0.5}$$

$$= 0.6$$

Antes de que observáramos la galleta, la probabilidad que le asignamos a Fred de haber elegido el tazón 1 era la probabilidad anterior, $P(H_1)$, que era 0,5. Después de observar la galleta, debemos revisar la probabilidad a $P(H_1 \mid E)$, que es 0,6.

Bibliografía

- Incertidumbre (https://es.wikipedia.org/wiki/Incertidumbre)
- Uncertainty (https://en.wikipedia.org/wiki/Uncertainty)
- Inferencia (https://es.wikipedia.org/wiki/Inferencia)
- Razonamiento deductivo (https://es.wikipedia.org/wiki/Razonamiento_deductivo)
- Razonamiento inductivo (https://es.wikipedia.org/wiki/Razonamiento_inductivo)
- Inferencia bayesiana (https://es.wikipedia.org/wiki/Inferencia_bayesiana)
- Statistical inference (https://en.wikipedia.org/wiki/Statistical_inference)
- Teorema de Bayes (https://es.wikipedia.org/wiki/Teorema_de_Bayes)
- Inferencia bayesiana (https://es.wikipedia.org/wiki/Inferencia_bayesiana)
- Probabilidad condicionada (https://es.wikipedia.org/wiki/Probabilidad_condicionada)
- Law of total probability (https://en.wikipedia.org/wiki/Law_of_total_probability)
- Certainty (https://en.wikipedia.org/wiki/Certainty)
- Certeza y opinión (https://es.wikipedia.org/wiki/Certeza_y_opini%C3%B3n)
- Epistemology (https://en.wikipedia.org/wiki/Epistemology)
- Inferencia bayesiana (https://es.wikipedia.org/wiki/Inferencia_bayesiana)
- Bayesian inference (https://en.wikipedia.org/wiki/Bayesian_inference)

- [Bayesian inference (https://www.statlect.com/fundamentals-of-statistics/Bayesian-inference)](https://www.statlect.com/fundamentals-of-statistics/Bayesian-inference)
- [Inferencia Bayesiana (https://seeing-theory.brown.edu/bayesian-inference/es.html)](https://seeing-theory.brown.edu/bayesian-inference/es.html)