Universidad de Cantabria

SISTEMAS EXPERTOS BASADOS EN PROBABILIDAD

En las aplicaciones, la incertidumbre es lo común y no la excepción. Por ejemplo, una pregunta típica en diagnóstico médico es: dado que el paciente presenta unos síntomas, ¿cuál de las enfermedades posibles es la que tiene el paciente? Esta situación implica incertidumbre puesto que:

- Los hechos o datos pueden no ser conocidos con exactitud. Por ejemplo, un paciente puede no estar seguro de haber tenido fiebre. Hay incertidumbre en la información asociada a cada paciente (subjetividad, imprecisión, errores, datos ausentes, etc.).
- El conocimiento no es determinista. Las relaciones entre las enfermedades y los síntomas no son deterministas, puesto que un mismo conjunto de síntomas puede estar asociado a diferentes enfermedades.

Por ello, son necesarios los sistemas expertos que traten situaciones de incertidumbre.

Inicialmente, se eligió la probabilidad como medida para tratar la incertidumbre, pero se encontraron problemas, debidos al uso incorrecto de las hipótesis de independencia, utilizadas para reducir la complejidad. Consecuentemente, surgieron medidas alternativas a la probabilidad (factores de certeza, credibilidades, plausibilidades, necesidades o posibilidades).

Lindley

"La única descripción satisfactoria de la incertidumbre es la probabilidad. Esto quiere decir que toda afirmación incierta debe estar en forma de una probabilidad, que varias incertidumbres deben ser combinadas usando las reglas de la probabilidad, y que el cálculo de probabilidades es adecuado para manejar situaciones que implican incertidumbre. En particular, las descripciones alternativas de la incertidumbre son innecesarias."

Definición 1 Medida de Probabilidad. *Una función p que proyecta los subconjuntos $A \subseteq S$ en el intervalo $[0,1]$ se llama medida de probabilidad si satisface los siguientes axiomas:*

- **Axioma 1 (Normalización):** $p(S) = 1$.
- **Axioma 2 (Aditividad):** *Para cualquier sucesión infinita, $A_1, A_2, \ldots$, de subconjuntos disjuntos de S, se cumple la igualdad*

$$p\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} p\,(A_i). \tag{1}$$

El Axioma 1 establece que, independientemente de nuestro grado de certeza, ocurrirá un elemento del conjunto universal S (es decir, el conjunto S es exhaustivo). El Axioma 2 es una fórmula de agregación que se usa para calcular la probabilidad de la unión de subconjuntos disjuntos. Establece que la incertidumbre de un cierto subconjunto es la suma de las incertidumbres de sus partes (disjuntas). Nótese que esta propiedad también se cumple para sucesiones finitas.

Universidad de Cantabria

PROPIEDADES DE LA PROBABILIDAD

- **Propiedad 1 (Normalización):** $p(\phi) = 0$.
- **Propiedad 2 (Monotonicidad):** Si $A \subseteq B \subseteq S$, entonces $p(A) \leq p(B)$.
- **Propiedad 3 (Continuidad-Consistencia):** Para toda sucesión creciente $A_1 \subseteq A_2 \subseteq \ldots$ o decreciente $A_1 \supseteq A_2 \supseteq \ldots$ de subconjuntos de S se tiene
$$\lim_{i\to\infty} p(A_i) = p(\lim_{i\to\infty} A_i).$$
- **Propiedad 4(Inclusión-Exclusión) :**

$$p(A \cup B) = p(A) + p(B) - p(A \cap B). \qquad (2)$$

La Propiedad 1 establece que la evidencia asociada a una ausencia completa de información es cero. La Propiedad 2 muestra que la evidencia de la pertenencia de un elemento a un conjunto debe ser al menos la evidencia de cualquiera de sus subconjuntos. La Propiedad 3 puede ser considerada como una propiedad de continuidad. La Propiedad 4 establece que $p(A), p(B), p(A \cup B)$ y $p(A \cap B)$ están relacionadas.

PROBABILIDAD CONDICIONAL E INDEPENDENCIA

Definición 2 Probabilidad condicional. *La probabilidad condicional de X dado $Y = y$ viene dada por*

$$p(X = x|Y = y) = p(x|y) = \frac{p(x,y)}{p(y)}. \tag{3}$$

Definición 3 Independencia de dos variables. *Se dice que X es independiente de Y si y solamente si*

$$p(x|y) = p(x), \tag{4}$$

para todos los valores posibles x e y de X e Y; en otro caso, X se dice dependiente de Y.

Si X es independiente de Y, resulta

$$p(x,y)/p(y) = p(x),$$

que implica

$$p(x,y) = p(x)p(y). \tag{5}$$

que es una definición alternativa de independencia, pero menos intuitiva.

DEPENDENCIA E INDEPENDENCIA CONDICIONAL

Definición 4 Dependencia e independencia condicional. *Sean X, Y y Z tres conjuntos disjuntos, entonces X se dice condicionalmente independiente de Y dado Z, si y sólo si*

$$p(x|z, y) = p(x|z), \tag{6}$$

En otro caso X e Y se dicen condicionalmente dependientes dado Z.

Cuando X e Y son condicionalmente independientes dado Z, se escribe $I(X, Y|Z)$. Similarmente, cuando X e Y son condicionalmente dependientes dado Z, se escribe $D(X, Y|Z)$.
La definición de independencia condicional lleva en sí la idea de que una vez que es conocida Z, el conocimiento de Y no altera la probabilidad de X. Es decir, si Z ya es conocida, el conocimiento de Y no añade información alguna sobre X.
Una definición alternativa, pero equivalente, de independencia condicional es

$$p(x, y|z) = p(x|z)p(y|z). \tag{7}$$

TEOREMA DE BAYES

El teorema de Bayes afirma que:

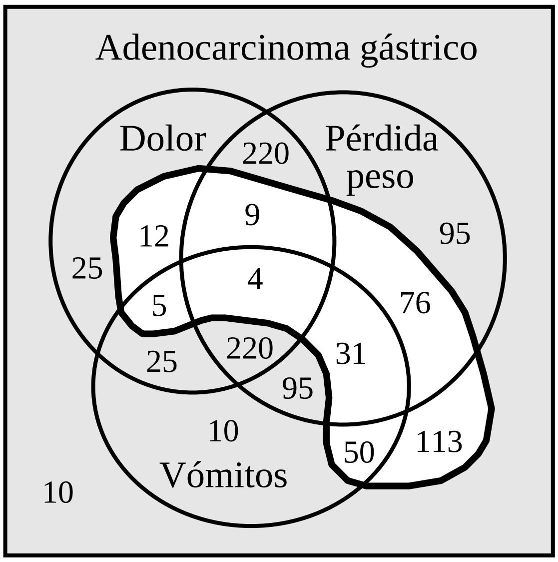
$$p(e_i|s_1,\ldots,s_k) = \frac{p(s_1,\ldots,s_k|e_i)p(e_i)}{\sum_{e_i} p(s_1,\ldots,s_k|e_i)p(e_i)}. \quad (8)$$

- La probabilidad $p(e_i)$ se llama probabilidad *marginal*, *prior*, *"a priori"* o *inicial* de la enfermedad $E = e_i$ puesto que puede ser obtenida *antes* de conocer los síntomas.
- La probabilidad $p(e_i|s_1,\ldots,s_k)$ es la probabilidad *posterior*, *"a posteriori"* o *condicional* de la enfermedad $E = e_i$, puesto que se calcula *después* de conocer los síntomas $S_1 = s_1,\ldots,S_k = s_k$.
- La probabilidad $p(s_1,\ldots,s_k|e_i)$ se conoce por el nombre de *verosimilitud* de que un paciente con la enfermedad $E = e_i$ tenga los síntomas $S_1 = s_1,\ldots,S_k = s_k$.

Por ello, se puede utilizar el teorema de Bayes para actualizar la probabilidad.

Universidad de Cantabria

ADENOCARCINOMA GASTRICO



- Probabilidad “a priori”: 440 de 1,000 pacientes vomitan: $p(v) = 0.44$.
- Verosimilitud: El 50% de los pacientes con la enfermedad vomitan: $p(v|g) = 350/700 = 0.5$, mientras que sólo 30% de los pacientes sin la enfermedad vomitan: $p(v|\bar{g}) = 0.3$.
- Verosimilitud: El 45% de los pacientes con la enfermedad vomitan y pierden peso, $p(v, p|g) = card(v, p, g)/card(g) = 315/700 = 0.45$, mientras que sólo el 12% de los que no tienen la enfermedad vomitan y pierden peso, $p(v, p|\bar{g}) \approx 0.12$.

ADENOCARCINOMA GASTRICO

Síntomas: vómitos y pérdida de peso:

- Probabilidad inicial: $p(g) = 0.7$.
- Tras observar que $V = v$:

$$p(g|v) = \frac{p(g)p(v|g)}{p(g)p(v|g) + p(\bar{g})p(v|\bar{g})}$$
$$= \frac{0.7 \times 0.5}{(0.7 \times 0.5) + (0.3 \times 0.3)} = 0.795.$$

- Tras observar que $V = v$ y $P = p$:

$$p(g|v,p) = \frac{p(g)p(v,p|g)}{p(g)p(v,p|g) + p(\bar{g})p(v,p|\bar{g})}$$
$$= \frac{0.7 \times 0.45}{(0.7 \times 0.45) + (0.3 \times 0.12)} = 0.9.$$

o también:

$$p(g|v,p) = \frac{p(g|v)p(p|g,v)}{p(g|v)p(p|g,v) + p(\bar{g}|v)p(p|\bar{g},v)}$$
$$= \frac{0.795 \times 0.9}{(0.795 \times 0.9) + (0.205 \times 0.389)} = 0.9,$$

TIPOS DE ERRORES

En un caso de diagnóstico médico, por ejemplo, los posibles errores son:

- *Error de Tipo I:* Un paciente no tiene la enfermedad pero el doctor concluye que la tiene.
- *Error de Tipo II:* Un paciente tiene la enfermedad pero el doctor concluye que no.

Decisión médica	Estado de la naturaleza
	Sí
Sí	Decisión correcta
No	Decisión incorrecta (Tipo II)

Decisión médica	Estado de la naturaleza
	No
Sí	Decisión incorrecta (Tipo I)
No	Decisión correcta

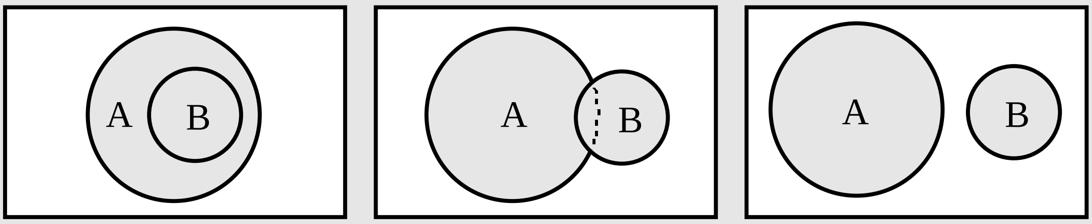
Las consecuencias de un error pueden ser mucho más graves que las consecuencias del otro. Por ejemplo, si la enfermedad sospechada es cáncer, se puede argüir que el error de Tipo II es más serio que el error de Tipo I.

REGLAS GENERALIZADAS

Implicación fuerte ($\theta = 1$): En la lógica clásica: *Si A es cierta, entonces B es cierta*, se puede decir que A implica B con probabilidad 1. Esto se ilustra en la Figura (a).

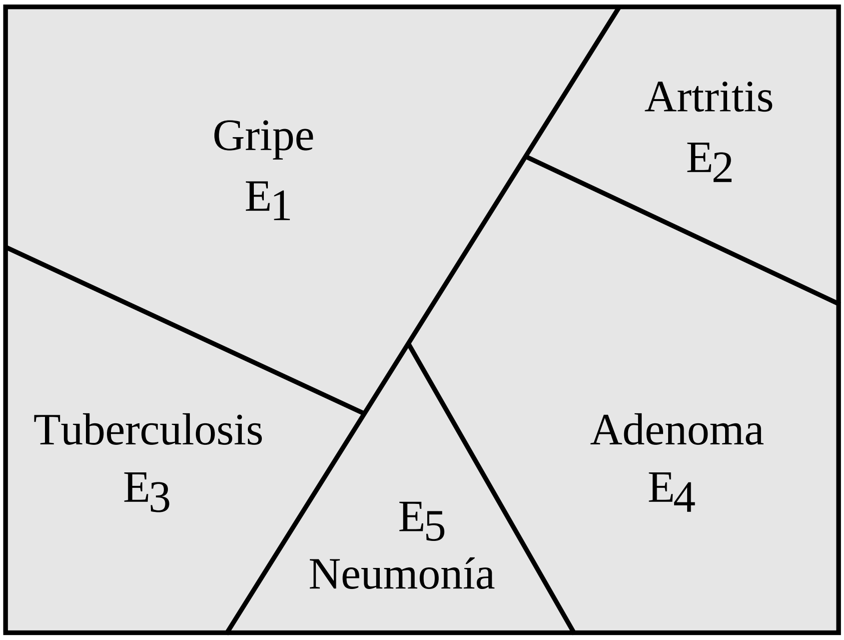
Implicación débil ($0 < \theta < 1$): La regla anterior puede ser vista en un sentido generalizado cuando A implica B sólo en algunas ocasiones. En este caso, se dice que A implica B con probabilidad $p(B = cierto|A = cierto)$ (Figura (b)).

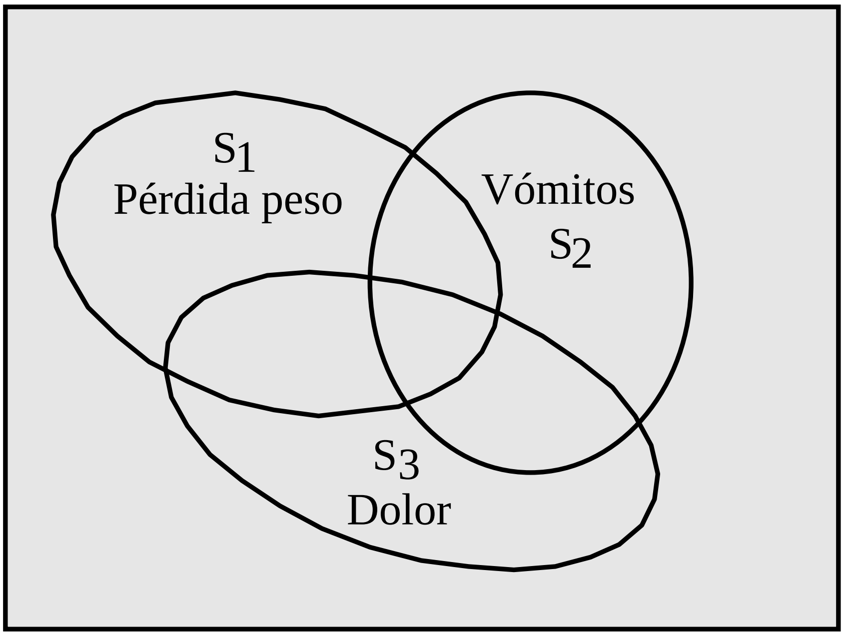
No implicación ($\theta = 0$): El caso en que A no implica B puede considerarse como que A implica B con probabilidad 0. Esto se ilustra en la Figura (c).



(a) p(B | A) = 1 (b) 0 < p(B | A) < 1 (c) p(B | A) = 0

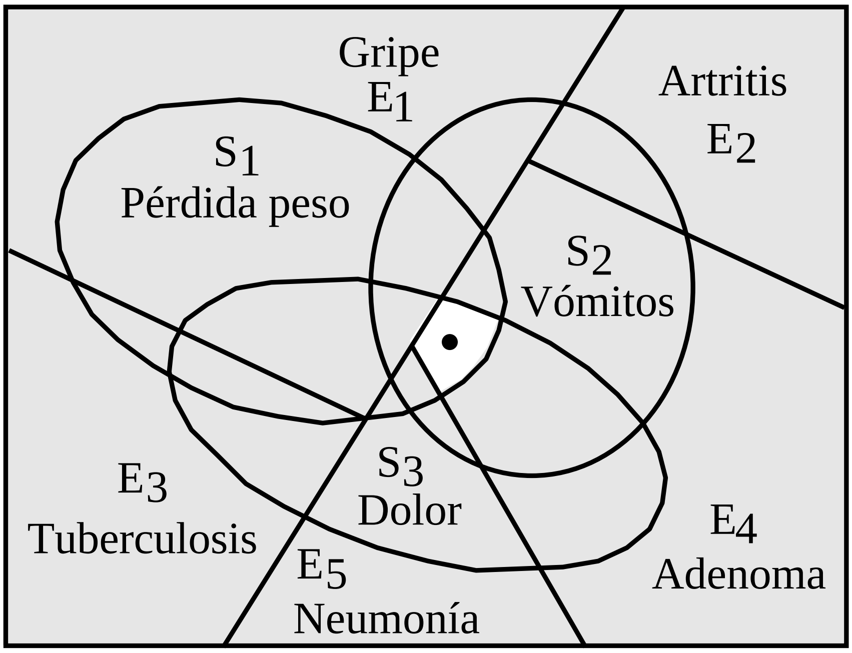
SINTOMAS Y ENFERMEDADES





Universidad de Cantabria

DIAGRAMA CONJUNTO SINTOMAS Y ENFERMEDADES



Paciente	Enfermedad E	Síntomas S_1	S_2	S_3
1	e_5	1	1	1
2	e_2	1	0	1
3	e_3	1	1	0
4	e_5	0	0	1
5	e_3	0	1	0
6	e_1	1	1	0
7	e_1	1	1	1
8	e_3	1	0	0

Universidad de Cantabria

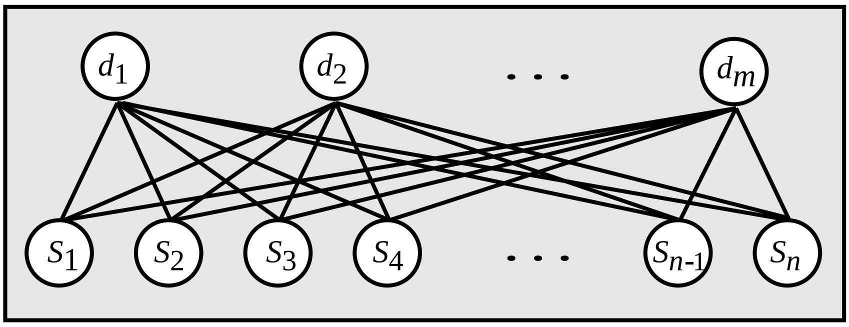
MODELOS

El modelo más general posible se basa en especificar directamente la función de probabilidad conjunta; es decir, asignar un valor numérico (parámetro) a cada una de las posibles combinaciones de valores de las variables. Desgraciadamente, la especificación directa de la función de probabilidad conjunta implica un gran número de parámetros.
Por tanto se usan los siguientes modelos simplificados:

1. **MSD: El Modelo de Síntomas Dependientes.**
2. **MSI: El Modelo de Síntomas Independientes.**
3. **MSRI: El Modelo de Síntomas Relevantes Independientes.**
4. **MSRD: El Modelo de Síntomas Relevantes Dependientes.**

MODELO DE SINTOMAS INDEPENDIENTES

Debido a la complicación del modelo DSM, es necesario simplificar. Una simplificación posible consiste en suponer que los síntomas son condicionalmente independientes dada la enfermedad, lo que conduce al modelo ISM, en la que los síntomas no están ligados (independencia).



Con esta hipótesis resulta:

$$p(s_1, \ldots, s_n | d_i) = \prod_{j=1}^{n} p(s_j | d_i). \tag{9}$$

por lo que

$$p(d_i | s_1, \ldots, s_n) = \frac{p(d_i) p(s_1, \ldots, s_n | d_i)}{p(s_1, \ldots, s_n)}$$

$$= \frac{p(d_i) \prod_{j=1}^{n} p(s_j | d_i)}{p(s_1, \ldots, s_n)} \propto p(d_i) \prod_{j=1}^{n} p(s_j | d_i). \tag{10}$$

MODELO DE SINTOMAS INDEPENDIENTES

La ecuación anterior muestra que la probabilidad inicial de la enfermedad d_i es $p(d_i)$, y que tras conocer el síntoma s_j, es proporcional a $p(s_j|d_i)$. Nótese que cada síntoma conduce a un nuevo factor y que $p(s_1, \ldots, s_n)$ es una constante de normalización, que no necesita calcularse.

Los parámetros necesarios para la base de conocimiento del modelo ISM son

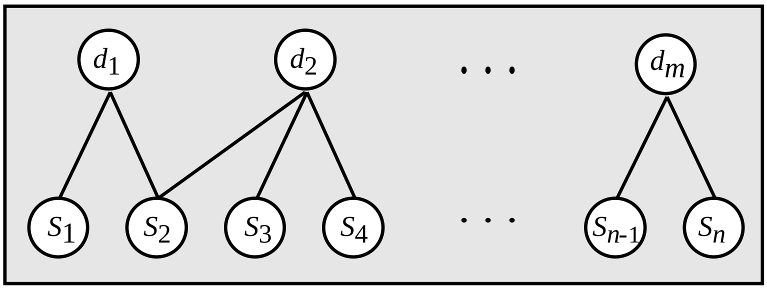
- Las probabilidades marginales $p(d_i)$, para todos los valores posibles de la enfermedad D.
- Las probabilidades condicionales $p(s_j|d_i)$, para todos los valores posibles de los síntomas S_j y la enfermedad D.

Por ello, con esta hipótesis se reduce considerablemente el número de parámetros. Con m enfermedades posibles y n síntomas binarios, el número total de parámetros es

$$\boxed{m(n+1)-1.}$$

MODELO DE SINTOMAS RELEVANTES INDEPENDIENTES

Se puede conseguir una redución aún mayor del número de parámetros si para cada valor e_i de la enfermedad E se seleccionan algunos síntomas relevantes $S_1, \ldots, S_r$ y los restantes síntomas se suponen independientes para ese valor de E.



Supóngase que $S_1, \ldots, S_{r_i}$ son relevantes para la enfermedad e_i y que los restantes síntomas $S_{r_i+1}, \ldots, S_n$ son irrelevantes. Se supone $p_j = p(s_j|e_i)$ idéntica para todos los síntomas que son irrelevantes para la enfermedad e_i. Entonces:

$$p(e_i|s_1, \ldots, s_n) = \frac{p(e_i)p(s_1,\ldots,s_n|e_i)}{p(s_1,\ldots,s_n)}$$

$$= \frac{p(e_i) \prod_{j=1}^{r_i} p(s_j|e_i) \prod_{j=r_i+1}^{n} p(s_j|e_i)}{p(s_1,\ldots,s_n)}$$

$$= \frac{p(e_i) \prod_{j=1}^{r_i} p(s_j|e_i) \prod_{j=r_i+1}^{n} p_j}{p(s_1,\ldots,s_n)} \propto p(e_i) \prod_{j=1}^{r_i} p(s_j|e_i) \prod_{j=r_i+1}^{n} p_j,$$

NODELO DE SINTOMAS RELEVANTES DEPENDIENTES

Es necesario almacenar las probabilidades siguientes en la base de conocimiento del MSRI:

- Las probabilidades marginales $p(e_i)$, para todos los valores posibles de la enfermedad E.
- Las probabilidades condicionales $p(s_j|e_i)$, para cada valor posible de E y cada uno de sus correspondientes síntomas relevantes.
- Las probabilidades p_j, para cada valor posible de E que tiene al menos un síntoma irrelevante.

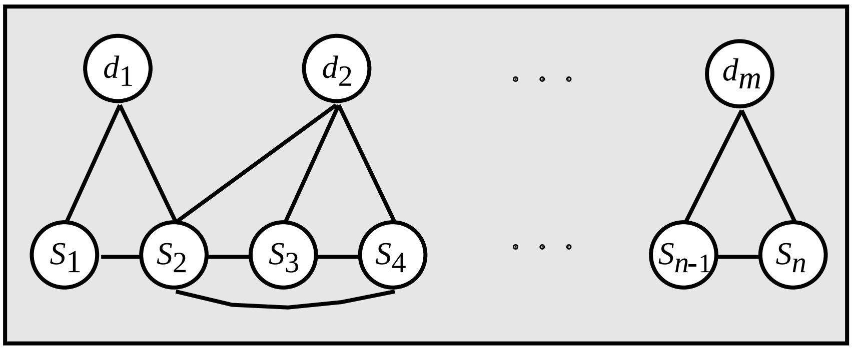
Por ello, si se tienen m posibles enfermedades y n síntomas binarios, el número de parámetros en el MSRI es

$$m - 1 + n - a + \sum_{i=1}^{m} r_i, \tag{11}$$

donde r_i es el número de síntomas relevantes para la enfermedad e_i y a es el número de síntomas que son relevantes para todas las enfermedades.

NODELO DE SINTOMAS RELEVANTES DEPENDIENTES

Aunque el MSRI reduce el número de parámetros, desgraciadamente, es poco realista, ya que los síntomas suelen producirse en grupos o síndromes. Por ello, es poco razonable suponer que los síntomas relevantes son independientes.



Supóngase que $S_1, \ldots, S_{r_i}$ son relevantes para la enfermedad e_i y que los restantes síntomas $S_{r_i+1}, \ldots, S_n$ son irrelevantes. Entonces:

$$
\begin{aligned}
p(e_i|s_1, \ldots, s_n) &= \frac{p(e_i)p(s_1, \ldots, s_{r_i}|e_i) \prod\limits_{j=r_i+1}^{n} p(s_j|e_i)}{p(s_1, \ldots, s_n)} \\
&= \frac{p(e_i)p(s_1, \ldots, s_{r_i}|e_i) \prod\limits_{j=r_i+1}^{n} p_j}{p(s_1, \ldots, s_n)} \\
&\propto p(e_i)p(s_1, \ldots, s_{r_i}|e_i) \prod\limits_{j=r_i+1}^{n} p_j.
\end{aligned}
$$

COMPARACION DE VARIOS MODELOS

Para este modelo, es necesario almacenar las siguientes probabilidades en la base de datos:

- Las probabilidades marginales $p(e_i)$, para todos las posibles valores de la enfermedad E.
- Las probabilidades $p(s_1, \ldots, s_{r_i}|e_i)$, para todos los posibles valores de la enfermedad E y sus síntomas relevantes $S_1, \ldots, S_{r_i}$.
- Las probabilidades p_j, para cada valor posible de E que tenga al menos un síntoma irrelevante. (Como en el MSRI, esto implica que $p_j = p(s_j|e_i)$ coincide para todos los síntomas irrelevantes para e_i.)

En consecuencia, para m enfermedades binarias y n síntomas binarios, el número total de parámetros en el MSRD es

$$m - 1 + n - a + \sum_{i=1}^{m} (2^{r_i} - 1) = n - 1 - a + \sum_{i=1}^{m} 2^{r_i}.$$

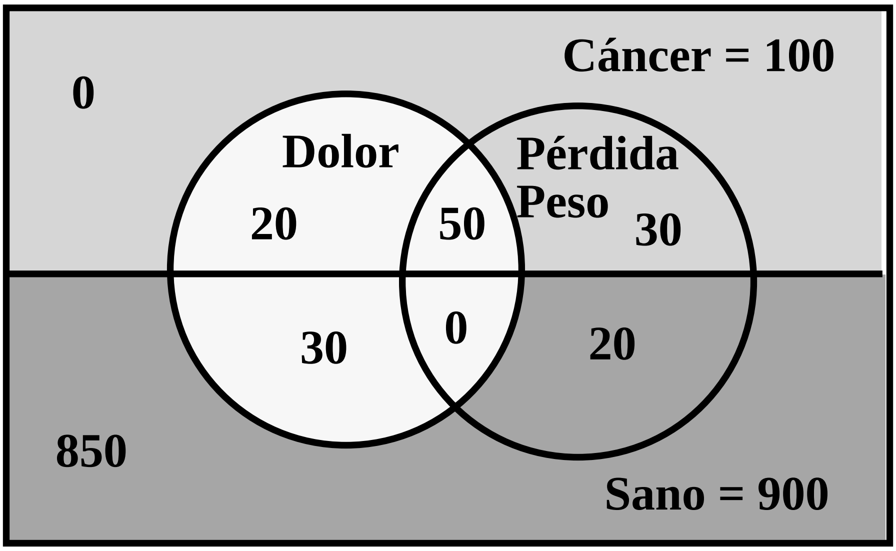
COMPARACION DE VARIOS MODELOS

La Tabla muestra una comparación de los números de parámetros requeridos para especificar los diferentes modelos en el caso de $m = 100$ enfermedades binarias, $n = 200$ síntomas binarios, y $r = 10$ síntomas relevantes por enfermedad.

Modelo	Número de parámetros	
	Fórmula	Valor
DSM	$m2^n - 1$	$> 10^{62}$
ISM	$m(n+1) - 1$	20,099
IRSM	$m(r+1) + n - 1$	1,299
DRSM	$m2^r + n - 1$	102,599

En el DRSM el número de parámetros se reduce notablemente comparado con el DSM, aunque el modelo es todavía realista por considerar dependencias entre los síntomas relevantes para cada enfermedad. Una reducción extra se consigue dividiendo el conjunto de síntomas relevantes en bloques, que son independientes entre sí.

ILUSTRACION PROBABILIDAD CONDICIONADA



P (Cáncer) = 100 / 1000 =0.1

$$P(\text{Cáncer / Dolor}) = \frac{P(\text{Cáncer y Dolor})}{P(\text{Dolor})} = 70 / 100 = 0.7$$

P (Cáncer / Dolor y Pérdida Peso) =

$$= \frac{P(\text{Cáncer, Dolor y Pérdida Peso})}{P(\text{Dolor y Pérdida Peso})} = 50 / 50 = 1$$