CI5438. Inteligencia Artificial II
Clase 9: Incertidumbre
Cap 13 Russel & Norvig

Ivette C. Martínez

Universidad Simón Bolívar

10 de Noviembre de 2008

Incertidumbre

Sea A_t = Salir hacia el aeropuerto t minutos antes del vuelo.
Será suficiente A_t para llegar a tiempo al aeropuerto?.

Problemas:

1. Parcialmente observable (estado de las vías, planes de otros conductores, etc.)
2. Sensores con ruido (reportes de tráfico)
3. Incertidumbre en el resultado de las acciones (cauchos gastados, etc.)
4. Gran complejidad para modelar y predecir el tráfico.

Por lo tanto, una aproximación púramente lógica puede:

1. Correr el riesgo de ser falsa: "A_{25} me llevará al aeropuerto a tiempo"
2. Llevarnos a conclusiones que son muy débiles para tomar decisiones: " A_{25} me llevara a tiempo al aeropuerto si no hay accidentes en el puente y si no llueve, y si mis cauchos están en buen estado, y si ..."

(A_{1440} puede parecer razonable para llegar a tiempo, pero no queremos pasar la noche en el aeropuerto)

Métodos para manejar la incertidumbre

Lógica no-monotónica:

- Asumo que mi carro no tiene los cauchos lisos
- Asumo que A_{25} funciona a menos que la evidencia lo contradiga.

Problemas:

- Qué suposiciones son razonables?
- Cómo manejar las contradicciones?

Métodos para manejar la incertidumbre

Reglas con factores estimados:

- $A_{25} \rightarrow_{0,3} AtAirportOnTime$
- $Sprinkler \rightarrow_{0,99} WetGrass$
- $WetGrass \rightarrow_{0,7} Rain$—

Problemas:

- Problemas con las combinaciones, e.j., *Sprinkler* causa *Rain*?

Probabilidades:

Dada la evidencia disponible: A_{25} me llevará a tiempo al aeropuerto con probabilidad 0,04

Probabilidades

Las afirmaciones probabilísticas resumen los efectos de:

- **pereza**: fallas para enumerar excepciones, calificaciones, etc
- **ignoracia**: falta de hechos relevantes, condiciones iniciales, etc.

Probabilidades Subjetivas o Bayesianas:
Probabilidades que relacionan proposiciones a los estados de creencia propios.
ejm. $P(A_{25} | no\_reported\_accidents) = 0{,}06$
No son aseveraciones de una "tendencia probabilística" en la situación actual (pero se pueden aprender de las experiencias pasadas en situaciones semejantes)
Las probabilidades de las propociones cambian con nuevas evidencias:
ejm. $P(A_{25} | no\_reported\_accidents, 5a.m.) = 0{,}15$

Supongamos que creo lo siguiente:
$P(A_{25}\_gets\_me\_there\_on\_time|\ldots) = 0{,}04$
$P(A_{90}\_gets\_me\_there\_on\_time|\ldots) = 0{,}70$
$P(A_{120}\_gets\_me\_there\_on\_time|\ldots) = 0{,}95$
$P(A_{1440}\_gets\_me\_there\_on\_time|\ldots) = 0{,}9999$
¿Cuál acción seleccionar?
Depende de mis **preferencias** para vuelos perdidos v.s. comida de aeropuerto, etc.

La **Teoría de Utilidad** es usada para representar e inferir preferencias.

Teoría de Decisión =
Teoría de Utilidad + Teoría de Probabilidades

Elementos de Probabilidades

Comencemos con el conjunto Ω - El espacio de muestras
ejm. 6 posibles lanzadas de un dado
$\omega \in \Omega$ es una muestra, evento

Un **espacio de probabilidades** o **modelo de probabilidades** es una asignación $P(\omega)$ para cada $\omega \in \Omega$. Estas asignaciones deben cumplir las siguientes restricciones:

- $0 \leq P(\omega) \leq 1$
- $\sum_{\omega} P(\omega) = 1$

ejm: $P(1) = P(2) = P(3) = P(4) = P(5) = P(6) = 1/6$

Un **evento** A es un subconjunto cualquiera de Ω
$P(A) = sum_{\omega \in A} P(\omega)$ ejm:
$P(die\_roll < 4) = P(1) + P(2) + P(3) = 1/2$

Variables Aleatorias

Una **variable aleatoria** es una función desde los puntos de ejemplo sobre algún rango. Ejm.: Los reales o los Booleanos.
ejm.: $Odd(1) = true$

P induce una distribución de probabilidades para cualquier variable aleatoria X:
$P(X = x_i) = \sum_{\{\omega : X(\omega) = x_i\}} P(\omega)$
Ejm: $P(Odd = true) = P(1) + P(3) + P(5) = 1/2$

Proposiciones

Pensemos en una proposición como un evento (conjunto de puntos de ejemplo) donde la proposición es cierta.

Dadas las variables booleanas aleatorias A y B

- el evento a = el conjunto de ejemplos donde $A(\omega) = true$
- el evento $\neg a$ = el conjunto de ejemplos donde $A(\omega) = false$
- el evento $a \wedge b$ = el conjunto de ejemplos donde $A(\omega) = true$ y $B(\omega) = true$

Usualmente en aplicaciones de IA, los puntos de ejemplo se definen por los valores del conjunto de variables aleatorias, i.e., el espacio muestral es el producto cartesiano de los rangos de las variables.

Con variables booleanas, punto de ejemplo = modelo lógico proposicional. ejm: $A = true, B = false$, o $a \wedge \neg b$

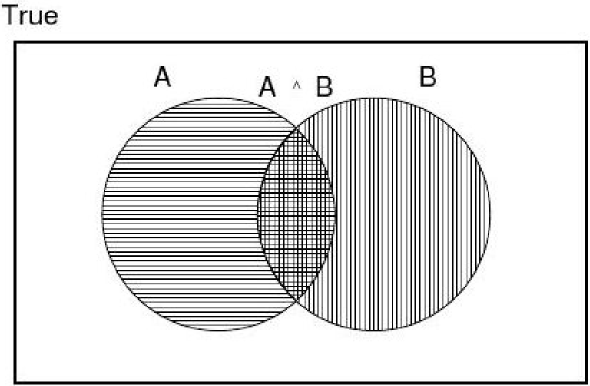
Proposición = Disyunción entre eventos atómicos en los que es cierta.

ejm: $(a \vee b \equiv (\neg a \wedge b) \vee (a \wedge \neg b) \vee (a \wedge b))$

$\Rightarrow P(a \vee b) = P(\neg a \wedge b) + P(a \wedge \neg b) + (a \wedge b))$

Por qué usar Probabilidades?

Las definiciones implican que los eventos relacionados lógicamente deben tener probabilidades relacionadas.
Ejm: $P(a \vee b) = P(a) + P(b) - P(a \wedge b)$



Sintaxis para las proposiciones

- Variable aleatorias Proposicionales o Booleanas
  ejm. *Cavity* (tengo una caries?)
  Cavity$=$ *true* es una proposición, que tambien se escribe *cavity*.
- Variables aleatorias Discretas (finitas o infinitas)
  ejm. *Weather* es $\{$*sunny,rain,cloudy, snow*$\}$
  Weather$=$ *sunny* es una proposición.
  Los valores deber ser exaustivos y mútuamente excluyentes.
- Variables aleatorias Contínuas (acotadas o no acotadas)
  ejm. *Temp*$=$ 21,6; tambien puede ser *Temp*$<$ 22
- Combinaciones booleanas de proposiciones básicas

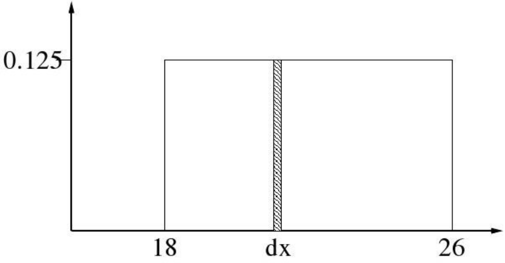
Probabilidades a Priori

- Probabilidades a Priori o no-condicionales.
  ejm.: $P(Cavity = true) = 0{,}1$ y $P(Weather = sunny) = 0{,}72$
  Se corresponden con las creencias previas a arribo de cualquier (nueva) evidencia.
- **Distribución de probabilidades**. Le otorga valores a todas las posibles asignaciones:
  $P(Weather) =< 0{,}72, 0{,}1, 0{,}08, 0{,}1 >$
- la **Distribución de probabilidades conjunta** para un conjunto de variables aleatorias le asigna probabilidades para cada evento atómico (i.e., cada punto de ejemplo).
  $P(Weather, Cavity)$ = la siguiente matriz de valores:

Weather =	sunny	rain	cloudy	snow
Cavity = true	0.144	0.02	0.016	0.02
Cavity = false	0.576	0.08	0.064	0.08

Probabilidades para variables contínuas

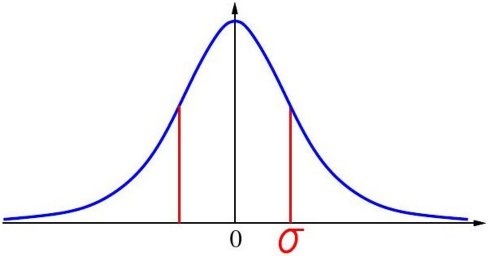
Expresar la distribución como una función parametrizada del valor:
$P(X = x) = U[18, 26](x) =$ densidad uniforme entre 18 y 26



Aqui P es una densidad; integra a 1.
$P(X = 20{,}5) = 0{,}125$ realmente significa:
$\lim_{dx \to 0} P(20{,}5 \leq X \leq 20{,}5 + dx) = 0{,}125$

Densidad Gaussiana

$$P(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-(x-\mu)^2/2\sigma^2}$$



Probabilidad Condicional

- Probabilidades Condicionales o a posteriori.
  ejm. $P(cavity|toothache) = 0{,}8$
  i.e., dado que "toothache" es todo lo que sabemos, NO "si toothache entonces hay un chance de 80 % de cavity"
- Si sabemos mas, ejm. Tambien tenemos *cavity*, entonces tenemos que $P(cavity|toothache, cavity) = 1$
- La nueva evidencia puede ser irrelevante, permitiendo ciertas simplificaciones,
  ejm:
  $P(cavity|toothache, 49ersWin) = P(cavity|toothache) = 0{,}8$.
  Este tipo de inferencias, sancionadas por el conocimiento del dominio es crucial.

Probabilidad Condicional

- Definición:

$$P(a|b) = \frac{P(a \wedge b)}{P(b)} \text{ si } P(b) \neq 0$$

- La regla del producto nos da una formulación alternativa:

$$P(a \wedge b) = P(a|b)P(b) = P(b|a)P(a)$$

- Una versión general se cumple para distribuciones completas. Ejm:

$$P(Weather, Cavity) = P(Weather|Cavity)P(Cavity)$$

Probabilidad Condicional

La **regla de la cadena** se deriva de aplicaciones sucesivas de la reglas del producto:

$P(X_1, ..., X_n) = P(X_1, ..., X_{n-1})P(X_n|X_1, ..., X_{n-1})$

$= P(X_1, ..., X_{n-2})P(X_{n-1}|X_1, ..., X_{n-2})$

$= P(X_n|X_1, ..., X_{n-1})$

$= ...$

$= \prod_{i=1}^{n} P(X_i|X_1, ..., X_{i-1})$

Inferencia por enumeración

Comencemos con la siguiente distribución conjunta:

	toothache		$\neg$ *toothache*	
	catch	$\neg$ *catch*	*catch*	$\neg$ *catch*
cavity	.108	.012	.072	.008
$\neg$ *cavity*	.016	.064	.144	.576

Para cualquier proposición θ, sumar los eventos atómicos en los que es cierta:

$$P(\phi) = \sum_{\omega:\omega\vdash\phi} P(\omega)$$

Inferencia por enumeración

Comencemos con la siguiente distribución conjunta:

	toothache		$\neg$ *toothache*	
	catch	$\neg$ *catch*	*catch*	$\neg$ *catch*
cavity	**.108**	**.012**	**.072**	**.008**
$\neg$ *cavity*	**.016**	**.064**	**.144**	**.576**

Para cualquier proposición θ, sumar los eventos atómicos en los que es cierta:

$$P(\phi) = \sum_{\omega:\omega\vdash\phi} P(\omega)$$

$$P(\mathit{toothache}) = 0{,}108 + 0{,}012 + 0{,}016 + 0{,}064 = 0{,}2$$

Inferencia por enumeración

Comencemos con la siguiente distribución conjunta:

	toothache		$\neg$ *toothache*	
	catch	$\neg$ *catch*	*catch*	$\neg$ *catch*
cavity	**.108**	**.012**	**.072**	**.008**
$\neg$ *cavity*	**.016**	**.064**	**.144**	**.576**

Para cualquier proposición θ, sumar los eventos atómicos en los que es cierta:

$$P(\phi) = \sum_{\omega:\omega\vdash\phi} P(\omega)$$

$P(cavity \vee toothache) = 0{,}108 + 0{,}012 + 0{,}072 + 0{,}008 + 0{,}016 + 0{,}064 = 0{,}28$

Inferencia por enumeración

Comencemos con la siguiente distribución conjunta:

	toothache		$\neg$ *toothache*	
	catch	$\neg$ *catch*	*catch*	$\neg$ *catch*
cavity	.108	.012	.072	.008
$\neg$ *cavity*	.016	.064	.144	.576

Tambien podemos calcular probabilidades condicionales:

$$P(\neg cavity|toothache) = \frac{P(\neg cacity \wedge toothache)}{P(toothache)}$$

$$P(\neg cavity|toothache) = \frac{0{,}016 + 0{,}064}{0{,}108 + 0{,}012 + 0{,}016 + 0{,}064} = 0{,}4$$

Normalización

	toothache		*¬ toothache*	
	catch	*¬ catch*	*catch*	*¬ catch*
cavity	$\boxed{.108}$	$\boxed{.012}$	.072	.008
¬ cavity	$\boxed{.016}$	$\boxed{.064}$	.144	.576

El denominador tambien puede ser visto como una **constante de normalización** α.

$P(Cavity|toothache) = \alpha P(Cavity, toothache)$

$= \alpha[P(Cavity, toothache, cath) + P(Cavity, toothache, \neg cath)]$

$= \alpha[< 0{,}108, 0{,}016 > + < 0{,}012, 0{,}064 >]$

$= \alpha < 0{,}12, 0{,}008 >$

$=< 0{,}6, 0{,}4 >$

Idea General: Calcular la distribución de la variable en cuestión fijando las "variables de evidencia" y sumando sobre las "variables ocultas".

Inferencia por enumeración

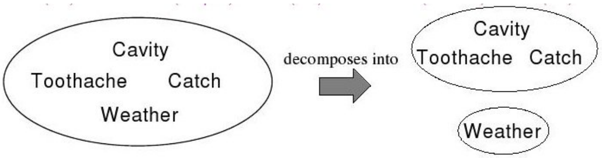
Sea X el conjunto de todas las variables. Usualmente queremos la dsitribución conjunta a posteriori de las variables de consulta Y dados valores específicos e de kas variables de evidencia E.
Sea H el conjunto de **variables ocultas**. $H = X - Y - E$
Entonces, la suma requerida de las entradas conjuntas se realiza sumando sobre las variables ocultas:

$$P(Y|E = e) = \alpha P(Y, E = e) = \alpha \sum_{h} P(Y, E = e, H = h)$$

Nótese que Y, E y H juntas cubren completamente el conjunto de variables aleatorias del dominio.

Independencia

A y B son independientes si y solo si:
$P(A|B) = P(A)$ o $P(B|A) = P(B)$ o $P(A, B) = P(A)P(B)$



$P(Toothache, Catch, Cavity, Weather) =$
$P(Toothache, Catch, Cavity)P(Weather)$

Independencia Condicional

$P(Toothache, Catch, Cavity)$ tiene $2^3 = 8$ entradas independientes.

- Si tengo una caries, la probabilidad de "catch" no depende de que tenga dolor de muelas:
  $P(catch|toothache, cavity) = P(catch|cavity)$
- La misma independencia se presenta si no tengo caries:
  $P(catch|toothache, \neg cavity) = P(catch|\neg cavity)$

Catch es independiente condicionalmente de *Toothache* dado *Cavity*:
$P(Catch|Toothache, Cavity) = P(Catch|Cavity)$

Independencia Condicional

Escribamos la distribución conjunta completa usando la regla de la cadena:
$P(Toothache, Catch, Cavity)$
$= P(Toothache|Catch, Cavity)P(Catch, Cavity)$
$= P(Toothache|Catch, Cavity)P(Catch|Cavity)P(Cavity)$
$= P(Toothache|Cavity)P(Catch|Cavity)P(Cavity)$
I.e., $2 + 2 + 1 = 5$ números independientes
En la mayoría de los casos el uso de la independencia condicional reduce el tamaño de la representación de la distribución conjunta de exponencial en n a lineal en n.
La independecia condicional es la forma más básica y robusta de conocimiento en ambientes con incertidumbre.

Regla de Bayes

Regla del producto: $P(a \wedge b) = P(a|b)P(b) = P(b|a)P(a)$

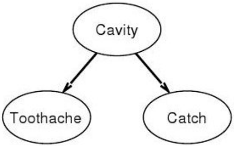
Regla de Bayes: $P(a|b) = \frac{P(b|a)P(a)}{P(b)}$

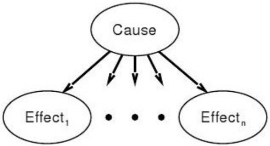
o en forma de distribución:
$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)} = \alpha P(X|Y)P(Y)$

Últil para calcular la probabilidad de un diagnóstico a partir de una probabilidad causal:
$P(Causa|Effect) = \frac{P(Effect|Cause)P(Cause)}{P(Effect)}$

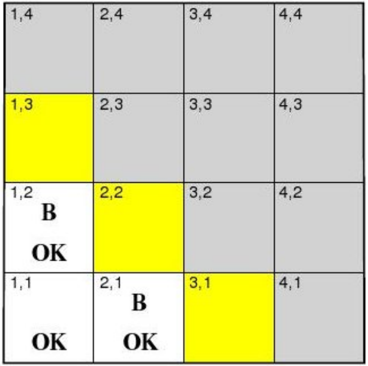
$P(Cavity|toothache \wedge catch)$
$= \alpha P(toothache \wedge catch|Cavity)P(Cavity)$
$= \alpha P(toothache|Cavity)P(catch|Cavity)P(Cavity)$

Este es un ejemplo del modelo de "Bayes simple":
$P(Cause, Effect_1, ..., Effect_n) = P(Cause) \prod_i P(Effect_i|Cause)$





Wumpus World



$P_{ij} = \textit{true}$ si y solo si $[i, j]$ contiene un hoyo
$B_{ij} = \textit{true}$ si y solo si $[i, j]$ tiene viento
incluimos solo $B_{1,1}$, $B_{1,2}$ y $B_{2,1}$ en el modelo

Especíﬁcando el modelo de probabilidades

La distribución conjunta completa es
$P(P_{1,1}, ...P_{4,4}, B_{1,1}, B_{1,2}, B_{2,1})$

Aplicando la regla del producto:
$P(B_{1,1}, B_{1,2}, B_{2,1}|P_{1,1}, ...P_{4,4})P(P_{1,1}, ...P_{4,4})$
(Esto se hace para tener P(Efecto—Causa)).

Primer término: 1 si los hoyos están adyacentes al viento, 0 de otra forma
Segundo término: los hoyos se ubican aleatóriamente, con probabilidad 0,2 por cuadrado:
$P(P_{1,1}, ...P_{4,4}) = \prod_{i,j=1,1}^{4,4} P(P_{i,j}) = 0{,}2^n x 0{,}8^{16-n}$ para n hoyos.

Observaciones y Preguntas

Conocemos los siguientes hechos:
$b = \neg b_{1,1} \wedge b_{1,2} \wedge b_{2,1}$ $known = \neg p_{1,1} \wedge p_{1,2} \wedge p_{2,1}$

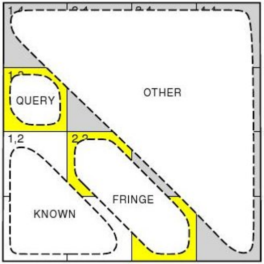
La pregunta es: $P(P_{1,3}|known, b)$

Definimos $Unknown = P'_{i,j}$ diferentes de $P_{1,3}$ y $known$

Para la inferencia por enumeración, tenemos:
$P(P_{1,3}|known, b) = \alpha \sum_{unknown} P(P_{1,3}, unknown, known, b)$
Crece exponencialmente con el número de cuadros!.

Usando independencia condicional

Las observaciones son independientes condicionalmente de otros cuadrados ocultos dadas vecidades de cuadrados ocultos



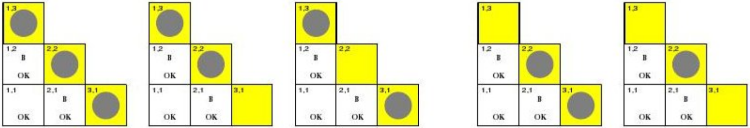
Definimos: $Unknown = Fringe \cup Other$

$P(b|P_{1,3}, Known, Unknown) = P(b|P_{1,3}, Known, Fringe)$

Usando independencia condicional

$$
\begin{aligned}
&\mathbf{P}(P_{1,3}|known, b) = \alpha \sum_{unknown} \mathbf{P}(P_{1,3}, unknown, known, b) \\
&= \alpha \sum_{unknown} \mathbf{P}(b|P_{1,3}, known, unknown, b)\mathbf{P}(P_{1,3}, known, unknown, b) \\
&= \\
&\alpha \sum_{fringe} \sum_{other} \mathbf{P}(b|known, P_{1,3}, fringe, other)\mathbf{P}(P_{1,3}, known, fringe, other) \\
&= \alpha \sum_{fringe} \sum_{other} \mathbf{P}(b|known, P_{1,3}, fringe)\mathbf{P}(P_{1,3}, known, fringe, other) \\
&= \alpha \sum_{fringe} \mathbf{P}(b|known, P_{1,3}, fringe) \sum_{other} \mathbf{P}(P_{1,3}, known, fringe, other) \\
&= \\
&\alpha \sum_{fringe} \mathbf{P}(b|known, P_{1,3}, fringe) \sum_{other} \mathbf{P}(P_{1,3})P(known)P(fringe)P(other) \\
&= \\
&\alpha P(known)\mathbf{P}(P_{1,3}) \sum_{fringe} \mathbf{P}(b|known, P_{1,3}, fringe)P(fringe) \sum_{other} P(other) \\
&= \alpha' \mathbf{P}(P_{1,3}) \sum_{fringe} \mathbf{P}(b|known, P_{1,3}, fringe)P(fringe)
\end{aligned}
$$

Usando independencia condicional (cont.)



0.2 x 0.2 = 0.04 | 0.2 x 0.8 = 0.16 | 0.8 x 0.2 = 0.16 | 0.2 x 0.2 = 0.04 | 0.2 x 0.8 = 0.16

$$P(P_{1,3}|known, b) = \alpha' < 0{,}2(0{,}04 + 0{,}16 + 0{,}16), 0{,}8(0{,}04 + 0{,}16) >$$
$$P(P_{1,3}|known, b) \approx < 0{,}31, 0{,}69 >$$

$$P(P_{2,2}|known, b) \approx < 0{,}86, 0{,}14 >$$

Resumen

- Las probabilidades son un formalismo riguroso para representar conocimiento incierto.
- La distribución de probabilidad conjunta especifica la probabilidad para cada evento atómico
- Podemos contestar preguntas sobre el domino sumando sobre los eventos atómicos
- Para dominios no triviales, debemos hallar una forma de reducir el tamaño de la distribución conjunta,
- La independencia y la independencia condicional nos propocionan estos mecanismos.