Modelos gráficos probabilísticos

Ejercicios de trabajo, parte 1

Curso 2022-2023

---

1. Teoría de la probabilidad

1. Se tiene una urna A con 3 bolas blancas y 7 rojas y otra urna B con con 5 bolas blancas y 5 negras. Se realiza el siguiente experimento: se lanza un dado y si sale menor o igual a 2, se elige la primera urna y en otro caso la segunda. Una vez elegida la urna, se elige una bola al azar de ella. Calcular las siguientes probabilidades:

   - que la urna sea la A y que la bola sea blanca
   - que la bola sea roja o que la urna sea la B
   - que ni la bola sea blanca ni la urna sea la B

2. Se considera el experimento de lanzar una moneda sin trucar. Si sale cara, entonces se lanza un dado de 4 caras; en caso contrario, se lanza un dado de 6 caras. Determinar la probabilidad de obtención de un 3.

3. La probabilidad de que un sábado por la noche una persona vaya el cine es de 0.3 y la probabilidad de que vaya a un restaurante es de 0.2, siendo 0.05 la probabilidad de que haga ambas cosas. Calcular las siguientes probabilidades:

   - probabilidad de que vaya al cine o al restaurante
   - probabilidad de que vaya al cine y no vaya al restaurante

4. Si estoy dispuesto a pagar 0.2 euros por una apuesta en un juego en el que recibo 1 euro si ocurre el suceso A, ¿qué se puede afirmar sobre mi creencia en el valor de probabilidad $P(A)$?

5. Se dispone de 3 urnas con la siguiente composición:

   - A, con 7 bolas rojas y 3 blancas
   - B, con 4 rojas y 6 blancas
   - C, con 1 roja y 9 blancas

   Se elige una urna al azar con distribución uniforme y se extrae una bola de la urna. Calcular:

   - la probabilidad de que la bola sea blanca
   - si la bola extraída ha sido roja, calcular la probabilidad de que la urna elegida haya sido la C
   - determinad la distribución conjunta de probabilidad para este problema
   - determinar las distribuciones marginales para la elección de la urna y para la bola extraída
   - determinar la distribución condicionada de la urna seleccionada dado el color de la bola

6. Se transmiten símbolos $\{0, 1\}$ por un canal de comunicación. Se sabe que la probabilidad de transmitir un 0 es 0.4 Si se transmite un 0 se recibe el símbolo correcto el 90 % de las veces y si se transmite un 1 la transmisión correcta se da en el 99 % de los casos. Calcular la probabilidad de que se haya emitido un 1 si se ha recibido un 1. Determinar la distribución conjunta de probabilidad para los símbolos emitidos y recibidos.

7. Un granjero sospecha que una de las vacas está preñada. Realiza un test de orina para comprobar si es realmente así. Hay 4 posibles resultados:

   - la vaca está preñada y el test es positivo
   - la vaca está preñada y el test es negativo
   - la vaca no está preñada y el test es positivo
   - la vaca no está preñada y el test es negativo

   La probabilidad a priori de que la vaca esté preñada es 0.05. El test tiene fiabilidad del 98 % si la vaca está preñada y del 99.99 % si no lo está. El resultado del test es positivo. ¿Cuál es la probabilidad a posteriori de que la vaca esté preñada?

8. Las salidas a pasear por el campo de un individuo dependen del tiempo que hace, de la siguiente forma:

   - si hace sol, la probabilidad de salir a pasear es 0.95
   - si está nublado, es de 0.85
   - si llueve, es de 0.3
   - en el lugar donde vive hace sol con probabilidad 0.7, está nublado con probabilidad 0.2 y llueve con probabilidad 0.1.

   Se pide modelizar el problema con las variables X (tiempo) e Y (salir a pasear). Determinar la distribución conjunta de estas variables, la distribución marginal de Y y la condicionada de X dado Y.

9. X e Y son variables independientes que toman valores en el conjunto $\{0, 1\}$. Se sabe que $P(X = 0) = 0.3$ y $P(Y = 0) = 0.8$. Determinar la distribución de probabilidad conjunta para estas variables.

10. Una variante genética (variable X) está presente en el 1 % de la población. Se sabe que esa variante genética predispone a dos enfermedades (variables Y y Z). Si la variante X está presente, la probabilidad de ocurrencia de la enfermedad Y es del 0.1 y si no está presente de 0.001. Si la variante X está presente, la probabilidad de ocurrencia de la enfermedad Z es de 0.3 y si no está presente baja a 0.01. Se sabe que Y y Z son independientes dado X. Determinar la distribución de probabilidad conjunta para las tres variables.

11. En una habitación hay dos interruptores que pueden estar en las posiciones 0 ó 1. Llamemos X e Y a sus posiciones. Supongamos que cuando los dos interruptores están en la misma posición la luz de la habitación está encendida y está apagada cuando están en posiciones diferentes. Sea Z la variable que representa si la luz están encendida o apagada. Si las posiciones de los interruptores siguen la distribución uniforme, determinar una distribución de probabilidad conjunta para las tres variables. ¿Qué relaciones de independencia se verifican y cuáles no?

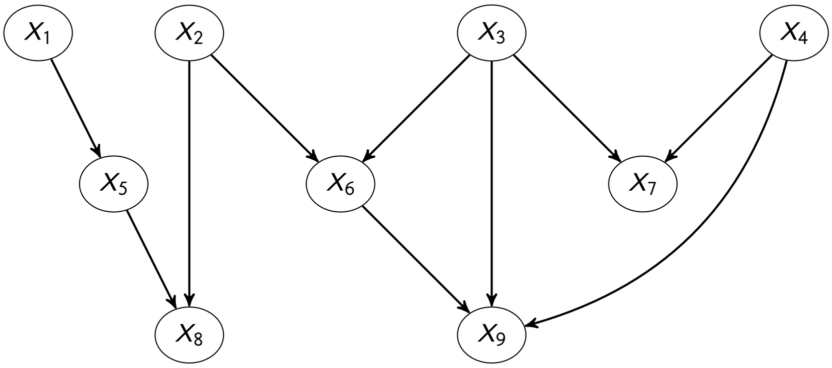
12. Consideremos las variables X, Y y Z, con posibles valores $\{0, 1\}$ y la siguiente distribución conjunta:

X	Y	Z	p
0	0	0	0.0104
0	0	1	0.0898
0	1	0	0.0078
0	1	1	0.432
1	0	0	0.014
1	0	1	0.126
1	1	0	0.032
1	1	1	0.288

   ¿Qué relaciones de independencia se dan entre estas tres variables?

2. D-separación

1. Dada la siguiente red:

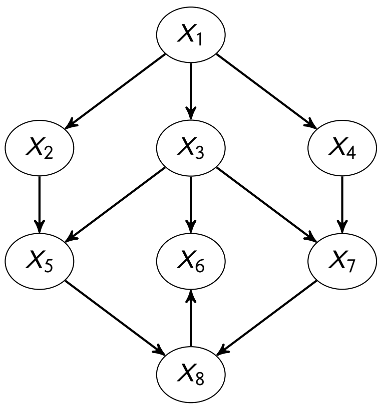


justificar si se cumplen las siguientes condiciones:

- $I(X_1, X_4|\varnothing)$
- $I(X_1, X_4|X_8, X_9)$
- $I(X_1, X_4|X_2, X_8, X_9)$
- $I(X_2, X_3|X_9)$
- $I(X_2, X_3|X_6)$

Observad que I indica independencia, que se cumple en el caso de haber *d-separación* entre las variables que aparecen indicadas en la parte izquierda asumiendo que las variables de la parte derecha forman parte del conjunto evidencia. Por ejemplo, para determinar si $I(X_1, X_4|X_8, X_9)$ debe comprobarse si X_1 y X_4 están $d-separadas$ dado que disponemos de evidencia sobre X_8 y X_9. Y el caso $I(X_1, X_4|\varnothing)$ requiere determinar si X_1 está $d-separada$ de X_4 asumiendo que no disponemos de evidencia sobre el resto de variables.

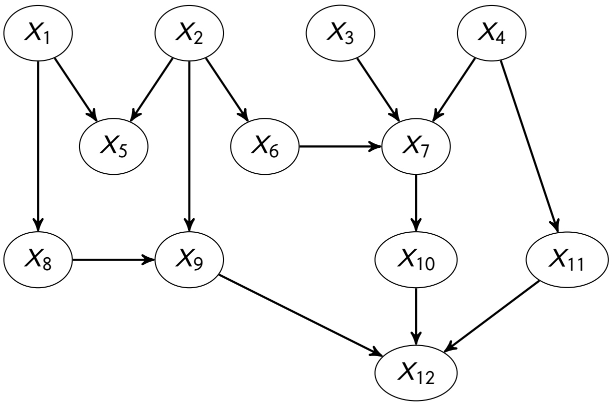
2. Dada la siguiente red:



Justifica si se cumplen las siguientes condiciones:

- $I(X_1, X_3|\varnothing)$
- $I(X_1, X_6|X_3, X_8)$
- $I(X_1, X_8|X_5, X_6, X_7)$

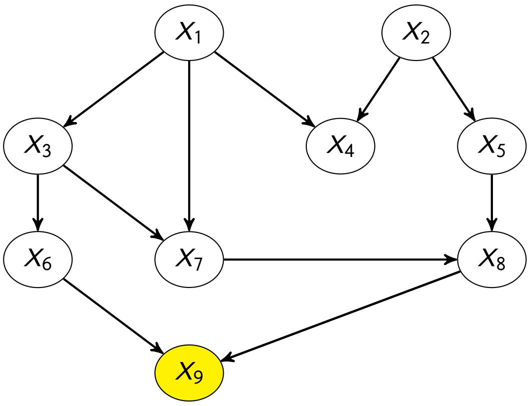
3. Dada la red de la figura inferior:



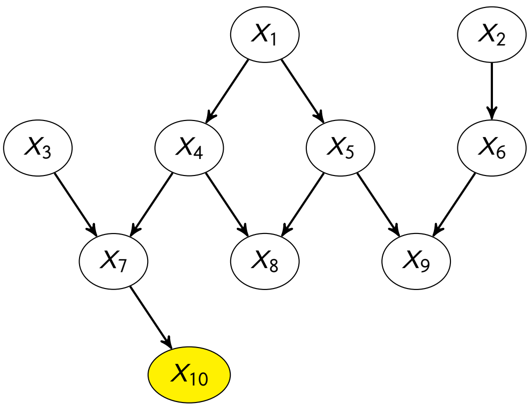
Justifica si se cumplen las siguientes condiciones:

- $I(X_2, X_4|\emptyset)$
- $I(X_1, X_2|X_2, X_9, X_{10})$
- $I(X_1 3 X_4|X_{10})$
- $I(X_1, X_2|X_9)$

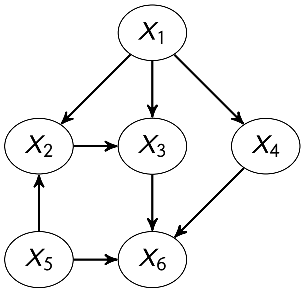
4. Dada la siguiente red, donde el nodo coloreado representa la evidencia, determina todas las variables d-separadas de X_9



5. Dada la siguiente red, donde el nodo coloreado representa la evidencia, determina todas las variables d-separadas de X_3



6. Dada la siguiente red, determina el manto de Markov de todas las variables de la red.



3. Modelización

1. Determina la estructura de una red bayesiana para el siguiente problema: un granjero quiere determinar si una vaca está preñada (P) después de una inseminación artificial. Para ello dispone de tres tipos de test diferentes: ecografía (TE), sangre (TS) y orina (TO). Los dos últimos dependen del nivel hormonal (H), que puede ser alto o bajo, dependiendo de si está o no preñada. Hay un tipo de sangre (X) especial que hace que el resultado del test de sangre sea siempre positivo, con independencia de si la vaca está o no preñada. Para descartar esta situación, se puede realizar otro test para comprobar el tipo de sangre (TX). Una vez determinado el modelo, indicar cómo serían las tablas de probabilidad necesarias para la especificación completa del modelo.

2. La probabilidad de que una vaca sufra mastitis (M) un día depende de varios factores: si ya padecía esta dolencia el día anterior (D), del número de días habitual para la enfermedad (N) y número de días que lleva recibiendo tratamiento (T). La enfermedad se diagnostica mediante la observación del estado general de la vaca (G) y de un test sobre la leche (L). El test puede no ser fiable si la vaca ha estado sometida a tratamiento durante más de tres días. Describe el grafo de dependencias compatible con esta descripción del problema.

3. Se considera el problema de transmitir palabras de longitud 5, formadas por el alfabeto $A = \{a, b\}$ sobre un canal de transmisión. Las palabras se transmiten símbolo a símbolo. Hay que tener en cuenta que la línea es susceptible de tener ruido, de forma que no siempre se recibe el símbolo emitido. Si se emite el símbolo a se recibe correctamente con probabilidad 0.8; si se emite una b la probabilidad de recepción correcta es 0.9. La probabilidad de error depende únicamente del símbolo emitido. Se sabe que las palabras emitidas no son aleatorias, de forma que el valor de un símbolo depende del símbolo emitido con anterioridad. Se trata de describir una red bayesiana que represente este problema, para la emisión de una palabra completa.

4. En una granja hay dos yeguas y un caballo, sin parentesco entre ellos. Van a nacer dos potros, uno de cada yegua. El caballo es el padre de ambos. Existe una enfermedad que está ligada a la presencia de un gen recesivo a (el gen normal es A). De esta forma, la carga genética de cada individuo puede ser AA, Aa y aa. La enfermedad solo se manifiesta si el individuo tiene carga genética aa. En caso de tener Aa es portador de la enfermedad, aunque no la manifiesta. Se trata de:

   - determinar una red bayesiana que exprese las dependencias entre la carga genética de cada uno de los caballos de la granja, incluyendo los potros por nacer.
   - si en la población general la probabilidad de ser portador es de 0.01 para los caballos y 0.02 para las yeguas, sabiendo que los potros heredan un gen de cada padre, detallar las distribuciones de probabilidad de la red.

5. Un determinado defecto genético (variable G) puede producir dos enfermedades (variables E_1, E_2). En presencia de dicho defecto, las enfermedades se manifiestan con una determinada probabilidad, pero no existe ninguna relación entre los mecanismos que dan lugar a las enfermedades. Es decir, el hecho de que una se manifieste no hace a la otra más o menos probable. Hay tres posibles síntomas asociados a las enfermedades (S_1, S_2, S_3). Los dos primeros se asocian a la enfermedad E_1 y los dos últimos a la enfermedad E_2. En la enfermedad E_1 la presencia del síntoma S_1 hace al síntoma S_2 más probable. En la enfermedad E_2 la presencia de uno de los síntomas no cambia la probabilidad de

aparición del otro. Se puede realizar una prueba de laboratorio P, cuyo resultado depende de forma conjunta de la presencia o ausencia de ambas enfermedades, pero tiene comportamiento distinto en hombres y mujeres (variable X). Se asume que X no tiene relación directa con ninguna otra variable del problema. Se pide:

- determinar un grafo dirigido con las variables anteriores que represente unas relaciones de independencia entre las variables que sean compatibles con las especificaciones anteriores.
- en el caso de incluir hipótesis adicionales, indicarlas de forma precisa.

6. Consideremos un modelo de la intereracción entre tres factores (hierba, herbívoros, carnívoros) (variables X_1, X_2 y X_3 respectivamente) en un sistema ecológico. Para ello se consideran tres instantes de tiempo y, en cada uno de ellos, los valores de estas tres variables (hay que hacer tres versiones de cada una de las variables, una en cada instante de tiempo). Cada variable tiene tres valores posibles ('escaso','normal','abundante'). Se trata de:

   - determinar una red bayesiana que represente el modelo, describiendo la parte cualitativa y cuantitativa.
   - para las probabilidades se deben de determinar valores numéricos que sean compatibles con la intuición (por ejemplo, si hay pocos carnívoros y muchos herbívoros en un momento dado, en el instante siguiente lo más probable es que el número de carnívoros sea normal).

7. Ha habido un crimen para el que hay un sospechoso y se pretende analizar la evidencia con una red bayesiana, considerando la relación previa del sospechoso con la víctima, una mancha de sangre encontrada en la escena y una huella de zapato. Las variables que vamos a considerar son:

   - R: el sospechoso tenía una relación sentimental con la víctima (con valores Si y No)
   - H: el sospechoso es culpable
   - G: la mancha de sangre es del criminal
   - T: la huella es del crimimal
   - S: la mancha de sangre es del sospechoso
   - U: la huella es del sospechoso
   - E: el análisis de DNA de la mancha coincide con el sospechoso
   - F: la huella del zapato es compatible con el sospechoso

   Dibujad un grafo dirigido acíclico que represente correctamente las independencias condicionadas existentes entre las variables del problema. Hacer explícitas hipótesis de independencia que se hayan usado y que resulten dudosas en el enunciado del problema.

8. Describid un grafo dirigido acíclico que represente las dependencias e independencias del siguiente problema, que relaciona la posibilidad de que una persona tenga una infección vírica en dos instantes, relacionándolo con la exposición al virus, las pruebas realizadas, la fortaleza de su sistema inmunológico y los síntomas. Las variables son:

   - $X1$ la persona tiene la infección vírica en $t = 1$
   - $X2$ la persona tiene la infección vírica en $t = 2$
   - $C1$ la persona ha estado en contacto con personas infectadas en un tiempo anterior a $t = 1$
   - $C2$ la persona ha estado en contacto con personas infectadas en un tiempo anterior a $t = 2$
   - $T1$ la persona tiene tos en $t = 1$
   - $T2$ la persona tiene tos en $t = 2$
   - $F1$ la persona tiene fiebre en $t = 1$
   - $F2$ la persona tiene fiebre en $t = 2$

- $PCR1$ la persona ha dado positivo en $t = 1$ a una prueba PCR
- $PCR2$ la persona ha dado positivo en $t = 2$ a una prueba PCR
- $A1$ la persona ha dado positivo a una prueba de anticuerpos en $t = 1$
- $A2$ la persona ha dado positivo a una prueba de anticuerpos en $t = 2$
- S fortaleza del sistema inmunitario

Se supone lo siguiente:

- si la persona está infectada en el instante 1, entonces puede seguir infectada en el instante 2, aunque no haya estado en contacto con una persona infectada entre el tiempo 1 y 2 (no se ha curado todavía)
- el test PCR sólo es sensible al hecho de que la persona tenga la infección en ese momento
- el test de anticuerpos mide si la persona ha tenido la enfermedad en este momento o en el pasado
- los enfermos pueden ser asintomáticos, pero la aparición de un síntoma hace al otro más probable
- los síntomas en el instante 2 solo dependen de si la persona está infectada en ese instante: es decir, conocida $X2$, los síntomas son independientes del resto de las variables.