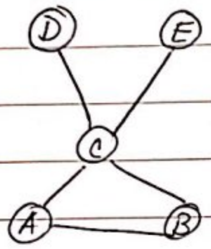
Tarea 05: de Est-46114 Métodos Multivariados

[1] Definan un grafo matemático con vértices $\{A, B, C, D, E\}$ tal que los conjuntos $\{A, B\}$ y $\{E\}$ sean separados por el conjunto $\{C, D\}$



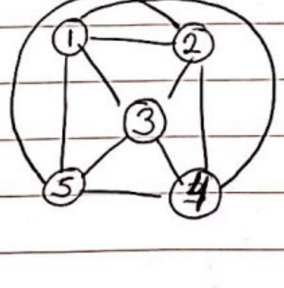
con los Nodos $= \{A, B, C, D, E\}$ y
las Bordes $= \{\{D, C\}, \{E, C\}, \{C, A\}, \{A, B\}, \{C, B\}\}$

[2] Definan un grafo matemático con vértices $\{A, B, C, D, E\}$ tal que los conjuntos $\{A, B\}$ y $\{E\}$ sean separados por el conjunto $\{C\}$ pero no por el conjunto $\{C, D\}$.

No existe una solución para este caso, ya que si estuvieran separados por el conjunto $\{C\}$ también deberían estar separados por el conjunto $\{C, D\}$.

[3] Escriban los elementos del grafo probabilístico saturado de una tabla de contingencias de 5 dimensiones donde cada dimensión tiene 2 categorías (i.e. modelo loglineal saturado 2^5).

Para ejemplificar un modelo loglineal saturado se tiene un grafo intuitivo como este:



donde todos los nodos están conectados entre sí.
Vértices $= \{1, 2, 3, 4, 5\}$

y por lo tanto sus bordes son las combinaciones a pares dadas por:

Bordes = { {1,2} {2,3} {3,4} {4,5}
{1,3} {2,4} {3,5}
{1,4} {2,5}
{1,5} }

Como se tienen 5 dimensiones, las variables son: X_1, X_2, X_3, X_4 y X_5 con la Probabilidad:

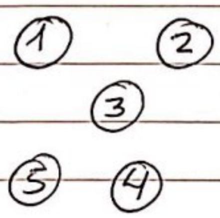
$$P(X_1, X_2, X_3, X_4, X_5) = \phi_{1,2}(X_1, X_2)\,\phi_{1,3}(X_1, X_3)\,\phi_{1,4}(X_1, X_4) \cdot \phi_{1,5}(X_1, X_5) \cdot \phi_{2,3}(X_2, X_3) \cdot \phi_{2,4}(X_2, X_4) \cdot \phi_{2,5}(X_2, X_5) \cdot \phi_{3,4}(X_3, X_4) \cdot \phi_{3,5}(X_3, X_5) \cdot \phi_{4,5}(X_4, X_5)$$

El modelo sería $= \log(P(X_1 = e(1), X_2 = e(2), X_3 = e(3), X_4 = e(4), X_5 = e(5)))$

Expresado como: $\log\theta_v = a^{\emptyset} + a^{1,2}_{e(1,2)} + a^{1,3}_{e(1,3)} + a^{1,4}_{e(1,4)} + a^{1,5}_{e(1,5)} + a^{2,3}_{e(2,3)} + a^{2,4}_{e(2,4)} + a^{2,5}_{e(2,5)} + a^{3,4}_{e(3,4)} + a^{3,5}_{e(3,5)} + a^{4,5}_{e(4,5)}$

4 Escriban los elementos del grafo probabilístico independiente de una tabla de contingencias donde cada dimensión tiene 2 categorías (i.e. modelo loglineal saturado 2^5).

El grafo asociado sería similar al anterior pero sin incluir las conexiones (bordes) entre los nodos ya que se trata de un modelo independiente



Los vértices son $= \{1, 2, 3, 4, 5\}$

Se tienen las mismas variables que en el modelo anterior : $\{X_1, X_2, X_3, X_4, X_5\}$ porque es un modelo 2^5

El modelo loglineal es: $P(X_1, X_2, X_3, X_4, X_5) = \Phi_1(X_1) \cdot \Phi_2(X_2) \cdot \Phi_3(X_3) \cdot \Phi_4(X_4) \cdot \Phi_5(X_5)$

Que se expresa como: $\log(P(X_1 = e(1), X_2 = e(2), X_3 = e(3), X_4 = e(4), X_5 = e(5))$

$$\log \theta_v = a^0 + a^1 + a^2 + a^3 + a^4 + a^5$$

$\boxed{5}$ Respecto a los modelos loglineales de la sección 4.4 de las notas de la sesión de hoy, contrasten, para el ejemplo de la tabla de contingencia de reptiles, las probabilidades muestrales del:

a) modelo independiente

vs.

b) las probabilidades muestrales.

Es decir, identifiquen si las probabilidades muestrales pueden expresarse como el producto de las probabilidades marginales de las tres dimensiones.

Se establece primero una tabla de contingencias dada por los conteos correspondientes (se extrae la tabla de la sesión 21 de Datos Categóricos y Modelos de Grafos

Con $X_3 \leq 4.75$ →

$X_1 \backslash X_2$	≤ 4	>4
anoli	86	35
dist	73	70

Con $X_3 \geq 4.75$ →

$X_1 \backslash X_2$	≤ 4	>4
anoli	32	11
dist	61	41

Las probabilidades del modelo independiente $P(X_1, X_2, X_3)$ son las siguientes:

$$\begin{aligned}
&\quad\ X_1 \quad X_2 \quad X_3 \\
&P(\text{anoli}, \leq 4, \leq 4.75) = \tfrac{86}{409} = 0.2103 \\
&P(\text{anoli}, \leq 4, > 4.75) = 32/409 = 0.0782 \\
&P(\text{anoli}, > 4, \leq 4.75) = 35/409 = 0.0856 \\
&P(\text{anoli}, > 4, > 4.75) = 11/409 = 0.0269
\end{aligned}$$

$$\begin{aligned}
&P(\text{dist}, \leq 4, \leq 4.75) = 73/409 = 0.1785 \\
&P(\text{dist}, \leq 4, > 4.75) = 61/409 = 0.1491 \\
&P(\text{dist}, > 4, \leq 4.75) = 70/409 = 0.1711 \\
&P(\text{dist}, > 4, > 4.75) = 41/409 = 0.1002
\end{aligned}$$

Para las probabilidades muestrales expresadas como el producto de las probabilidades marginales de las 3 dimensiones $P(X_1 = x_1) \cdot P(X_2 = x_2) \cdot P(X_3 = x_3)$ se tiene:

a) Probabilidades Marginales)

$$P(X_1 = \text{anoli}) = 164/409 = 0.4010$$
$$P(X_2 = \text{dist}) = 245/409 = 0.5990$$

$$P(X_2 = \text{"}\leq 4\text{"}) = 252/409 = 0.6161$$
$$P(X_2 = \text{"}> 4\text{"}) = 157/409 = 0.3839$$

$$P(X_3 = \text{"}\leq 4.75\text{"}) = 264/409 = 0.6455$$
$$P\,X_3 = \text{"}> 4.75\text{"}) = 145/409 = 0.3545$$

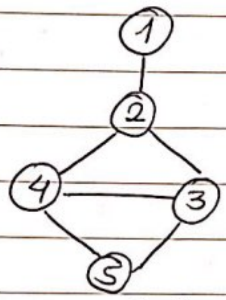
b) Producto de Marginales:

$$\begin{aligned}
&P(\text{anoli}, \leq 4, \leq 4.75) = 0.1594 \\
&P(\text{anoli}, \leq 4, > 4.75) = 0.0875 \\
&P(\text{anoli}, > 4, \leq 4.75) = 0.0993 \\
&P(\text{anoli}, > 4, > 4.75) = 0.0545 \\
&P(\text{dist}, \leq 4, \leq 4.75) = 0.2382 \\
&P(\text{dist}, \leq 4, > 4.75) = 0.1308 \\
&P(\text{dist}, > 4, \leq 4.75) = 0.1484 \\
&P(\text{dist}, > 4, > 4.75) = 0.0815
\end{aligned}$$

Al hacer el contraste de las probabilidades, se observa que son diferentes y por lo tanto se podría decir que con estos resultados, el modelo no parece ser independiente.

6. Para el grafo g_3 del ejemplo visto en clase, identifiquen si $\{X_1\}$ es condicionalmente independiente a $\{X_3, X_5\}$ dado $\{X_2\}$

El grafo g_3 :



Nodos = $\{1,2,3,4,5\}$
Vértices = $\{X_1, X_2, X_3, X_4, X_5\}$
(variables)

Si se tienen las probabilidades para el caso de interés dejando al nodo 4 aparte (marginalizándolo):

$$\Theta_V = \overset{1}{\boxed{\Theta_2(X_1,X_2)}} \cdot \overset{2}{\boxed{\Theta_{2,3}(X_2,X_3) \cdot \Theta_{3,5}(X_3,X_5)}}$$

Se tiene que Θ_{12} se puede separar del producto $(\Theta_{2,3}) \cdot (\Theta_{3,5})$. Por lo tanto, sí es posible tener $\{X_3, X_5\}$ condicionalmente independiente

7. Para el mismo grafo g_3, identifiquen si $\{X_1, X_2\}$ son condicionalmente independientes a $\{X_5\}$ dado $\{X_3, X_4\}$

Con el grafo de 6) se tienen las probabilidades dadas por:

$$\Theta_V = \overset{1}{\boxed{\Theta_{12}(X_1,X_2)\,\Theta_{2,3,4}(X_2,X_3,X_4)}} \overset{2}{\boxed{\Theta_{4,5}(X_4,X_5) \cdot \Theta_{3,5}(X_3,X_5)}}$$

Se observa que se puede separar en 2 partes a través del conjunto $\{X_3, X_4\}$ y por lo tanto se puede decir que $\{X_1, X_2\}$ sí son condicionalmente independientes a $\{X_5\}$ dado $\{X_3, X_4\}$