UNIVERSIDAD DE BUENOS AIRES
FACULTAD DE
CIENCIAS ECONOMICAS

TEORIA DE LA DESICION

Estructura Bayesianas

Teoría de la información

Profesor: Pedro F. J. Pavesi

CODIGO: 161

Facultad de Ciencias Económicas

Departamento de Teoría de la Administración

INSTITUTO DE INVESTIGACIONES ADMINISTRATIVAS

ESTRUCTURAS BAYESIANAS Y TEORÍA DE LA INFORMACIÓN

Consideraciones acerca de la reducción de la incertidumbre por la información

(Versión preliminar)

Pedro F. J. Pavesi
Profesor Titular Regular
Enero 1986

ESTRUCTURAS BAYESIANAS Y TEORÍA DE LA INFORMACIÓN

I N D I C E

1. Introducción pág. 3

2. Definiciones pág. 3

3. Problemas en el uso de estructuras bayesianas pág. 6

4. Un teorema fundamental: una estructura bayesiana nunca aumenta la incertidumbre pág. 7

5. La certeza a priori pág. 11

6. La información que aumenta la incertidumbre pág. 13

7. Información y restricción: las estructuras de información pág. 15

8. La flecha del tiempo pág. 16

Apéndice A:
Recordatorio de probabilidades - Teorema de Bayes - pág. 18

Apendice B:
La convexidad de las funciones logarítmicas pág. 27

Apéndice C:
Ejemplos pág. 25

Bibliografía pág. 33

ESTRUCTURAS BAYESIANAS Y TEORÍA DE LA INFORMACIÓN

CONSIDERACIONES ACERCA DE LA REDUCCIÓN DE LA INCERTIDUMBRE POR LA INFORMACIÓN

1. Introducción

En situación de incertidumbre, se usa buscar información para reducirla. El teorema de Bayes aporta a la Teoría de la Decisión una importante herramienta para analizar si el precio de la información adicional a obtener es conveniente, debiendo ser menor (o, a lo sumo, igual) al valor de esa información, es decir a los beneficios que la misma puede reportar.

Si bien es cierto que en la Teoría de la Información se usa definir información como todo lo que reduce la incertidumbre y que la medida de ambas es la misma , también es cierto que se presentan innumerables ocasiones en las cuales la información adicional aporta mayor incertidumbre, en lugar de reducirla. Por otra parte, el análisis Bayesiano se revela inútil cuando se parte de una situación de certeza subjetiva.

El presente trabajo establece lo siguiente:

(1) Demuestra la siguiente regla general: en situaciones universales denominadas *estructuras de información* y *estructuras bayesianas*, la incertidumbre promedio a posteriori (de una eventual toma de información adicional) es siempre menor o, en el peor de los casos, igual, a la incertidumbre existente a priori.

(2) Analiza el caso de certeza subjetiva a priori en estructuras bayesianas, la que -de acuerdo con la regla general mencionada anteriormente- no puede ser modificada por la información adicional relevante.

(3) Extiende la regla general a cualquier caso de toma de información adicional, postulando que cualquier situación de información que aumente la incertidumbre está obligatoriamente acompañada por situaciones de información que la disminuyen, cumpliéndose finalmente la mencionada regla general.

(4) Precisa definiciones formales de situaciones de flujo de información de alcance muy general.

El presente trabajo es una versión previa para discusión.

2. Definiciones

En el Apéndice A se incluye un breve y sencillo recordatorio de conceptos de probabilidad y otros correspondientes al Teorema de Bayes, que serán utilizados en este trabajo. La simbolización también es la utilizada en el Apéndice.

Las referencias a las definiciones del punto 3 del Apéndice A se indicarán por la letra A seguida del número de la definición.

(1) El universo es un conjunto de variables que pueden asumir distintos valores, niveles o grados a través del tiempo e interrelacionadas entre sí.
El universo es definido por un observador dado, en un momento determinado y en posesión de cierta información (PAVESI, 1986, (b)).

(2) En un momento dado y para un observador determinado, una variable exhibe incertidumbre (es "incierta") cuando puede asumir en un momento posterior más de un comportamiento. El comportamiento de una variable es el conjunto de valores, niveles o grados que asume a través del tiempo. Un universo es incierto cuando por lo menos una de sus variables es incierta.

(3) Información es, en general, conocimiento acerca del comportamiento de una variable, en un momento dado, para un observador determinado y sus propias circunstancias en ese momento: personalidad, ideología, instrumentos de observación, prejuicios. Debe tomarse en su doble sentido de variable de stock (el conocimiento acumulado en un momento dado) o de flujo (aumento o disminución, en un lapso dado, del conocimiento acumulado al comienzo de dicho lapso).

(4) Una estructura $< N , Z , P (N_j) , P (Z_i) >$ (donde las probabilidades cumplen con los requisitos del cálculo respectivo) es una *estructura de información.*

$N : \{N_j\}$: Variable del universo que puede asumir los estados N_j para los $j = 1, 2 \ldots m$, los que constituyen una partición (exhaustivos y mutuamente excluyentes), asociados cada uno de ellos a una probabilidad de acontecer $P (N_j) = p (j)$

$Z : \{Z_i\}$: Variable del universo que puede asumir los estados Z_i para los $i = 1, 2 \ldots k \ldots n$, los que constituyen una partición (exhaustivos y mutuamente excluyentes), asociados cada uno de ellos a una probabilidad de acontecer $p (i)$

$P (N_j \cap Z_i)$: Probabilidad del acontecimiento conjunto de N_j y $Z_i = p (i j)$

(5) Una estructura de información $< N, Z, P (N_j), P (Z_i / N_j) >$ es una ***estructura bayesiana*** donde se atribuye a los elementos que la constituyen los siguientes significados:

N:	Conjuntos de estados inciertos del universo
Z:	Conjunto de mensajes que pueden obtenerse del universo
$P (Z_i / N_j)$:	$p (i j)$ probabilidad condicional que el universo emita el mensaje Z_i si se encuentra en el estado N_j.

La atribución de significados a N y a Z es arbitraria y se efectúa solamente a los fines de facilitar la exposición. En realidad ambas variables pueden jugar ambos roles y éstos tampoco son necesarios a los fines de modelo formal sugerido.

(6) La incertidumbre asociada a una estructura de información y a sus variables se mide con la ***entropía*** (SHANNON)

La incertidumbre asociada a una variable N se mide por la entropía (marginal):

$$H(N) = \sum_j p(j) . \log\ p(j)$$

donde

$H(N)$ = Entropía de la variable N

$p(j)$ = Probabilidad del estado N_j, es decir $P(N_j)$

log = Logaritmo en base 2

(7) La incertidumbre asociada al suceso conjunto de los distintos estados de las variables N y Z (de una estructura de información) se mide por la entropía conjunta:

$$H(N,Z) = \sum_i^n \sum_j^m p(ij) . \log\ p(ij)$$

(8) La incertidumbre asociada a la variable N luego del acontecimiento particular de Z_k se mide por la entropía condicional específica (a posteriori).

$$H(N/Z_k) = \sum_j^m p(j/k) . \log\ p(j/k)$$

(9) La incertidumbre promedio asociada a las distintas distribuciones de los estados N_j luego del acontecimiento de todos los estados Z_i se mide por la entropía condicional promedio (a posteriori).

$$H(N/Z) = \sum_i^n p(i) . H(N/Z_i)$$

$$= \sum_i^n p(i) \sum_j^m p(j/i) . \log\ p(j/i)$$

$$= \sum_i^n \sum_j^m p(ij) . \log\ p(j/i) \qquad \text{(por A.3)}$$

Se trata de la entropía (8) ponderada por las $P(Z_i)$ para las distintas Z_i. SHANNON denomina a esa medida con la palabra inglesa "equivocation". Para la teoría de la Información, mide la distorsión del mensaje enviado como Z_i y recibido como N_j.

Del mismo modo puede definirse la entropía $H(Z/N)$:

$$H(Z/N) = \sum_{i}^{n} \sum_{j}^{m} p(ij) . \log p(i/j)$$

Las definiciones de las distintas entropías están tomadas de SHANNON y adaptadas a la simbología de las estructuras bayesianas utilizada aquí.

3. Problemas en el uso de estructuras bayesianas

Entre otros, el análisis de estructuras bayesianas presenta tres problemas de relevancia.

El primero se refiere a que, en la práctica, aparece la incertidumbre promedio de todas las probabilidades a posteriori, originadas por la influencia de todos los mensajes pertinentes Z_i, es menor (o a lo sumo igual) que la incertidumbre original. En otras palabras, el análisis bayesiano siempre disminuye, en promedio, la incertidumbre previa (o a lo sumo, la mantiene idéntica). Ello choca el sentido común. ¿Existe alguna regla general acerca de la reducción de la incertidumbre por una estructura bayesiana en particular (o una estructura de información en general) ?

El segundo se refiere a que una estructura bayesiana es indiferente a un cambio relevante de información cuando el estado inicial es de certeza (subjetiva). En efecto, aplicando la regla general anterior, la incertidumbre inicial sería nula y por consiguiente la incertidumbre a posteriori de la información adicional sólo puede ser nula, cualquiera sea la información recibida. ¿Es ello una debilidad del análisis bayesiano?

El tercero se refiere a que es común que cierta información adicional aumente la incertidumbre en lugar de reducirla. ¿ Cómo se concilia este hecho con la regla general enunciada en primer término? .

Es importante destacar algunos aspectos de esta problemática. La inevitable disminución de la incertidumbre a través del análisis bayesiano no ha sido tratada por la literatura clásica sobre dicho análisis (CHERNOFF y MOSES , RAIFFA y SCHLAIFFER, SCHLAIFFER, MORGAN, HADLEY, entre otros) no obstante su importancia técnica y práctica.

Ello se debe quizá a que dicha literatura se ha concentrado sobre el valor de la información adicional, sin percatarse que no existe *valor negativo* de la información. Seguramente, ha influido un deficiente y a veces inexistente manejo de la *medida* de la incertidumbre. Uno pocos autores la mencionan como MORGAN, HADLEY, quienes

utilizan para ello la varianza. En realidad, no tratan la incertidumbre directamente sino la información como su contrario, midiéndola por la recíproca de la varianza. Esa medida es imperfecta, casi diríamos rudimentaria (PAVESI, 1986, (B)).

La superioridad de la entropía como medida de la incertidumbre, tanto por su eficacia como el aparato teórico que la respalda, ha sido desconocida. Los teóricos bayesianos surgidos especialmente del campo de la estadística, han ignorado la teoría de la Información de SHANNON. Evidentemente, esta última tiene otro destino pero este trabajo parte del principio que una estructura bayesiana es una estructura de información, como la utilizada (como matriz de probabilidades conjuntas) por SHANNON. Este enfoque de la decisión bayesiana desde el punto de vista de la Teoría de la Información no sólo es legítimo sino que se revela como sumamente fértil.

4. Un teorema fundamental: una estructura bayesiana nunca aumenta la incertidumbre

En este trabajo se entiende por decisión bayesiana el proceso secuencial por el cual, dada una estructura bayesiana:

(1) Se evalúa si la información adicional es conveniente (en el sentido que su costo debe ser menor o a lo sumo igual al beneficio que aporta) y se decide si adquirirla o no.

(2) Adquirida la información adicional se toma la decisión principal en base a las posibilidades a posteriori luego analizar la mencionada información.

En la estructura bayesiana se determinan las probabilidades a priori $P(N_j)$ (por cualquier medio) y conocer las probabilidades condicionales $P(Z_i / N_j)$ de que siendo N_j el estado *verdadero* del universo, aparezca el mensaje Z_i. De allí se derivan probabilidades a posteriori. $P(N_j / Z_i)$, afectadas por la información adicional acarreada por Z_i

Debe recordarse que, de todos modos, el análisis es *a priori* de la toma de información, digamos en el momento *t*. En ese mismo momento *t* se establecen las probabilidades *a priori* y las posibilidades *a posteriori* de la toma de información; por consiguiente esas probabilidades a posteriori son calculadas a priori, de allí que se usa llamar "preposterior" al análisis donde se establecen las probabilidades a posteriori.

Una vez tomada la información, la evaluación de la misma no tiene sentido para la toma de decisión. Sólo cabe calcular las nuevas probabilidades (a posteriori) para la información específica y real obtenida.

En este trabajo no tratamos el problema del valor y del precio de la información adicional. Nos interesa solamente saber si la incertidumbre disminuye o no. Podemos imaginar una situación en la cual, por la especial conformación de la matriz de decisión original, disminuya la incertidumbre promedio sin que se modifique el valor esperado

de la decisión a tomarse. Una reducción de la incertidumbre no implica obligatoriamente que la información adicional que la provocó tenga valor.

La incertidumbre y su variación es el único objeto de este trabajo. La misma se mide con la entropía, utilizada por SHANNON, fundador de la teoría de la información tanto para la incertidumbre como para la información. (Mayor desarrollo del tema de la entropía en PAVESI, 1986, (a) y 1986 (b).

Los autores de la llamada decisión estadística llegan a una conclusión que no debe confundirse con la nuestra: a partir de una situación de incertidumbre absoluta resuelta por los criterios de LAPLACE o del Minimax de WALD, establecen que cualquier criterio basado sobre una asignación subjetiva de probabilidades (criterio denominado "bayesiano", en un sentido muy diferente al nuestro) brinda un mejor valor esperado. Este resultado, que aparece como natural y conforme al sentido común, es cuasi trivial y nada tiene que ver con la regla general que estableceremos. Esta, en efecto, nos dice que cualquiera sea la distribución a priori, su entropía es mayor o igual a la entropía promedio de las distribuciones a posteriori.

La demostración del Teorema fundamental está basado sobre la argumentación de SHANNON (pág. 21 y siguientes) . Pero SHANNON no demostró el Teorema, por lo menos en su trabajo fundamental. Recurrió perversamente a la clásica fórmula "se demuestra fácilmente que ... ". ¡Verá el lector que la demostración no es tan trivial !

Teorema

Dada una estructura de información $< N_j , Z_i , P (N_j \cap Z_i) >$, entonces:

$$H(N/Z) \leq H(N)$$

La incertidumbre promedio de las distribuciones aleatorias a posteriori de los estados N_j de una variable N, si se producen todos los estados aleatorios Z_i , es menor o igual a la incertidumbre existente antes de la recepción de la mencionada información.

En otras palabras, en una estructura de información o bayesiana, toda la información adicional (aportada por la Z_i) reduce (o a lo sumo mantiene) en promedio la incertidumbre existente.

(1) $H(N/Z) = - \sum_i \sum_j p(ij) . \log p(j/i)$ por definición (9)

(2) $H(N) = - \sum_j p(j) . \log p(j)$ por definición (6)

(3) $p(j) = - \sum_i p(ij)$ por definición (A.6)

Entonces, reemplazando p (j) :

(4) $$H(N) = -\sum_i \sum_j p(ij) . \log\ p(j)$$

Queremos demostrar que:

$$H(N/Z) \leq H(N)$$

es decir que:

(5) $$H(N/Z) - H(N) \leq 0$$

Reemplazando el primer miembro por (1) y (4)

(6) $$-\sum_I \sum_j p(ij) . \log\ p(j/i) + \sum_i \sum_j p(ij) . \log\ p(j) \leq 0$$

El primer miembro se transforma:

(7) $$\sum_i \sum_j p(ij) . \{ \log\ p(j) - \log\ p(ij) \} \leq 0$$

(8) $$\sum_i \sum_j p(ij) . \log \{ p(j)/p(ij) \} \leq 0$$

Utilizaremos una característica de la funciones logarítmicas, desarrollada en el Apéndice B

(9) $$\log x \leq (x - 1) . \log e$$

Por consiguiente si aplicamos (9) a (8) y comparamos, tendremos:

(10) $$\sum_i \sum_j p(ij) . \log \{ p(j)/p(i/j) \} \leq \sum_i \sum_j p(ij) . \{ p(j)/p(i/j) - 1 \} . \log e$$

Es fácil demostrar que el miembro derecho de la desigualdad es igual a cero. En efecto, recordando que:

$$p(ij) = p(i)/p(j/i) \quad (A.3)$$

tendremos que dicho miembro es igual a:

$$(11)\quad \sum_i \sum_j \left\{ p(i)/p(j/i) \cdot p(j)/p(i/j) - p(ij) \right\} \cdot \log e =$$

$$= \sum_i \sum_j \left\{ p(i) \cdot p(j) - p(ij) \right\} \cdot \log e =$$

$$= \sum_i \left\{ p(i) \cdot \sum_j p(j) - \sum_j p(ij) \right\} \cdot \log e =$$

$$(12)\quad = \sum_i \left\{ p(i) - p(i) \right\} \cdot \log e = 0 \quad \text{(Por A.6)}$$

Reemplazando en (10) y recordando que el miembro izquierdo de esa inigualdad es igual a (5)

$$(13)\quad H(N/Z) - H(N) \leq 0$$

$$(14)\quad \text{y} \quad H(N/Z) \leq H(N)$$

Quedando así demostrado nuestro Teorema. En los ejemplos del Apéndice C se verá claramente que significa este resultado.

Cuando la incertidumbre promedio a posteriori es igual a la incertidumbre a priori, los mensajes del universo no aportan ninguna información o, más precisamente, los mensajes no afectan a la incertidumbre. ¿Cuándo sucede eso ?

Ya hemos demostrados que en (10) el miembro de la derecha es igual a cero. Vemos cuando el miembro de la izquierda también es igual a cero.

Para ello la expresión

$\log \left\{ p(j)/p(j/i) \right\}$ debe ser igual a cero

y por consiguiente debe darse que

$$p(j)/p(j/i) = 1$$

o sea

$$(15)\quad p(j) = p(j/i)$$

La igualdad (15) sólo sucede cuando $p(j)$ y $p(i)$ son independientes y mutuamente excluyentes (A.4). En ese caso, el mismo sentido común indica que no puede existir

información ya que mensajes y universos son variables que no se influyen mutuamente, no acarreando por lo tanto los mensajes información alguna.

En consecuencia:

$H(N/Z) = H(N)$ cuando $p(j/i) = p(j)$ (o cuando $p(ij) = p(i) \cdot p(j)$ por (A.4)

Un caso particular relevante de independencia sucede cuando la situación inicial es de certeza por lo cual la entropía a priori mínima es igual a cero. Por la aplicación del teorema, la entropia a posteriori también debe ser igual a cero. Este tema se analiza en el punto siguiente.

El teorema demostrado anteriormente es importante porque implica que una estructura de información (y desde luego una estructura bayesiana) reduce siempre, en promedio, la incertidumbre salvo que sea trivial (es decir que los mensajes sean independientes del universo no pudiendo así acarrear información alguna, caso en que la incertidumbre se mantiene).

El resultado parece contradecir la experiencia de información que aumenta la incertidumbre. Trataremos este importante tema en el punto 6. Aquí sólo recordaremos que la entropía a posteriori es solamente un promedio y que si un mensaje Z_k determinado aumenta la incertidumbre, debe existir otro mensaje (potencial) Z_h que la disminuya más.

Es importante insistir sobre una diferenciación sutil. La entropía promedio a posteriori $H(N/Z)$ es un concepto virtual, potencial. Sólo tiene sentido práctico si se lo calcula a priori, antes de tomar la información adicional. Una vez tomada esa información, la entropía real habrá aumentado o disminuido de acuerdo con el mensaje recibido.

El resultado es importante porque es general. En efecto, sólo requiere que las probabilidades de los mensajes Z_i y de los estados del universo N_j conformen una matriz de probabilidades conjuntas en el sentido expuesto en el Apéndice A. Esa matriz - suficiente para generar una estructura bayesiana - es prácticamente de aplicación universal, lo que da al teorema un alcance de gran relevancia. Ese alcance puede ser tal de constituir la base para una nueva flecha del tiempo, como veremos en el punto final.

5. La certeza a priori

El análisis bayesiano es relevante particularmente en situaciones de incertidumbre subjetiva. Dentro de esas situaciones , en las cuales la propensión a suceder de los distintos estados de un universo son grados de creencia, vale la pena detenerse en un caso extremo: la certeza subjetiva.

Si una variable N puede adoptar los estados N_1 , N_2 ... N_j ... N_m y el observador esta convencido, por cualquier razón, que el estado que ha de suceder será el N_k , asignará $P(N_k) = 1$ y $P(N_j) = 0$ para $j \neq k$. Es evidente, en este caso, que una estructura bayesiana no aportará ningún cambio y que no hay información que pueda modificar ese estado inicial de certeza subjetiva. Al observarse la fórmula de Bayes (Apéndice A) se ve de inmediato que un $P(N_j) = 0$ o un $P(N_j) = 1$ originarán indefectiblemente $P(N_j / Z_i) = P(Z_i)$ para cualquier Z_i . En situaciones de certeza a priori, las entropías a priori y a posteriori son iguales ambas en cero.

Esta situación puede ser fácilmente admitida en caso de certeza objetiva (¡si es que existe tal cosa !). En efecto, la estructura de información subyacente a la estructura bayesiana no admite otra posibilidad.

Pero no lo es tanto en caso de certeza subjetiva: existe tendencia a resistirse a la absoluta falta de sensibilidad de una estructura bayesiana a información adicional relevante en una situación de ese tipo.

Sin embargo, el problema parece simple. Si el observador se rehusa a admitir información relevante contraria a su creencia original, o desconfía de la veracidad de esa información con fundamentos o es irracional (BOULDING justamente define la racionalidad por la capacidad de modificar una creencia frente a información adicional relevante).

En estos casos, el análisis bayesiano no hace más que confirmar la posición del observador: nadie puede enjuiciar su terquedad y la fórmula de Bayes se somete legítimamente a ella. Su imposibilidad de adaptarse a la capacidad de recepción de nueva información del observador es calificada por LINDLEY (pag. 128) como *absurda* y disentimos totalmente con esa calificación. Si el observador esta dispuesto a aceptar la información que la va a modificar su certeza inicial, no puede exigir que el análisis bayesiano lo acompañe sin modificar las condiciones iniciales del sistema: la asignación de $P(N_k) = 1$ ya no representa lo que el observador realmente cree.

En efecto, si el observador está dispuesto, no ya a aceptar información adicional, sino simplemente a considerar la posibilidad de buscarla, entonces su creencia no es tan firme y el asignar una probabilidad nula a todos los estados del universo salvo uno, no representa fielmente su punto de vista.

El solo hecho de considerar la posibilidad de comprar información adicional (¡y con más razón el compraría !) demuestra que su asignación de $P(N_k) = 1$ ya no puede mantenerse. Por ínfima que sea, conscientemente o no, asigna un "coeficiente de duda" a los otros estados. Simbolicemos ε por ese coeficiente y hagámoslo tan pequeño como se quiera. La probabilidad asignada al estado que se creía cierto será ahora $P(N_k) = (1 - \varepsilon)$.

Definiremos:

$$0 \leq \varepsilon \leq 1$$

$$\varepsilon: \{\varepsilon_j\} \text{ para todo } j \neq k$$

Como puede verse en el ejemplo del Apéndice C, la introducción de ese *coeficiente de duda* es suficiente para que la estructura bayesiana se haga receptiva a la información adicional. Nótese que no se trata de un simple truco computacional: se trata de expresar, de alguna forma, la creencia del observador.

LINDLEY llega a una conclusión que creemos valida a medias: si alguien le dice que la luna es de queso verde, dele algo de probabilidad, aun cuando sea muy pequeña, porque si no un ejército de astronautas volviendo de la luna con muestras de ese queso no lograrían hacerle cambiar de opinión.
No crea en nada de forma absoluta, deje un margen de duda. Esta arenga es válida como consejo práctico. Desde el punto de vista de los fundamentos de la Teoría de la Decisión, la negativa en creer que la luna está hecha de queso verde es legítima.

Por otra parte, los casos de certeza subjetiva inicial se originan en estructuras de información no relevantes (independencia de las variables). Este problema de la posición subjetiva, en el caso diametralmente opuesto, se presenta también en las situaciones de incertidumbre total(véase las supuestas paradojas en LUCE y RAIFFA, pag. 278 y ss.).

6. La información que aumenta la incertidumbre

Un observador piensa que Juan y / o Pedro pueden encontrarse en una habitación cerrada. Existen cuatro estados N_j posibles: se encuentran en la habitación Juan y Pedro o ambos o ninguno.

El observador cree razonablemente que éstos son los únicos cuatro estados posibles. Esto significa simplemente que ha asignado una propensión a suceder nula, una probabilidad cero a cualquier otro estado (originado, por ejemplo, en el hecho que puede haber una persona más incluida en la situación).

Supongamos justamente que existe otro individuo: Andrés y que alguién se lo informa al observador y que éste acepta como verídica la información. Existen ahora ocho estados posible de la variable N. (Se agregan las posibilidades de encontrarse en la habitación, además de los casos anteriores: Andrés; Juan y Andrés; Pedro y Andrés; Juan, Pedro y Andrés).

Indudablemente la incertidumbre del observador ha aumentado a raíz de la información (si los estados son equiprobables, pasa de una entropía de 2 bits a otra de 3 bits). ¿Ello contradice la regla general de Teorema expuesto anteriormente? Evidentemente no.

En primer lugar, observemos que el modificar el número de estados posibles de una variable incierta se reduce modificar la distribución de su propensión a suceder: se aumentan los estados asignando una probabilidad mayor que cero a los que tenían previamente asignados una probabilidad igual a cero y viceversa. Nuestras definiciones de estructuras de información y bayesianas no prohiben las probabilidades nulas.

En segundo lugar, el ejemplo anterior sólo describe una parte de la realidad al proponer una información determinada (hay 8 estados posible en lugar de 4). Siempre existe otra información aún cuando más no fuera la *no información*. A priori no se puede eliminar la posibilidad de que alguien le confirme al observador su creencia. O que alguien le informe que son 16 los estados posibles o ninguno! Y todas estas informaciones diversas tienen una propensión a ser emitidas, una probabilidad asignada, aún cuando sea desconocida por el observador. Ya lo estableció SHANNON: para que haya información deben existir por lo menos dos mensajes. Siempre debe existir, asociado a un mensaje determinado, el mensaje contrario: sino, no hay información. En efecto, con un solo estado Z_k , un solo mensaje las p (i) y las p (j) son independientes.

Por lo tanto , siempre existe a priori una estructura de información y por consiguiente siempre es válido el Teorema por el cual la entropía promedio a posteriori es menor o igual a la entropía a priori.

La entropía promedio a posteriori pondera las distintas entropías correspondientes a cada mensaje (definición 9). Algunas de esas entropías a posteriori pueden ser mayores que la entropía a priori (como en el caso del ejemplo), otras o todas pueden ser inferiores. Pero en promedio siempre la entropía final (como ente virtual, potencial, calculado ex-ante de la toma de información) es menor o igual a la entropía inicial, si bien la entropía real a posteriori correspondiente a un mensaje determinado puede ser superior.

Supongamos que a priori, haya solo un mensaje posible. Entonces, deja de ser posible para transformarse en necesario y por consiguiente deja de ser un mensaje: se trata de un estado único y cierto del universo, conocido por el observados por definición subjetiva de la incertidumbre. Si el observador lo repudia y se mantiene en una creencia contraria a la evidencia entonces esa evidencia no existe: lo que existe y es válido es la creencia del observador en su universo percibido. Esta *paradoja de la terquedad* es una característica inevitable del enfoque subjetivo de las cosas del conocimiento. Si bien puede molestar al estudioso, es absolutamente cohorente y no contradice ningún principio lógico (y por lo tanto, no es estrictamente una paradoja). El análisis bayesiano y, en general, el enfoque subjetivo de la información se someten a ese hecho indiscutible.

Por lo tanto, es legítimo decir que pueden existir mensajes específicos acarreando información que aumenta la incertidumbre. Pero todos los mensajes específicos posibles -y en general siempre hay por lo menos dos posibles- en promedio disminuyen la incertidumbre, salvo que no sean verdaderos mensajes (es decir, salvo que su o-

currencia sea independiente de los estados de la variable incierta que interesa al observador).

Ni es cierto que toda información reduce la incertidumbre ni es cierto que alguna estructura de información pueda prometer, a priori, la certeza de aumentar, en promedio, la incertidumbre.

7. Información y restricción: las estructuras de información

Toda situación real puede modelizarse con una *estructura de información.*

Esta será *irrelevante* si las variables N y Z tienen un comportamiento *independiente* , es decir si $p(ij) = p(i) \cdot p(j)$. En este caso, no hay información y se verifica que:

$$H(N/Z) = H(N)$$

$$H(N \cap Z) = H(N) + H(Z)$$

(la segunda expresión también fue planteada por SHANNON y puede demostrarse en forma similar a la utilizada para la primera).

La estructura de información será *relevante* cuando las variables N y Z son *interdependientes* , es decir que $p(ij) \neq p(i) \cdot p(j)$. En este caso, hay información y se verifica que:

$$H(N/Z) < H(N)$$

$$H(N \cap Z) < H(N) + H(Z)$$

Por consiguiente, para un determinado observador en un momento dado, una variable Z acarrea información acerca de otra variable N cuando $P(N \cap Z) \neq P(N) \cdot P(Z)$, de acuerdo con la percepción del observador.

Puede demostrarse que la entropía conjunta de una estructura de información relevante es menor que la entropía conjunta de una estructura de información no relevante, es decir:

Si $$H(N \cap Z) = \sum_i \sum_j p(ij) \cdot \log p(ij)$$

y $$H'(N \cap Z) = \sum_i \sum_j p(i) \cdot p(j) \cdot \log \{ (pi) \cdot (pj) \}$$

donde $p(ij) \neq p(i) \cdot p(j)$

entonces:

$$H(N \cap Z) < H'(N \cap Z)$$

La no existencia de independencia entre N y Z indica la existencia de restricciones sobre su comportamiento mutuo. De modo que también se puede definir una *restricción mutua* entre N y Z cuando $P(N \cap Z) \neq P(N) \cdot P(Z)$. (Evidentemente, no se sabe cual restringe a cual).

Por consiguiente, *restricción* e *información* son conceptos formalmente idénticos lo que explica que no hay información sin restricción y viceversa (PAVESI, 1986 (b)).

Una restricción introduce una deformación, un pliegue en la estructura.

La expresión:

$$H(N \cap Z) - H'(N \cap Z)$$

revela el grado de *deformación* de una estructura relevante dada con respecto a una estructura no relevante con iguales probabilidades marginales. Se puede demostrar que la medida de esa deformación es también la medida de la información aportada por la estructura. Finalmente:

$$H(N \cap Z) - H'(N \cap Z) = H(N/Z) - H(N) = I(N/Z)$$

Esta conclusión es muy interesante si bien no debería sorprendernos. Es justamente la estructura de información que da lugar a la estructura bayesiana y ésta sólo puede aportar la información incierta en su estructura.

Puede compararse el eficiente uso de la entropía H con los coeficientes de contingencia de Pearson y otras medidas de asociación de atributos. (Véase, por ejemplo, YULE y KENDALL, capítulos 2 y 3)

8. La flecha del tiempo

La física clásica, típicamente determinista, no tiene en cuenta el tiempo. Se llega a deducir la definición del tiempo de la segunda Ley de Termodinámica: el tiempo sigue el sentido de entropía. Si el mismo universo presenta en dos momentos diferentes (pero sin orden definido) dos niveles de entropía, el tiempo ha transcurrido del estado de menor entropía al estado de mayor entropía. Este sentido del tiempo ha sido bautizado por EDDINGTON, la *flecha del tiempo*.

La universalidad de la segunda ley ha sido duramente discutida pero ha sobrellevado airosamente los ataques: hoy nadie la repudia si bien se ha demostrado la existencia de fenómenos que, en el mismo sentido que la evolución biológica y que la organización, combaten localmente con éxito la tendencia a la entropía creciente (estructuras disipativas de PRIGOGINE, autopoiesis de VARELA Y MATURANA, autocatálisis, hiperciclos).

Se trata de establecer así otras variables que, además de la entropía, puedan constituir un confiable indicador de la flecha del tiempo. El tema es trascendente ya que la mayor parte de esas variables se refieren al incremento de magnitudes cuyo comportamiento es sustancialmente asimétrico y que llevan en forma irreversible al orden o al desorden. La segunda ley asegura un inexorable caos creciente. La leyes biológicas prometen un incremento permanente en la complejidad, de la diversidad, de la especificidad, es decir del orden. La expansión cósmica define un universo en constante expansión con galaxias huidizas que extienden sin fin el espacio.

Surge entonces la idea que la información, el conocimiento, constituye también una variable en expansión, de comportamiento asimétrico e irreversible (LAYER, BRILLOUIN). La información acumulada en el universo va creciendo siempre, en forma inexorable: ello implica automáticamente la existencia de mayores restricciones, de mayor orden, de mayor organización.

La aparente contradicción entre las leyes biológicas -que siguen un sentido parecido al supuesto de la información- ha sido eliminada por el trabajo pionero de SCHRÖDINGER y otros. Pero ¿cómo podría reconciliarse un universo en desorden físico creciente simultáneamente con un universo en orden de información y organización también creciente?

La hipótesis de una ley de información creciente es tratada despectivamente por algunos autores como una metáfora. BUNGE (pag. 85) sostiene que la incertidumbre crece a la par de la información (lo que no contradice la supuesta Ley de información creciente).

Creemos que la regla general por la cual la incertidumbre promedio a posteriori es siempre menor (o a lo sumo igual) que la incertidumbre a priori sostiene la mencionada hipótesis. Las estructuras de información en la forma que la hemos definido son sumamente generales y cualquier par de variables del universo constituye una (las manchas solares y las cosechas, la luna llena y las mareas, los cortes generales de energía eléctrica y los nacimientos, los planes de inversión y el producto bruto, la trayectoria de los astros y la felicidad, la crítica de cine y la concurrencia a las salas, el número de pétalos de la margarita y el amor del ser querido, etc.)

Basta que su comportamiento no sea independiente para que en promedio, la incertidumbre se reduzca y el conocimiento aumente. En el peor de los casos (independencia de los comportamientos) nada cambia.

La universalidad de las estructuras de información y su intenso uso que miles de millones de individuos garantizan - a través de la ley de los grandes números - que todos los mensajes posibles se produzcan cumpliéndose al promedio de la entropía y por consiguiente el aumento permanente de la información.

El mecanismo de incremento del orden no es fluido ni continuo ni predecible en cada situación. Pero la vastedad del campo de la aplicación es tal que podemos imaginar que desde un punto de vista macroscópico y a través de los años, el incremento permanente de la información puede ser confirmado como ley estadística.

APÉNDICE A

RECORDATORIO DE PROBABILIDADES - TEOREMA DE BAYES

Se trata de un breve resumen al solo fin de facilitar la lectura del texto principal.

1. Variables y eventos

Trabajaremos con dos variables: N y Z.

N puede tomar m valores: N_j : N_1 , N_2 . . . N_j . . . N_m

Z puede tomar n valores: Z_i : Z_1 , Z_2 . . . Z_i . . . Z_n

Cada uno de esos valores es considerado un evento. Todos los valores de *una misma* variable son mutuamente exclusivos de modo que:

$N_i \cap N_j = \varnothing$ para todo $i \neq j$

$Z_i \cap Z_j = \varnothing$ para todo $i \neq j$

Ello quiere decir que cada variable asume uno y solo uno de sus valores.

Pero *entre las dos variable* , no se define la exclusión de modo que:

$Z_i \cap N_j = \varnothing$ ó $\neq \varnothing$

Cuando $Z_i \cap N_j = \varnothing$, la aparición de un valor determinado de una variable no influye sobre la aparición de un valor determinado de la otra variable. Ambas son independientes.

Cuando $Z_i \cap N_j \neq \varnothing$, el acontecimiento de uno influye sobre el acontecimiento de otro. Ambas son interdependientes.

2. Probabilidades

Los valores de una variable tienen cierta propensión a suceder que se mide por la probabilidad. Simbolizaremos:

$P(N_j) = p(j)$ = probabilidad del evento N_j

$P(Z_i) = p(i)$ = probabilidad del evento Z_i

La probabilidad cumple con dos restricciones:

(1) $p(x_i) \geq 0$

(2) $\sum_i p(x_i) = 1$

Nos interesan tres clases de probabilidades:

$P(N_i \cup N_j)$ = probabilidad que N_i ó N_j sucedan.

$P(Z_i \cap N_j) = p(ij)$ = probabilidad que Z_i ó N_j sucedan conjuntamente o probabilidad *conjunta.*

$P(Z_i / N_j) = p(i/j)$ = probabilidad que suceda Z_i si N_j ya sucedió o probabilidad *condicional.*

3. Relaciones fundamentales

(1) $P(Z_i \cup N_j) = P(Z_i) + P(N_j) - P(Z_i \cap N_j)$

(2) Cuando $Z_i \cap N_j = \varnothing$, es decir cuando son independientes

$$P(Z_i \cup N_j) = P(Z_i) + P(N_j)$$

(3) $P(Z_i \cap N_j) = P(Z_i) \cdot P(N_j / Z_i) = P(N_j) \cdot P(Z_i / P_j)$

Utilizando la simbología abreviada:

$p(ij) = p(i) \cdot p(j/i) = p(j) \cdot p(i/j)$

(4) Cuando $Z_i \cap N_j = \varnothing$, entonces ninguno de ellos depende del otro. Por lo tanto:

$$P(N_j / Z_i) = P(N_j); \quad P(Z_i / N_j) = P(Z_i)$$

y $$P(Z_i \cap N_j) = P(N_j) \cdot P(Z_i) = P(Z_i) \cdot P(N_j)$$

o bien $$p(ij) = p(i) \cdot p(j) = p(j) \cdot p(i)$$

También puede expresarse la independencia diciendo:

$$\frac{P(N_j / Z_i)}{P(N_j)} = 1$$

Z_i no aporta ninguna modificación en el acontecer de N_j, no influye en el suceso.

(5) Por consiguiente, se deduce de (3):

$$p(i/j) = p(ij)/\ p(j)$$

$$p(j/i) = p(ij)/\ p(i)$$

Las relaciones siguientes se apreciarán mejor a través del ejemplo del punto siguiente.

(6) $\sum_j p(ij) = p(i)$ $\qquad\qquad$ $\sum_i p(ij) = p(j)$

por consiguiente, por (3):

$$\sum_j p(i)\ .\ p(j/i) = p(i).\sum_j p(j/i) = p(i) \text{ donde } \sum_j p(j/i) = 1$$

$$\sum_i p(j)\ .\ p(i/j) = p(j).\sum_i p(i/j) = p(j) \text{ donde } \sum_i p(i/j) = 1$$

(7) Por (5):

$$\sum_j p(j/i) = 1\ ; \qquad \sum_i p(i/j) = 1$$

(8) **Teorema de Bayes**

$$P(N_j / Z_i) = \frac{P(N_j).P(Z_i / N_j)}{\sum_j P(N_j).P\ (Z_i / N_j)}$$

Por (5):

$$P(N_j / Z_i) = \frac{P(N_j \cap Z_i)}{P(Z_i)} = \frac{P(N_j).P(Z_i / N_j)}{P(Z_i)}$$

Por (6):

$$P(N_j / Z_i) = \frac{P(N_j).P(Z_i / N_j)}{\sum_j P(N_j).P\ (Z_i / N_j)}$$

Se le da, en este trabajo, la siguiente interpretación al "Teorema" (en realidad una fórmula) de Bayes:

N_j : Estados del universo

Z_i : Mensajes emitidos por -u obtenidos del- universo

$P(N_j)$: Probabilidad a priori : estimación en el momento t de la propensión a suceder del estado N_j . Conjuntamente con $P(Z_i)$ constituye una probabilidad marginal (por aparecer en el "margen" de un cuadro del tipo utilizado en el ejemplo del punto siguiente)

$P(N_j / Z_i)$: Probabilidades a posteriori : nueva estimación en el momento t + d de la propensión a suceder del estado N_j , luego de considerar la influencia de la información acarreada por el mensaje Z_i .

$P(Z_i / N_j)$: Verosimilitud , probabilidad de que siendo el estado del universo el N_j , se detecte el mensaje Z_i .

4. Ejemplo

Adoptaremos un ejemplo clásico de la urna. Una urna tiene 100 bolillas: 50 blancas, 30 verdes y 20 rojas. A su vez 60 de estas bolillas están numeradas y 40 no lo están. La situación se resume en el cuadro siguiente:

		Número		
		Sí	No	Total
Color	Blancas	25	25	50
	Verdes	20	10	30
	Rojas	15	5	20
	Total	60	40	100

El cuadro indica, por ejemplo, que de las 20 bolillas rojas, 12 están numeradas y 8 no lo están.

Simbolizaremos la situación del siguiente modo: la variable *Numeración de la bolilla* tiene dos estados: N_1 ; bolilla numerada; N_2 : bolilla no numerada.

La variable *Color de bolilla* tiene tres estados. Z_1 : Blanca; Z_2 : Roja; Z_3 : Verde.

Las probabilidades conjuntas $P(N_j \cap Z_i)$ (es decir, por ejemplo, la probabilidad que una bolilla roja sea no numerada) están dadas por la siguiente tabla, donde todos lo datos de la anterior se han dividido por 100.

$p(ij)$

	N_1	N_2	
Z_1	0,25	0,25	0,50
Z_2	0,20	0,10	0,30
Z_3	0,15	0,05	0,20
	0,60	0,40	
	$P(N_1)$	$P(N_2)$	

La matriz de probabilidades conjuntas anterior constituye una estructura de información.

Las $P(N_j)$ y las $P(Z_i)$ son probabilidades marginales por su ubicación en el cuadro.

La relación (6) del punto 3 de este Apéndice se comprueba:

$$\sum_j p(ij) = p(i) \qquad \sum_i p(ij) = p(j)$$

$$\sum_j p(ij) = 0{,}25 + 0.25 = 0{,}50 = p\ (i=1)$$

$$\sum_i p\ (i2) = 0{,}25+0.10+0{,}05 = 0{,}40 = p\ (j=2)$$

Las probabilidades condicionales $P(Z_i / N_j)$ y $P(N_j / Z_i)$ se obtienen aplicando la relación (5) dividiendo, para las primeras, las probabilidades conjuntas de una columna por la probabilidad marginal $P(N_j)$ que es igual a la suma de esta columna y, para las segundas, las probabilidades conjuntas de una fila por la probabilidad marginal $P(Z_i)$ que es la suma de esa fila.

	N_1	N_2
Z_1	5 / 12	5 / 8
Z_2	4 / 12	2 / 8
Z_3	3 / 12	1 / 8
Σ	1	1

	N_1	N_2	Σ
Z_1	1 / 2	1 / 2	1
Z_2	2 / 3	1 / 3	1
Z_3	3 / 4	1 / 4	1

Se comprueban las relaciones (7)

$$\sum_i p(i/j) = 1 \qquad \sum_j p(i/j) = 1$$

$$\sum_i p(i/j) = 5/12 + 4/12 + 3/12 = 1 \qquad \sum_j p(2/j) = 2/3 + 1/3 = 1$$

¿ Cuál es la probabilidad de que una bolilla extraída que resulta ser verde no sea numerada ?

$$P(N_2 / Z_2) = 1/3$$

Si sólo poseemos como información la posibilidades a priori $P(N_j)$ y la matriz de verosimilitud $P(Z_i / N_j)$ la aplicación del Teorema de Bayes nos da las probabilidades a posteriori $P(N_j / Z_i)$.

Si extraemos una bolilla y resulta ser verde, las probabilidades a posteriori serán, evidentemente:

$$P(N_1 / Z_2) = 2/3 \qquad P(N_2 / Z_2) = 1/3$$

El Teorema de Bayes resulta banal con un ejemplo de urna. Comienza a tomar cierta trascendencia cuando las Z_i son mensajes que acarrean información acerca de las N_j. En nuestro ejemplo, *antes* de extraer la bolilla, la probabilidad de extraer una numerada era 0,60. Luego de haber extraído una verde, la probabilidad que sea numerada aumentó a 2 / 3 = 0,67.

Si las Z_i y las N_j fueran independientes, la urna presentaría la siguiente conformación:

	N_1	N_2	
Z_1	30	20	50
Z_2	18	12	30
Z_3	12	8	20
	60	40	100

Puede comprobarse fácilmente que $P(N_j) = P(N_j / Z_i)$ y que por consiguiente $P(N_j \cap Z_i) = P(N_j).P(Z_i)$.

APÉNDICE B

LA CONVEXIDAD DE LAS FUNCIONES LOGARÍTMICAS

Daremos solamente una interpretación geométrica de la convexidad.

1. Una función f(x) es convexa cuando:

$$a\,f(x_1) + (1 - a)\,f(x_2) \leq f\left[\, ax_1 + (1 - a)\,x_2 \,\right] \text{ para } 0 \leq a \leq 1$$

Esto significa que la cuerda conectando $f(x_1)$ y $f(x_2)$ se encuentra siempre por debajo de la curva f(x).

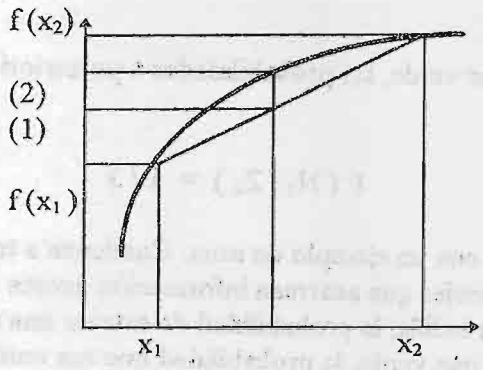


Figura 1

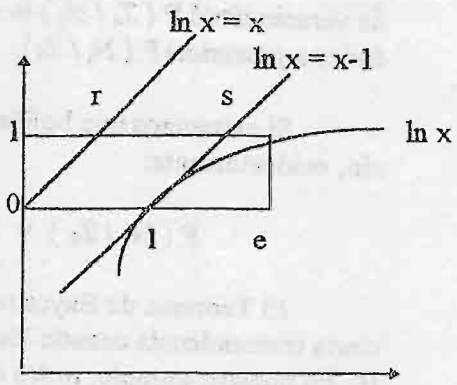


Figura 2

En la figura 1:

$$(1) = a\,f(x_1) + (1 - a)\,f(x_2) \qquad (2) = f\left[\, ax_1 + (1 - a)\,x_2 \,\right]$$

Se puede demostrar que una función logarítmica cumple ese requerimiento. Por ejemplo, para a = 1 / 2

$$1/2\ (\log x_1 + \log x_2) \leq \log\left[\, (x_1 + x_2)/2 \,\right]$$

2. En la figura 2 hemos tomado (arbitrariamente) el logaritmo neperiano de base e.

La recta r representa la hipotética igualdad ln x = x. La recta s, tangente a ln 1 representa la hipotética igualdad ln x = x - 1.

Es evidente que por la convexidad: ln x ≤ x - 1
Esto se aplica para cualquier base del logaritmo. En efecto:

$$\log x = \log_2 x \,/\, \log_2 e \leq (x - 1)$$
$$\log_2 x \leq (x - 1)\log_2 e$$

APÉNDICE C

EJEMPLOS

A. El Teorema

1. Tomaremos como base el ejemplo del Apéndice A.

Estructura de información

p (i j)	N_1	N_2	$P(Z_i)$
Z_1	0,25	0,25	0,50
Z_2	0,20	0,10	0,30
Z_3	0,15	0,05	0,20
$P(N_j)$	0,60	0,40	

Estructura bayesiana

Verosimilitud p (i j)	N_1	N_2	Probab. a posteriori p (j / i)	N_1
Z_1	5 / 12	5 / 8	Z_1	1 / 2
Z_2	4 / 12	2 / 8	Z_2	2 / 3
Z_3	3 / 12	1 / 8	Z_3	3 / 4
	1	1		

Entropía

a priori: $H(N) = 0,60 \log 0,60 + 0,40 . \log 0,40 = 0,9710$ bit

a posteriori :

$H(N / Z_1) = 0,50 \log 0,50 + 0,50 \log 0,50 = 1$ bit

$H(N / Z_2) = 2 / 3 . \log 2 / 3 + 1 / 3 . \log 1 / 3 = 0,9483$ bit

$H(N / Z_3) = 3 / 4 . \log 3 / 4 + 1 / 4 . \log 1 / 4 = 0,8113$ bit

$H(N / Z_4) = 0,50 . 1 + 0,30 . 0,9183 + 0,20. 0,8113 = 0,9378$ bit

Información promedio aportada:

$$H(N) - H(N/Z) = 0{,}9710 - 0{,}9378 = 0{,}0332 \text{ bit}$$

Obsérvese que:

- La entropía promedio a posteriori es menor que la entropía a priori. La diferencia es la medida de la información aportada.

- Para los tres mensajes posibles Z_i, el primero Z_1 aumenta la incertidumbre. La entropía pasa de 0,9710 bit a 1 bit. De una distribución de probabilidades 0,6 - 0,4 a 0,5-0,5. Para Z_2 se pasa de 0,6-0,4 a 0,67 - 0,33 y para Z_3 se pasa de 0,6 - 0,4 a 0,75 - 0,25. Es evidente que Z_3 reduce la incertidumbre más que Z_2.

2. Se aprecia claramente que $H(Z) = \sum_j p(ij) . \log p(ij) = \sum_i \sum_j p(ij) . \log \sum p(ij)$

$$H(Z) = 0{,}50 . \log 0{,}50 + 0{,}30 . \log 0{,}30 + 0{,}20 . \log 0{,}20 = 1{,}4855$$

$$= 0{,}25 . \log (0{,}25 + 0{,}25) + 0{,}25 . \log (0{,}25 + 0{,}25) + 0{,}20 . \log (0{,}20 + 0{,}10) + 0{,}10 . \log (0{,}20 + 0{,}10) + 0{,}15 . \log (0{,}15 + 0{,}05) + 0{,}05 . \log (0{,}15 + 0{,}05) = 1{,}4855$$

3. Siendo una estructura de información relevante, se da que :

$$H(N \cap Z) < H(N) + H(Z)$$

$$H(N \cap Z) = 0{,}25 . \log 0{,}25 + 0{,}25 . \log 0{,}25 + 0{,}20 . \log 0{,}20 + 0{,}10 . \log 0{,}10 + 0{,}15 . \log 0{,}15 + 0{,}05 . \log 0{,}05 = 2{,}4233$$

$$H(N) = 0{,}9710$$

$$H(Z) = 1{,}4855$$

$$H(N) + H(Z) = H'(N \cap Z) = 2{,}4565 > 2{,}4233 = H(N \cap Z)$$

$$H'(N \cap Z) - H(N \cap Z) = 2{,}4565 - 2{,}4233 = 0{,}0332$$

Evidentemente, la *distorsión* de la estructura de información revela el monto de información disponible en la misma y que hemos utilizado al usar Bayes en el punto 1 de este Apéndice.

4. Si la estructura fuese no relevante, es decir si el comportamiento de N y Z es independiente, la estructura de información (matriz de probabilidades conjuntas) sería, con la mismas $P(N_j)$ y $P(Z_i)$.

	N_1	N_2	$P(Z_i)$
Z_1	0,30	0,20	0,50
Z_2	0,18	0,12	0,30
Z_3	0,12	0,08	0,20
$P(N_j)$:	0,60	0,40	

La entropía total

$$H'(N \cap Z) = 0{,}30 \cdot \log 0{,}30 + 0{,}20 \cdot \log 0{,}20 + 0{,}18 \cdot \log 0{,}18 + 0{,}12 \cdot \log 0{,}12 + 0{,}12 \cdot \log 0{,}12 + 0{,}08 \cdot \log 0{,}08 = 2{,}4565$$

$$= H(N) + H(Z)$$

La entropía conjunta de la estructura no relevante, $H'(N \cap Z)$ puede obtenerse tanto por la suma de las entropías marginales como por la aplicación de la fórmula general de la entropía.

Evidentemente, la información que surge decisión de la estructuras relevantes y no relevantes con probabilidades marginales iguales tiene la misma cantidad que la que surge del análisis bayesiano.

B. La certeza a priori

1. Tomemos el ejemplo del punto 6. El observador DECISIÓN cree que sólo pueden encontrarse Juan o Pedro en (o ambos o ninguno) en la habitación cerrada. Llamamos N_1 al conjunto de estas cuatro posibilidades y N_2 a cualquier evento distinto.

El observador está convencido que N_1 es la verdadera situación. Por la tanto asignará $P(N_1) = 1$ y $P(N_2) = 0$.

Por la aplicación de Bayes, cualquier información en contrario, aún la más confiable e impactante, no modificará esa asignación.

Si D está dispuesto a escuchar información o si cualquier circunstancia lo hace dudar, deberá modificar su asignación de $P(N_j)$. Supongamos que fija $P(N_1) = 1 - \varepsilon$ y $P(N_2) = \varepsilon$.

Alguien en quien confía le asegura que puede informar fehacientemente si la

situación es N_1 o N_2. Ello implica la siguiente estructura bayesianas (verosimilitud):

	N_1	N_2
Z_1	1	0
Z_2	0	1

Si la información es Z_1, $P(N_1 / Z_1) = 1$ y $P(N_2 / Z_1) = 0$

Si la información es Z_2, $P(N_1 / Z_1) = 0$ y $P(N_2 / Z_1) = 1$

En ambos casos, la incertidumbre disminuye ya que

$$H(1,0) = H(0,1) < H[(1 - \varepsilon), \varepsilon]$$

El análisis bayesiano recoge la información contraria o la creencia inicial (Z_2) y la de toda su importancia ya que para una $P(N_2) = \varepsilon$ (muy pequeña) pasa a una $P(N_2 / Z_2) = 1$.

Lo mismo sucede con una verosimilitud de información imperfecta.

2. Es interesante analizar la cantidad de información necesaria para hacer cambiar totalmente de opinión al observador. Ello puede interpretarse así (MASSEY):

$$P(N_j / Z_k) = 1 - P(N_j)$$

Supongamos la siguiente distribución a priori :

$$P(N_1) = 0{,}9 \qquad P(N_2) = 0{,}1$$

Se podría suponer que para obtener las probabilidades a posteriori :

$$P(N_1 / Z_k) = 0{,}1 \qquad P(N_2 / Z_k) = 0{,}9$$

se necesita mucha información. Sin embargo no es así ya que en este caso;

$$H(N_j) = H(N_j / Z_k)$$

Es decir que la incertidumbre no cambia, no obstante el cambio radical de creencia. La entropía es indiferente al ordenamiento de los estados. Esta conclusión es paradójica pero el enfoque parcial efectuado aquí puede conducir a error. En efecto, es necesario ver que pasa, teniendo en cuenta los mensajes posibles.

3. Supongamos que $Z = Z_1 , Z_2$.

Para que los valores de las probabilidades de N_1 y N_2 se inviertan, tal como hemos visto anteriormente, una matriz de verosimilitud $P (Z_i / N_j)$ posible es la siguiente:

	N_1	N_2
Z_1	0,012	0,988
Z_2	0,988	0,012

Tendremos:

Con Z_1:

$$P (N_1 / Z_1) = 0,10 \qquad P (N_2 / Z_1) = 0,90$$

Con Z_2:

$$P (N_1 / Z_2) = 0,9986 \qquad P (N_2 / Z_1) = 0,0014$$

Se ve como Z_1 invierte la creencia del observador pero Z_2 la reafirma considerablemente.

$H (N_j) = 0,469$

$H (N_j / Z_1) = 0,469$

$H (N_j / Z_2) = 0,015$

$H (N_j / Z_i) = 0,065$

$H (N_j) - H (N_j / Z_i) = 0,469 - 0,065 = 0,404$

La incertidumbre promedio ha disminuido (con Z_1 se mantiene constante, con Z_2 disminuye) y la información aportada, $I (N_j / Z_i)$, es de 0,404 bits.
Se trata de una cantidad reducida si se tiene en cuenta que existe una probabilidad del 11% de revertir totalmente la opinión.

4. Para un estado N_h dado y un mensaje Z_k dado, puede definirse la información aportada por Z_k como:

$$I (N_h / Z_k) = \log P (N_h / Z_k) - \log P (N_h) = \log \{ P (N_h / Z_k) / P (N_j) \}$$

(Se mide así la incertidumbre vinculada a cada estado -y no ya a toda la distribución- por el log de su probabilidad.

Cuando la incertidumbre disminuye, la información es positiva y cuando crece es negativa.

Tendremos en este caso:

Si aparece Z_1:

$$I(N_1 / Z_1) = \log 0,1 - \log 0,9 = -3,17$$

$$I(N_2 / Z_1) = \log 0,9 - \log 0,1 = 3,17$$

Si aparece Z_2:

$$I(N_1 / Z_2) = \log 0,9986 - \log 0,9 = 0,15$$

$$I(N_2 / Z_2) = \log 0,0014 - \log 0,1 = -6,158$$

C. **La información que aumenta la incertidumbre**

1. Siguiendo el ejemplo del punto 6, el observador tiene la seguridad que sólo 2 personas están involucradas y está seguro que cualquier información que solicitará le confirmará su creencia. Entonces no hay nada que hacer: el análisis bayesiano no puede modificar la terquedad del observador. Es sólo un instrumento que obedece.

2. Supongamos que el observador está dispuesto a admitir cierta duda y que además está dispuesto a aceptar que alguna información puede indicarle que son 3 las personas involucradas (con 8 estados aleatorios):

N_1 : hay sólo dos personas (4 estados) $\quad P(N_1) = 0,9$

N_2 : hay tres personas (8 estados) $\quad P(N_2) = 0,1$
Z_1 : le informan que hay 2 personas $\quad P(Z_1) = 0,9$
Z_2 : le informan que hay 3 personas $\quad P(Z_2) = 0,1$

La *estructura de información* (probabilidades conjuntas) es la siguiente:

	N_1	N_2	$P(Z_i)$
Z_1	0,9	--	0,9
Z_2	--	0,1	0,1
$P(N_j)$:	0,9	0,1	1

La entropía de esta estructura H (N, Z) = 0,4690

Si N y Z fuesen independientes, la entropía H' (N, Z) = 0,938 (¡el doble !). Es decir que la información promedio que brinda por la estructura es de 0,4690 bits. Si le dan como información Z_2, menos de 1/2 bit es suficiente para hacerlo cambiar de opinión. En efecto:

$P(N_1 / Z_1) = 1$

$P(N_2 / Z_2) = 1$

La entropía inicial $H(N) = 0,469$

La entropía final $H(N / Z) = 0$

En este caso, la información adicional le reduce la incertidumbre, cualquiera sea el mensaje (de los dos posibles) recibido.

Lo importante, para nuestro argumento, que es imposible (o mejor dicho no tiene sentido) pensar en un solo mensaje.

2. Supongamos ahora que el observador asigna:

$P(N_1) = 0,98 \qquad P(N_2) = 0,02$

y que piensa que puede recibir tres mensajes: Z_1 y Z_2, al igual que el ejemplo anterior, le confirman que hay sólo 2 personas o le informan que hay 3. El tercer mensaje Z_3 corresponde a la eventualidad de una información ambigua, indefinida (el informante no sabe, no contesta).

La estructura de información es la siguiente (probabilidades conjuntas):

	N_1	N_2	$P(Z_i)$
Z_1	0,974	0,005	0,979
Z_2	0,005	0,010	0,011
$P(N_j)$	0,98	0,02	1

La incertidumbre inicial es $H(N) = 0,14144$

La incertidumbre a posteriori es:

Con Z_1, $P(N_1 / Z_1) = 0{,}995$, $P(N_2 / Z_1) = 0{,}05$, $H(N / Z_1) = 0{,}0454$

Con Z_2, $P(N_1 / Z_2) = 0{,}10$, $P(N_2 / Z_2) = 0{,}90$, $H(N / Z_2) = 0{,}4690$

Con Z_3, $P(N_1 / Z_3) = 0{,}50$, $P(N_2 / Z_3) = 0{,}50$, $H(N / Z_3) = 1$

La incertidumbre promedio final es $H(N / Z) = 0{,}0596$. Con Z_2 y Z_3, la incertidumbre aumenta (con Z_3 es máxima), sólo disminuye con Z_1, que confirma la posición a priori del observador. Con Z_2, la creencia se invierte (¡o debería invertirse!) con menos de 1/3 de bit. Pero, inexorablemente, en promedio, la incertidumbre total disminuye. Lo importante es insistir que no puede dejarse de considerar más de un mensaje posible. Un solo mensaje posible es un mensaje obligatorio y no modifica la situación. Si un mensaje modifica la incertidumbre aumentándola, necesariamente debe existir por lo menos dos (porque si fuere único, no puede afectar la incertidumbre) y éste otro reduce la incertidumbre más que el aumento provocado por el primero a través de la ponderación.

Diseño y diagramación general:

BIBLIOGRAFÍA

BRILLOUIN, León, "La información y la incertidumbre en la ciencia", Universidad Autónoma de Méjico, Méjico, 1969.

BUNGE, Mario, "Treatise on Basic Philosophy: Epistemology and Methodology. Vol.5: Exploring the World", Reidel, Dordrecht, 1983.

HADLEY, G., "Probabilidad y estadística: una introducción a la Teoría de la Decisión", Fondo de Cultura Económica, Méjico, 1979 (1a. edición en inglés,1967).

LAYZER, David, "The Arrow of Time", Scientific American, 223, (1975).56/59.

LUCE, Duncan y RAIFFA, Howard, "Games and Decisions", Wiley, New York, 1957.

MATURANA, Humberto, "Autopoiesis", en: ZELENY, Milan, 1980

MORGAN, Bruce, "An Introduction to Bayesian Statiscal Decision Processes", Prentice Hall, Englewood Cliffs, 1968.

PAVESI, Pedro F.J., "Orden y Entropía: de la Autopoiesis a la Mathesis Universalis", Instituto de Investigaciones Administrativas, 1986 (a).

PAVESI, Pedro F.J., "La incertidumbre del Universo", Instituto de Investigaciones Administrativas, Facultad de Ciencias Económicas, U.B.A., Buenos Aires, 1986. (b), publicado por el CECE.

PRIGOGINE, Ilya, "Tan sólo una ilusión", Tusquets Editores, Barcelona, 1984.

SCHRÖDINGER, Erwin, "Qué es la vida", Tusquets Editores, Barcelona, 1983

SHANNON, Claude, "The Matematical Theory of Communication", The University of Illinois Press, Urbana, 1949.

VARELA, Francisco, "Describing the logic of the living: the adequacy and limitations of the idea of autopoiesis". En: ZELENY, Milan, 1980

ZELENY,Milan, "What is Autopoiesis" en ZELENY, Milan (1980)

ZELENY, Milan, (comp), "Autopoiesis: A Theory of living organisation", North Holland, New York, 1980.