# Sergio David Salazar Isairias – 201914381

Lucia Gómez Garzón - 201816285

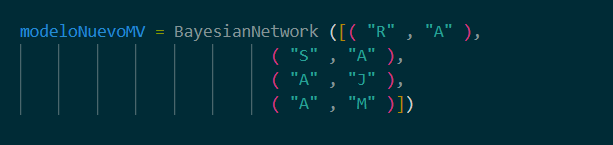
# Generando datos y estimando parámetros

1. Usando este modelo, use el módulo de muestreo de pgmpy para generar 100000 número aleatorios de acuerdo con la distribución definida por el grafo

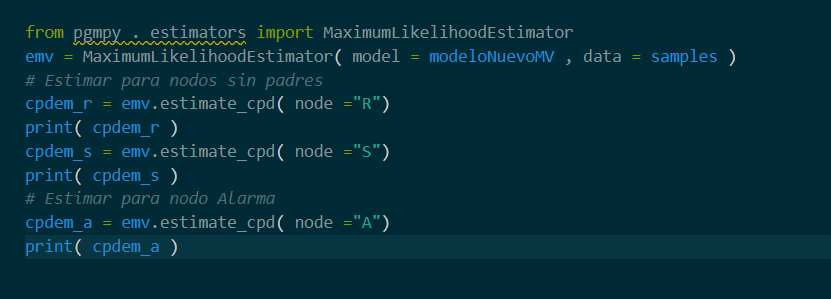
Imagen que contiene Texto

Descripción generada automáticamente

1. Cree un nuevo modelo con la misma **estructura** del inicial, pero **sin parámetros**.



1. Emplee el módulo de ajuste de pgmpy para ajustar algunas CPDs del nuevo modelo. Incluya el resultado en su **reporte**.



Tabla

Descripción generada automáticamente con confianza media

1. Emplee el método fit para ajustar todas las CPDs al tiempo y explore el resultado. Incluya el resultado en su **reporte**.

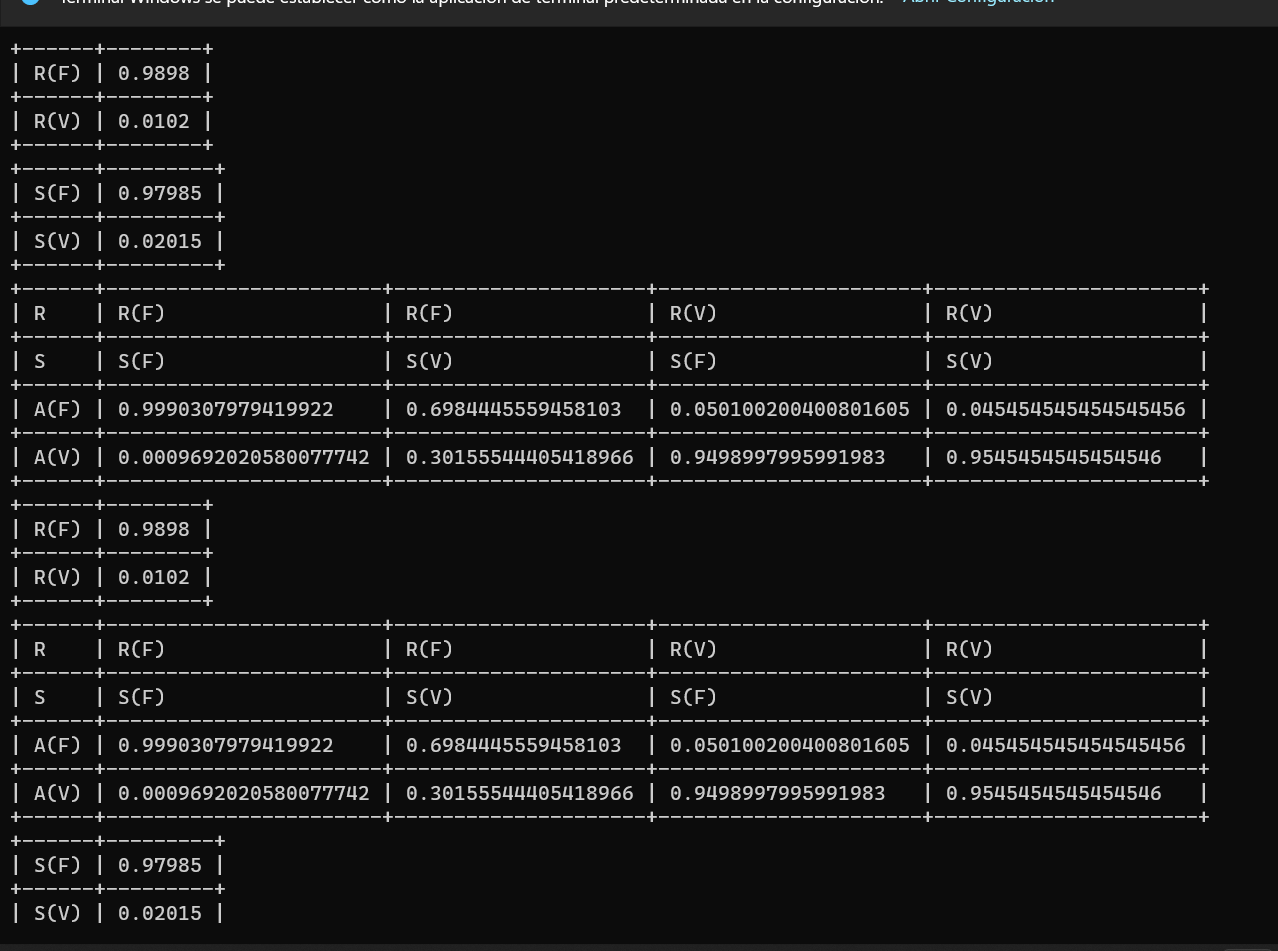


Imagen que contiene Texto

Descripción generada automáticamente

Al evaluar los estimadores de este medio podemos evidenciar que las probabilidades no son tan parejas, esto ocurre por la naturaleza del método de máxima verosimilitud que busca encontrar un parámetro que maximice la probabilidad según los datos (en este caso los aleatorios), de esta manera parece que el estimador de máxima verosimilitud que se toma es la probabilidad de que sea falso (cada nodo/variable), porque por esta razón las probabilidades para cada nodo con un estado falso tienen un porcentaje mayor que cuando es verdadero.

1. Realice nuevamente la estimación del modelo usando un prior

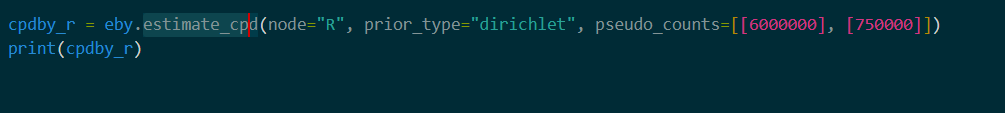
Texto

Descripción generada automáticamente

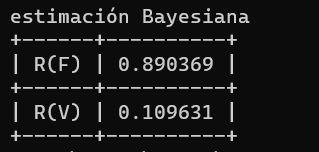
Se evidencia que la estimación del robo al ser la que se estima por este método bayesiano cambia con respecto al anterior, en cambios los demás datos efectivamente quedan muy similares. Por otro lado, se puede encontrar un mayor equilibrio en la estimación de las probabilidades de robo, ya que este método de estimación hace uso de pesos para cada estado posible en la variable que en este caso no son muy lejanos uno el otro.

8.Modifique los valores de pseudocounts e interprete los resultados. Incluya sus resultados y observaciones en su **reporte**.

En este caso se cambiaron los pseudocounts de esta manera:



Ahora al correrlo se espera que las probabilidades de Robo no sean tan balanceadas, sino que por el contrario una va a tener mayor peso que la otra:



Entonces efectivamente la probabilidad de que sea falso un robo termina teniendo mucho más peso que la probabilidad verdadera.

# Ahora con otra red y datos

Realice el proceso de estimación de la red en la grafica, empleando los datos del archivo adjunto. Estudie los parámetros obtenidos en su reporte.

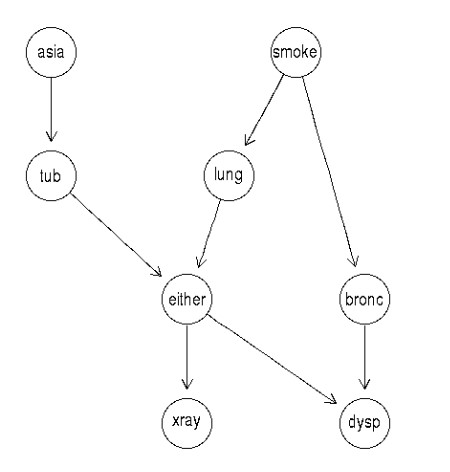
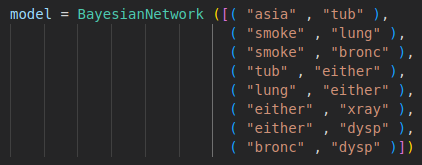
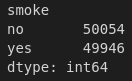
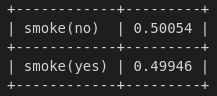


Figure 1: Grafo de una red Bayesiana para diagnosticar cáncer de pulmón

En primer lugar, se definió la red bayesiana.

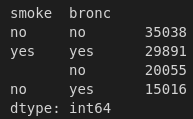
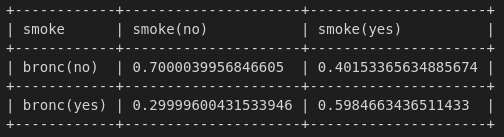


Luego, se utilizó el estimador de máxima verosimilitud junto a los datos data\_asia.csv para estimar las funciones de distribución condiciona de algunos nodos. En la siguiente imagen se encuentra la distribución de probabilidad estimada para la variable ‘smoke’.



En los datos la variable ‘smoke’ toma el valor de ‘no’ en 50.054 ocasiones y el valor de ‘si’ en 49.946 ocasiones. Ahora, si se calcula el porcentaje que representan dichos valores en los resultados totales, la respuesta es igual a la probabilidad estimada por máxima verosimilitud, lo cual concuerda con la expresión de los estimadores por máxima verosimilitud vista en clase.

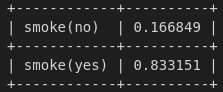
Asimismo, se utilizó el estimador para encontrar la función de probabilidad condicional de la variable ‘bronc’, la cual tiene un nodo padre (la variable ‘smoke’). El resultado se encuentra a continuación.



A partir de los datos se estimado que la probabilidad de que una persona no tenga bronquitis, dado que no fuma es de aproximadamente el 70%. De nuevo, para corroborar los resultados con la teoría, se identificó la cantidad de veces que en los datos la variable ‘bronc’ toma el valor de ‘no’ dado que la variable ‘smoke’ tomo el valor de ‘no’. Como resultado, el valor teórico de la probabilidad condicional es: 35.038/50.054, que es igual a la probabilidad obtenida por el estimador. Es decir, la probabilidad estimada coincide con el concepto de estimación por máxima verosimilitud.

**Estimador Bayesiano**

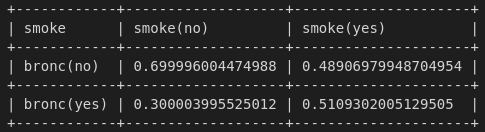
Luego, con los mismos datos se utilizó un estimador bayesiano con prior dirichlet para estimar las CPDs de los nodos del modelo. En particular, se estimó la distribución de probabilidad de la variable ‘smoke’ cuando se tiene 1 observación virtual del valor ‘no’ y 2e5 observaciones virtuales del valor ‘si’, el resultado es el siguiente.



Nótese que las probabilidades estimadas cambiaron. En particular, la probabilidad de que una persona sea fumadora aumento, lo cual se debe a que las observaciones del valor ‘si’ son mucho mayores que aquellas del valor ‘no’. Si se calcula el porcentaje de valores ‘no’ con relación al total, incluyendo los valores virtuales, el resultado es:

Es decir, la probabilidad estimada coincide con el valor teórico de las fórmulas de clase.

También se utilizó el estimador bayesiano para estimar la CPD de la variable ‘bronc’.



Si bien la distribución de probabilidad dado que la variable ‘smoke’ toma el valor de ‘no’ no cambio en gran medida, la distribución cuando ‘smoke’ es ‘si’ presento un cambio significativo. Con el estimador por máxima verosimilitud la probabilidad de que una persona no tenga bronquitis dado que si fuma es de aproximadamente 40%. Ahora, con el estimador bayesiano y el prior dirichlet dicha probabilidad aumento, de modo que la probabilidad de tener o no bronquitis es aproximadamente igual cuando la persona es fumadora.