

## Práctica 4. Redes Bayesianas

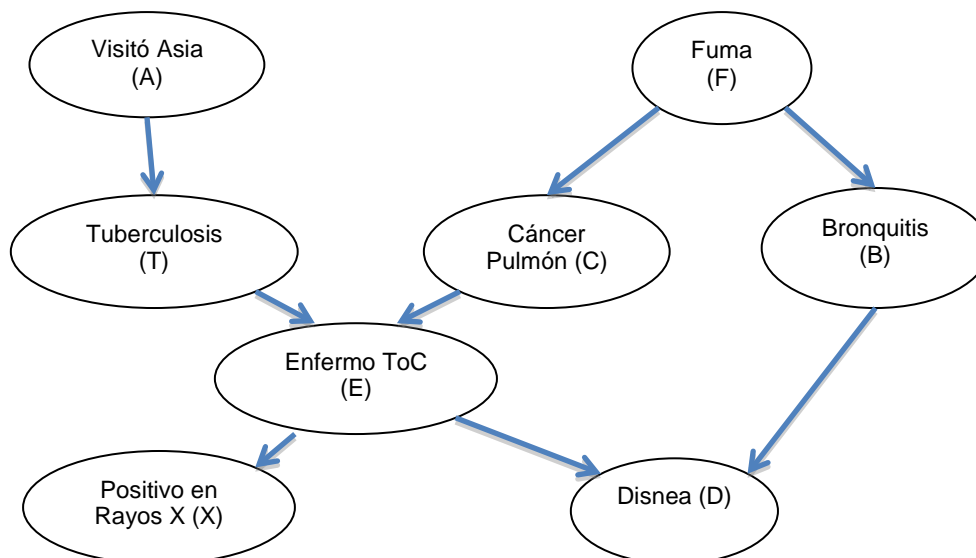
### Objetivo

El objetivo de esta práctica es aprender a modelar dominios con incertidumbre mediante redes Bayesianas y utilizar técnicas de razonamiento probabilista. Utilizaremos la herramienta de dominio público OpenMarkov, desarrollada en la UNED: <http://www.openmarkov.org/>

Nota: cuidado con OpenMarkov, cuando dibuja una red de 3 nodos, los dibuja alineados, puede haber una flecha entre el primer nodo y el tercero que sólo verás si mueves alguno de los nodos.

### Estudio previo (en papel)

1. Consideremos esta red Bayesiana para diagnóstico de enfermedades del pulmón (conocida como red Asia), con las tablas de probabilidades indicadas.



$P(a) = 0.01$	$P(e c, t) = 1$	$P(f) = 0.50$	$P(d e, b) = 0.90$	$P(b f) = 0.60$
	$P(e c, -t) = 1$		$P(d e, -b) = 0.70$	$P(b -f) = 0.30$
$P(t a) = 0.05$	$P(e -c, t) = 1$	$P(c f) = 0.10$	$P(d -e, b) = 0.80$	
$P(t -a) = 0.01$	$P(e -c, -t) = 0$	$P(c -f) = 0.01$	$P(d -e, -b) = 0.10$	

$$P(x|e) = 0.98$$

$$P(x|-e) = 0.05$$

- Explica **razonadamente** cuáles de estas propiedades son ciertas:
  1. A y B son independientes dado X
  2. A y B son independientes dados E y F
- Calcula siguiendo el algoritmo de eliminación de variables la probabilidad de que tenga tuberculosis un paciente que ha dado positivo en rayos X y fuma, si se sabe que no tiene cáncer de pulmón. Si hay alguna variable que se pueda ignorar, explica claramente por qué.

2. En un estudio sobre discriminación de género en la admisión de la Universidad de Berkeley, California, se recogieron los siguientes datos:

	Hombres		Mujeres	
Carrera	Solicitudes	Admitidos	Solicitudes	Admitidas
A	825	512	108	89
B	560	353	25	17
C	325	120	593	202
D	417	138	375	131

Se pide:

- Representa el problema con una **Red Bayesiana causal**. Piensa muy bien cuáles son las variables aleatorias necesarias, y cuáles las relaciones que pueden aparecer entre ellas. No hagas ninguna hipótesis de independencia entre variables salvo que esté soportada por los datos de la tabla.
- Escribe las tablas de probabilidad de cada nodo, estimando las probabilidades a partir de los datos de la tabla.
- A la vista de las tablas de probabilidad, ¿ha podido haber discriminación de género en la admisión?. Razona la respuesta.
- Utilizando el algoritmo de **eliminación de variables**, calcula la probabilidad de admisión para una mujer.

### Desarrollo de la práctica

Copia en tu ordenador los ficheros del archivo p4.zip y comprueba que puedes ejecutar OpenMarkov (necesita Java 7). Responde a las siguientes cuestiones:

- Crea la red Bayesiana para las enfermedades del pulmón (red Asia), e introduce sus tablas de probabilidades.
  - Haz las pruebas necesarias para comprobar si son ciertas las dos propiedades de independencia condicional del apartado 1.a) del estudio previo. Explica en el informe cómo lo has hecho e incluye capturas de pantalla con el resultado.
  - Calcula la probabilidad pedida en el apartado 1.b) del estudio previo. Comprueba que las variables que pensabas que se pueden ignorar, no influyen en el resultado. Explica en el informe cómo lo has hecho e incluye capturas de pantalla con el resultado.
- El fichero DatosBerkeley.xls contiene datos simulados de admisión de alumnos en Berkeley. Prueba a aprender la Red Bayesiana con OpenMarkov a partir de estos datos (menú Herramientas → Aprendizaje + Botón de opción “Aprendizaje Automático”)
  - Comprueba si la red y las tablas de probabilidad son como habías previsto. Si no lo son, discute las diferencias.
  - Calcula la probabilidad de admisión para un hombre y para una mujer.
  - Si son distintas, ¿es porque hubo discriminación de género, o puede haber algún otro motivo?

- 
3. El fichero DatosAsia.xls contiene datos simulados del ejemplo de enfermedades del pulmón. Prueba a aprender la red a partir de estos datos, con la opción de aprendizaje automático
- ¿Se obtiene una red causal?
  - Con la red obtenida, calcula la probabilidad de que tenga tuberculosis un paciente que ha dado positivo en rayos X, no fuma, y no ha visitado Asia. Haz una captura de pantalla con la red para el informe.
  - Repite el proceso con la opción de aprendizaje interactivo, forzando manualmente a que aparezcan sólo enlaces causales.
  - Calcula con la nueva red la probabilidad de que tenga tuberculosis un paciente que ha dado positivo en rayos X, no fuma, y no ha visitado Asia. ¿Sale lo mismo que en el apartado b)? Analiza y discute por qué. Incluye captura de pantalla en el informe.

**A entregar**

Un único fichero p4.pdf con los resultados de todos los apartados. El estudio previo puedes escanearlo o fotografiarlo con el móvil.