Universidad de La Habana
Facultad de Matemática y Computación

Causalidad como complemento al procesamiento estadístico de los datos

Autor: **Antonio Jesús Otaño Barrera**

Tutor: **Dr. Luciano García Garrido**

Trabajo de Diploma
presentado en opción al título de
Licenciado en Ciencias de la Computación

Noviembre de 2021

Agradecimientos

Quisiera agradecer, por el apoyo que me han brindado de distintas formas a lo largo de toda la carrera, a las siguientes personas:

A mi madre, que no se rindió cuando yo sí lo hice.

A mi padre, por apoyarme, a su manera.

A mis tíos Rafael y Juana María y mi primo David, que siempre estuvieron ahí para lo que hiciera falta.

A José Carlos, Jonathan y Luis Angel. Primer año no hubiese sido lo mismo sin ustedes.

A Eziel, quien ha resultado ser un verdadero amigo.

A Gilberto, compañero de mil proyectos, y sobre todo, amigo. También agradezco a su madre, por todo el apoyo y atención que me brindó.

A mis profesores Wilfredo, Idania y Somoza, quienes me trasmitieron valiosas enseñanzas, no sólo de la materia que impartían.

A mi tutor Luciano García, por aceptarme como tesista y dedicarme parte de su tiempo y atención.

A todos con los que fui injusto y no mencioné pero que de una forma u otra me ayudaron a ser mejor persona.

Opinión del tutor

La causalidad ha sido desarrollada por los seres pensantes como una forma de explicar y por consiguiente conocer el mundo que lo rodea y actuar transformándolo. Por ello ha sido objeto de estudio a través de la historia del pensamiento humano, habiendo sido objeto de valiosas reflexiones fundamentalmente en el campo de la filosofía. Pero por muchos años continuó siendo considerada una modalidad cualitativa del pensamiento humano hasta que la llegada de la estadística y algunas deficiencias en el análisis cuantitativo de los datos puso en evidencia la necesidad de desarrollar la causalidad sobre bases cuantitativas como complemento de la correlación y demás procederes estadísticos.

El trabajo de diploma que presenta para su defensa el estudiante Antonio Jesús Otaño Barrera introduce, de manera innovadora, recientes resultados sobre la modelación cuantitativa de la causalidad utilizando específicamente modelos estructurales causales acompañados de una implementación con una interfaz amistosa con la cual es posible experimentar diferentes modalidades de la causalidad.

Es de señalar la disciplina y el rigor con el que el estudiante enfrentó la tarea sobre un tema que le era totalmente desconocido y como fue satisfaciendo todos los requisitos que le fueron planteados para su realización. Creemos que el estudiante Antonio Jesús Otaño Barrera ha alcanzado el nivel de profesionalidad que exige alcanzar el título de Lic. en Ciencia de la Computación y por tal motivo solicitamos la calificación de Excelente (5) para su trabajo de diploma.

La Habana, 20 de noviembre de 2021.

Dr. Luciano García Garrido
Profesor Titular Consultante
Facultad de Matemática y Computación
Universidad de La Habana, Cuba

Resumen

La estadística tradicional se ocupa principalmente de la asociación entre variables pero no revela información acerca de las relaciones de causalidad entre estas. A partir de estas limitaciones es necesario la construcción de una teoría que formalice matemáticamente los procesos causales. Dicha teoría permitirá modelar situaciones de la vida cotidiana y responder preguntas causales acerca de estas. Un objetivo más ambicioso consiste en la simulación del razonamiento causal humano, el cual se cree que debe ser una pieza fundamental en la construcción de una inteligencia artificial fuerte. En el presente trabajo se presentarán los principales desarrollos recientes de teorías matemático computacionales de la causalidad, con especial énfasis en el trabajo de Judea Pearl, y se proveerá una implementación de un modelo causal capaz de responder distintos tipos de preguntas causales. Por último, se exponen posibles escenarios donde se puede utilizar el programa propuesto.

Abstract

Traditional statistics deals mainly with the association between variables but does not reveal information about the causal relationships between them. Based on these limitations, it is necessary to build a theory that mathematically formalizes causal processes. This theory will allow modeling everyday life situations and answering causal questions about them. A more ambitious goal is the simulation of human causal reasoning, which is believed to be a fundamental piece in building strong artificial intelligence. In this paper, the main recent developments of computational mathematical theories of causality will be presented, with special emphasis on the work of Judea Pearl, and an implementation of a causal model capable of answering different types of causal questions will be provided. Finally, possible scenarios are exposed where the proposed program can be used.

Índice general

Índice de figuras

Índice de tablas

Introducción

Desde tiempos de antaño el hombre se ha visto en la necesidad imperiosa de manejar de la manera más adecuada posible sus recursos, no solo para conseguir prolongar la vida de estos, sino además para contar con una aceptada utilización de los mismos. Uno de los principales problemas que aqueja a la sociedad actual es la dificultad que se presenta para garantizar un óptimo manejo de nuestro *tiempo*.

Hace unos 70000 años, nuestra especie, *Homo Sapiens*, experimentó un salto evolutivo que le dotó de capacidades para socializarse y modificar su entorno como ninguna otra sobre la Tierra. Algunos autores llaman a este salto la Revolución Cognitiva. En su libro *Sapiens: Una breve historia de la humanidad* , el historiador Yuval Noah Harari plantea que este salto consistió en la adquisición de la capacidad de imaginar entidades ficticias, lo cual permitió que grandes cantidades de desconocidos que compartían las mismas creencias cooperaran entre sí y conquistaran el planeta.[1]

La capacidad de imaginar está estrechamente vinculada con el pensamiento causal. Al imaginar las consecuencias de una acción con anticipación, aun sin llegar a ejecutarla, estamos estimando el efecto de una *intervención*. Por otra parte, formular preguntas acerca de mundos alternativos a partir de experiencias observadas en el real, se conoce como *contrafactual*. Antes de la Revolución Cognitiva, nuestra especie se limitaba a observar el mundo, infiriendo hechos a partir de las experiencias vividas. Sin embargo, la adquisición de estas nuevas habilidades transformó a *Homo Sapiens* en un razonador causal. Hasta ahora el único ser vivo que se conoce que tiene esta capacidad es nuestra especie y por tanto podría ser un factor clave si queremos simular la inteligencia humana.

En la actualidad, los algoritmos de IA se basan principalmente en el reconocimiento de patrones. Para ello necesitan ser entrenados con grandes cantidades de datos. Las increíbles habilidades que han demostrado estos algoritmos en algunas tareas ha llevado a pensar a algunos que la inteligencia humana se reduce al mero procesamiento de datos. Sin embargo, es evidente que la IA se encuentra todavía lejos de igualar a la inteligencia humana en muchos aspectos. Algunos autores, entre ellos Judea Pearl, plantean que la pieza faltante para lograr construir una verdadera inteligencia artificial consiste en dotar a las máquinas de razonamiento causal.[2]

A partir de la afirmación de Pearl, sería oportuno realizarse algunas preguntas:

- ¿Qué ventajas posee dotar de razonamiento causal a un programa de inteligencia artificial ?
- ¿Cuáles son los principales desarrollos alacanzados hasta la fecha en el área científica de la causalidad ?
- ¿Cuales son las aplicaciones que tiene la causalidad en la actualidad ?

El presente trabajo pretende arrojar algunas luces en torno a estas cuestiones. En particular, se demostrará cómo, con los desarrollos alcanzados en esta área, se pueden desarrollar herramientas que basándose en la inferencia causal permitan resolver problemas prácticos en la actualidad.

A continuación se recorrerán los momentos más importantes de la historia de la causalidad en la filosofía y la ciencia. Se verá como esta fue excluida de la ciencia por los estadísticos, y los problemas que surgieron a partir de esta exclusión. Luego se resaltarán los principales exponentes de la teoría de la causalidad junto con sus más notables avances.

0.1. Un poco de historia

Definir la causalidad fue una tarea que comenzó en manos de los filósofos. Ya en la antigua Grecia, Aristóteles atribuyó a la palabra causa varios significados [3]. En particular, la causa eficiente o causa de movimiento se define como "la fuente del primer comienzo de cambio o descanso"y fue adoptada por muchos filósofos posteriores como Nicolás Maquiavelo y Francis Bacon.

A principios de la época moderna, David Hume concebía la causalidad como un hecho psicológico, producto de nuestra experiencia. Entre la causa y el efecto existe una conexión necesaria que condiciona al efecto a sólo ocurrir después de la causa [4]. Las ideas de Hume influenciaron a muchos de los posteriores filósofos y matemáticos que trabajaron en el área de la causalidad. Hume fue uno de los defensores del análisis de regularidad para explicar la causalidad. Otro partidario de este método fue John Stuart Mill (1806-1873).

0.1.1. La expulsión de la causalidad de la estadística y sus efectos

La historia de la causalidad desde el punto de matemático comienza en el siglo XIX. Francis Galton (1822-1911) en sus investigaciones acerca de la estabilidad de la dotación genética de las poblaciones descubrió el fenómeno de regresión a la media [5]. Inicialmente, trató de encontrarle una explicación causal, pero en el proceso descubrió la correlación, abriendo el camino para una nueva rama de la matemática: la estadística. Eventualmente, Galton se distanció de la causalidad y centró sus esfuerzos en esta otra rama.

Karl Pearson, alumno de Galton, continuó con la tarea de excluir a la causalidad no solo de la estadística, sino de la ciencia misma. Para Pearson, la causalidad era prescindible para la ciencia. Argumentaba que las respuestas a

todas las cuestiones se encontraban en los datos y que, consecuentemente, la correlación era un descriptor mucho más preciso de los procesos de la naturaleza que la causación. Parecía que la causalidad iba a ser sepultada completamente.

Sin embargo, Pearson mismo no pudo desprenderse completamente de la causalidad. Escribió varios artículos junto a su asistente George Udny Yule(1871-1951) acerca de la "correlación espúrea ", un término que no es posible explicar sin hacer uso de la causalidad. Pearson se dio cuenta de que era relativamente fácil de encontrar ejemplos en los que la correlación carecía de sentido.

Un caso típico de correlación espúrea es la ocurrencia de *variables de confusión*(confounding), que ocurre cuando existe correlación entre dos variables pero ninguna es causa de la otra, sino que ambas son influenciadas a la vez por terceras variables. Un ejemplo famoso de este tipo consiste en que existe una fuerte correlación entre el consumo de chocolate per cápita en una nación y su número de ganadores del Premio Nobel. En respuesta a esta situación uno puede argumentar que el consumo de chocolate era abundante en los países de Occidente, donde también eran elegidos preferentemente los premios Nobel. Pero esta es una explicación causal, la cual, según Pearson, era irrelevante para el razonamiento científico.

Otro ejemplo de correlación espúrea fue la encontrada por Yule al detectar una fuerte correlación entre la tasa de mortalidad de Inglaterra en un año y el porciento de matrimonios contraídos en ese mismo año en la Iglesia de Inglaterra. Lo que ocurrió fue que dos tendencias históricas coincidieron: el decrecimiento de la tasa de mortalidad y el de la membresía a la Iglesia de Inglaterra. Al disminuir ambos a la vez, existía una correlación positiva entre ambos, pero no una conexión causal.

Pearson descubrió en 1889 uno de los tipos más famosos de correlación espúrea, que ocurre cuando dos poblaciones heterogéneas son mezcladas en una. Obtuvo medidas de la longitud y el ancho del cráneo de 806 hombres y 340 mujeres de las Catacumbas de París. Primero computó la correlación entre el ancho y la longitud en la población de los cráneos de hombres y luego hizo con la de las mujeres. En ambos casos no obtuvo un valor de correlación significativo. Sin embargo, al mezclar ambas poblaciones sí obtuvo una correlación significativa. Esto tiene sentido, porque un cráneo de longitud pequeña tiene más probabilidades de pertenecer a una mujer y que a su vez el ancho de este sea pequeño. Sin embargo, Pearson lo atribuyó a haber mezclado incorrectamente las poblaciones. Este ejemplo es un caso de un fenómeno más general conocido como la paradoja de Simpson.[2]

0.1.2. Sewall Wright y los diagramas de caminos

Un punto de inflexión en la historia de la causalidad lo marcó el genetista Sewall Wright (1889-1988) cuando en 1920 publicó el artículo *The relative importance of heredity and enviroment in determining the piebald pattern of guinea-pigs* [6]. El objetivo del trabajo de Wright por aquel entonces era determinar la causa de las variaciones en los patrones del color del pelo en los conejillos de indias. Wright dudaba que las variaciones genéticas fueran la úni-

ca causa de este comportamiento y postuló que los factores de desarrollo en el útero de la madre y los factores ambientales eran también causantes de algunas variaciones. Al efecto que ejercía cada una de estas variables se le llama cantidades causales. Wright desarrolló entonces un modelo, conocido como diagrama de caminos (path diagram), que permitió calcular el valor de dichas cantidades. A partir de correlaciones medidas en los datos recopilados, se podía determinar el valor de las cantidades causales mediante ecuaciones algebraicas. Al final, Wright demostró que los factores de desarrollo erán mas importantes que los hereditarios.

El diagrama de caminos es el primer modelo causal que se haya publicado. Estableció por primera vez un puente entre las probabilidades y la causalidad. A pesar de que no se necesita conocer todas las relaciones causales entre las variables de interés para sacar conclusiones causales, Wright deja un punto bien claro: no es posible llegar a conclusiones causales sin previamente establecer hipótesis causales. Esto implica que es imposible responder a una pregunta causal únicamente a partir de datos, lo cual supone una ruptura con los argumentos de los estadísticos de la época.

Obviamente las hipótesis causales asumidas pueden ser incorrectas. En ese caso Wright argumentó que se podía usar su modelo en "modo exploratorio". Si los resultados que arrojaba resultaban contradictorios con los datos, entonces las hipótesis asumidas eran incorrectas y el modelo debía ser corregido.

Contrario a lo que muchos piensan, el análisis causal no pretende demostrar que una variable causa a otra, o encontrar la causa de una variable. Ese es el terreno del descubrimiento causal. Wright comprendió desde el principio que el descubrimiento causal era mucho más difícil y quizá hasta imposible. En contraste, el análisis causal pretende representar el conocimiento causal en un lenguaje matemático, combinarlo con los datos, y responder a preguntas causales de valor práctico.[2]

0.1.3. Modelos de causalidad más relevantes

Hans Reichenbach introdujo el *Principio de la Causa Común*(RCCP), el cual fue incorporado a una teoría probabilística de la causalidad. Este principio es significativo porque establece una conexión entre la estructura causal y la correlación probabilística, facilitando la inferencia causal a partir de correlaciones observadas. A menudo es visto como un antecedente de la *Condición Causal de Markov*, la cual juega un papel fundamental en los modelos causales y en la inferencia causal. Sin embargo, el RCCP ha sido controvertido, siendo propuestos varios contra-ejemplos. Estos no refutan el principio completamente, sino que dan lugar a debates sobre su alcance e interpretación.[7]

Patrick Suppes se basó en las ideas de Hume para desarrollar su propia teoría de la causalidad. En su libro *A probabilistic theory of causality*(1970)[8], Suppes plantea que el carácter de la causalidad en el lenguaje ordinario no es propiamente determinista, y que la principal debilidad en el análisis de Hume fue la omisión de consideraciones probabilísticas. Así mismo, da a conocer su propia definición de causalidad: un evento es la causa de otro si la aparición

del primero es seguido con una alta probabilidad por la aparición del segundo, y no hay un tercer evento que se pueda utilizar para determinar la relación de probabilidad entre el primer y el segundo evento. Suppes propone entonces un framework probabilístico para el análisis de las relaciones causales. Presenta varias definiciones para los diferentes tipos de causas en un intento de distinguir las causas genuinas de las espúreas y las causas directas de las indirectas.

Sin embargo, su trabajo no estuvo exento de críticas. En el artículo *A critique of Suppes' theory of probabilistic causality* de Richard Otte, se ofrecen un conjunto de contra-ejemplos que muestran que la teoría de Suppes es defectuosa. En particular, destaca que las definiciones de Suppes fallan al diferenciar las causas genuinas de las espúreas, y las causas directas de las indirectas. Otte argumenta que no es posible diferenciar correctamente dichas causas utilizando únicamente relaciones probabilísticas, por lo que cualquier modificación menor en las definiciones de Suppes no logrará resolver estas dificultades.[9]

El test de causalidad de Granger fue propuesto en 1969 por Clive Granger para determinar cuando una serie de tiempo es útil para predecir otra [10]. Granger argumenta que en términos económicos la causalidad puede ser probada midiendo la capacidad de una serie de tiempo de predecir los valores futuros de otra serie de tiempo. En todo caso, los econometristas afirman que el test de Granger determina solo "causalidad predictiva". Granger enfatizó además que algunos estudios realizados en áreas fuera de la economía donde se utilizó el test de Granger arrojaron resultados ridículos. Esto se debe a que la definición de causalidad de Granger es demasiado restrictiva. El test de causalidad de Granger está diseñado para manejar relaciones entre dos variables, y puede arrojar resultados erróneos cuando las verdaderas relaciones de causalidad involucran tres o más variables. Además no puede capturar relaciones de causalidad no lineales. Sin embargo, se mantiene como un popular método de análisis causal en series de tiempo dada su simplicidad computacional. [11].

Los desarrollos de Judea Pearl

Judea Pearl propone una jerarquía para clasificar los distintos niveles de inferencia causal, a la cual denomina "la escalera de la causalidad". Los tres niveles corresponden a las capacidades cognitivas que abarcan el razonamiento causal: ver, hacer e imaginar.

El primer nivel pertenece a la asociación. La pregunta que lo caracteriza es: ¿Qué pasa si veo...?. Por ejemplo, un doctor luego de examinar a un paciente podría formularse la pregunta: ¿Dados los síntomas presentes, cuán probable es que el paciente posea una variante de la gripe?. La expresión $P(A \mid B)$ indica cuán probable es que el suceso A ocurra dado que se sabe que el suceso B ocurrió. La introducción de la Fórmula de la Probabilidad Total y la Fórmula de Bayes dan lugar al surgimiento del enfoque bayesiano de la Estadística en contraste con el enfoque frecuentista. La estadística bayesiana en conjunto con los modelos gráficos probabilistas sientan las bases para la modelación probabilista de la causalidad.

El segundo nivel está determinado por las intervenciones. Preguntas repre-

sentativas son: ¿Qué pasa si hacemos... ? o ¿Cómo?. A diferencia del primer nivel, se formulan preguntas en donde los hechos son forzados a ocurrir en vez de simplemente ser observados. Una pregunta de este nivel podría ser, por ejemplo: ¿Qué ocurriría si aumentamos el precio del producto X ? La expresión $P(A \mid do(B))$ nos dice cuán probable es la ocurrencia de A dado que forzamos el suceso B a ocurrir. La prueba aleatoria controlada (en inglés, *Randomized Controlled Trial*, RCT) ha sido la herramienta tradicional en la estadística para determinar el valor de esta expresión. Sin embargo, este método presenta sus limitantes al poder ser muy costoso en algunos casos o ser imposible de realizar por cuestiones éticas. Uno de los avances en el terreno de la causalidad ha sido precisamente el poder computar el efecto de una intervención sin tener que realizarla.

El tercer y último nivel corresponde a los contrafactuales. Preguntas representativas son: ¿Qué pasaría si en vez de...? o ¿Por qué ? Responder a estas preguntas requiere imaginar mundos alternativos y compararlos con el mundo real. Es en este nivel donde se formulan las preguntas más complejas e interesantes. Por ejemplo, luego de tomar un medicamento y aliviar su malestar, un paciente podría preguntarse: ¿Debo mi recuperación a haber tomado el medicamento ? [2]

Judea Pearl propone en los años 80 las redes bayesianas [12], un modelo que opera en el nivel 1 en su versión original. Estas se apoyan en la independencia condicional para construir una representación compacta de las relaciones entre las variables. A partir de un conjunto de evidencias, permiten estimar la probabilidad conjunta de un conjunto de variables de la red.

Los desarrollos posteriores de Pearl se adentran en abarcar los otros dos niveles de la escalera de causalidad. El modelo causal estructural es uno de sus resultados más importantes [13]. Este contiene los desarrollos teóricos necesarios para computar tanto intervenciones como contrafactuales. Precisamente en este modelo, conjuntamente con las redes bayesianas, se basa el marco teórico del presente trabajo.

0.2. Objetivos

El presente trabajo tiene como objetivo general el diseño y programación de una aplicación que permita su utilización en la solución de diversos problemas causales relevantes de un modelo estructural causal. La aplicación debe facilitar a usuarios no expertos en programación su empleo en la solución de diversos problemas de causalidad que surgen en la investigación científica. A su vez, se pueden identificar varios objetivos específicos:

1. Resumir los aspectos teóricos fundamentales acerca de la teoría de grafos y la teoría de las probabilidades, que serán la base de los desarrollos teóricos que se expondrán durante la tesis. Hacer especial énfasis en el concepto de independencia condicional y sus principales propiedades y resultados derivados.

2. Desarrollo de la teoría de las redes bayesianas como antecedente del modelo causal estructural de Pearl. Descripción de los algoritmos de inferencia en redes bayesianas.

3. Desarrollo del modelo causal estructural de Pearl. Exposición de los diferentes criterios y algoritmos para calcular intervenciones y contrafactuales. Desarrollo de los conceptos de mediación y atribución y sus principales aplicaciones.

4. Proponer algoritmos que haciendo uso de las redes bayesianas resuelvan el problema de la inferencia causal en un modelo causal estructural.

5. Diseño e implementación del programa. Este debe permitir la creación de modelos causales estructurales, en los que poder ejecutar algoritmos de inferencia causal que permitan responder a preguntas causales correspondientes a predicciones, intervenciones, contrafactuales, atribución y análisis de mediación.

6. Experimentación. Exposición de un problema práctico en el que se muestren las características del software desarrollado.

El presente trabajo se estructura como se detalla a continuación: En el capítulo 2 presentará toda la literatura consultada que contiene los resultados teóricos necesarios para la construcción del software. En el capítulo 3 se presenta la propuesta de software para resolver problemas de inferencia causal, mostrando las funcionalidades que ofrece así como los detalles técnicos de la implementación. Por último, en el capítulo 4 se muestra una aplicación práctica del programa en la que se utilizará un modelo causal para responder a preguntas causales acerca de la COVID 19.

Capítulo 1

Estado del Arte

El tema del manejo, creación y gestión de un sistema de horarios ha sido abordado por un gran número de personas, en muchos ámbitos diferentes.

1.1. Herramientas de gestión

El ejemplo más concreto que se analizó antes de la creación del presente trabajo fue un sistema desarrollado en 2019 entre un grupo de universidades de América del Norte y Europa: UniTime.

Este sistema ofrece soporte en un gran número de escenarios, dígase por ejemplo la creación de una especie de salas para la planificación de eventos así como el manejo de plnificaciones individuales por estudiantes. Además cuenta con una pequña comunidad que brinda soporte al mismo además de que se le han realizado diversas versiones y modificaciones periódicamente.

La principal cuestión que motivó el desarrollo de un sistema propio de la universidad y la no utilización de estas herramientas que ya existían fue la posibilidad de ofrecer un manejo de restricciones sobre todas las entidades del horario; restricciones que ofrecen un grado de *felicidad* al ser evaluadas y permiten la apreciación por parte del creador del horario de que tan bueno resulta la distribución brindada a los turnos de clases; que viene siendo, en definitiva, el punto central de todas estas herramientas.

Otros autores manejan en sus sistemas un concepto similar al de *restricción*, pero estas están en todos los casos relacionadas con un profesor y un turno de clase; en cambio en el presente software se permite asociarlas a cualquier entidad definida dentro del sistema, dígase por ejemplo: *local*, *departamento*.

1.2. Restricciones

La idea detrás del manejo e implementación de las restricciones surge a través del trabajo de diploma desarrollado por el estudiante *Joel Rey Travieso Sosa* en el año 2012.

Se manejan varios tipos principales de restricciones:

- Restricción de requerimiento de cuenta simple.
- Restricción de requerimiento de cuenta de condiciones.
- Restricción de requerimiento de distribución de atributos.
- Restricción de requerimiento relacional.

En los capítulos siguientes se ofrecerá una explicación más detallada de todo los relacionado con estas condiciones impuestas sobre el sistema así como la adecuada definición y formulación de todos los conceptos antes expuestos.

1.3. Tecnologías

Para la confección del sistema se analizaron diversos escenarios para considerar el que propiciara la obtención de un mejor producto y a la vez que se contara con los conocimientos suficientes en el área para obtener los mejores resultados.

El sistema está desarrollado en *NestJS*,

1.4. Definiciones previas

Los modelos que se estudiarán en esta tesis tendrán como estructura subyacente un grafo dirigido y acíclico (DAG) 1.1.

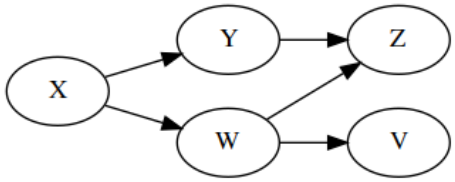


Figura 1.1: Grafo dirigido y acíclico(DAG)

Denotaremos las variables aleatorias con letras mayúsculas (X, Y, Z). A su vez los valores que estas pueden tomar se denotarán con letras minúsculas (x, y, z). Al conjunto de valores que puede tomar una variable X se le denotará $Val(X)$. Para referirnos a un conjunto de variables aleatorias se utilizarán letras mayúsculas en negritas ($\mathbf{X}, \mathbf{Y}$) y para denotar asignaciones a estos conjuntos se emplearán letras minúsculas en negritas ($\mathbf{x}$, $\mathbf{y}$). Por último, el conjunto de todas las asignaciones que puede darse a un conjunto de variables aleatorias $\mathbf{X}$ se denota como $Val(\mathbf{X})$.

De especial interés es la Fórmula de Bayes, que permite obtener el valor de la probabilidad condicional $P(X \mid Y)$ a partir del conocimiento previo de $P(Y \mid X)$:

Teorema 1.4.1 (Fórmula de Bayes) *Sean X, Y dos variables aleatorias. Entonces la probabilidad condicional $P(X \mid Y)$ se puede calcular como:*

$$P(X \mid Y) = \frac{P(Y \mid X) \cdot P(X)}{P(Y)}$$

La Ley de Probabilidad Total expresa la distribución de una variable en base a la probabilidad condicional respecto a otra variable:

Teorema 1.4.2 (Ley de Probabilidad Total)

$$P(X = x) = \sum_{y} P(X = x \mid Y = y)P(Y = y)$$

Esta también puede ser aplicada al caso de la probabilidad condicional:

$$P(X = x \mid Y = y) = \sum_{z} P(X = x \mid Y = y, Z = z)P(Z = z \mid Y = y)$$

La regla de la cadena define la probabilidad conjunta de un conjunto de variables en términos de la probabilidad condicional:

Teorema 1.4.3 (Regla de la cadena) *Sean $X_1, X_2, ..., X_n$ un conjunto de variables aleatorias. Entonces se cumple que:*

$$P(X_1, X_2, ..., X_n) = P(X_1 \mid X_2, ..., X_n)P(X_2 \mid X_3, ..., X_n)...P(X_{n-1} \mid X_n)P(X_n)$$

1.5. Independencia

La teoría de la independencia es fundamental en los modelos gráficos probabilistas. Estos son construidos a partir de las relaciones de independencia que se establecen entre las variables.

Definición 1.5.1 (Independencia) *Sean las variables aleatorias X, Y. Se dice que X es independiente (marginalmente)(incondicionalmente) de Y y se denota $X \perp Y$ si $P(X \mid Y) = P(X)$ o si $P(Y) = 0$*

Una definición alternativa podría ser:

Proposición 1.5.1 $X \perp Y \iff P(X,Y) = P(X)P(Y)$

En ocasiones dos variables no son independientes por sí solas pero sí lo son a través de una tercera variable. El concepto de independencia condicional recoge estos casos.

Definición 1.5.2 (Independencia condicional) *Sean X,Y,Z variables aleatorias. Se dice que X es condicionalmente independiente de Y dado Z y se denota $(X \perp Y \mid Z)$ si $P(X \mid Y, Z) = P(X \mid Z)$ o si $P(Y, Z) = 0, \forall x \in Val(X)$.*

Similarmente al caso de la independencia marginal, se puede brindar una definición alternativa para la independencia condicional:

Proposición 1.5.2 $(X \perp Y \mid Z) \Longleftrightarrow P(X,Y \mid Z) = P(X \mid Z)P(Y \mid Z)$

La independencia condicional satisface un conjunto de propiedades que resultan útiles:

- **Simetría**: $(X \perp Y \mid Z) \Longrightarrow (Y \perp X \mid Z)$
- **Descomposición**: $(X \perp Y,W \mid Z) \Longrightarrow (X \perp Y \mid Z)$
- **Unión débil**: $(X \perp Y,W \mid Z) \Longrightarrow (X \perp Y \mid Z,W)$
- **Contracción**: $(X \perp W \mid Z,Y)\&(X \perp Y \mid Z) \Longrightarrow (X \perp Y,W \mid Z)$

Definición 1.5.3 *Una distribución P sobre una variable X se dice que es positiva si para todo $x \in Val(X)$ se cumple que $P(X = x) > 0$*

- **Intersección**: Si $P(X)$, $P(Y)$, $P(W)$ y $P(Z)$ son positivas:

$$(X \perp Y \mid Z,W)\ \&\ (X \perp W \mid Z,Y) \Longrightarrow (X \perp Y,W \mid Z)$$

Los conceptos y propiedades vistos hasta el momento pueden extenderse a conjuntos de variables aleatorias.

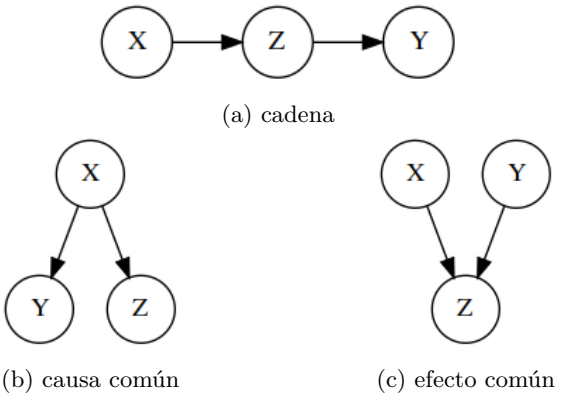
1.5.1. Independencia condicional en modelos gráficos probabilistas

Existen tres tipos de conexiones en los modelos gráficos probabilistas que juegan un papel importante en el análisis de la independencia condicional. Sean X, Y, Z tres variables cualesquiera. Una *cadena* consiste en una conexión del tipo $X \rightarrow Y \rightarrow Z$(Figura 1.2a). Una *causa común* consiste en una conexión del tipo $Y \leftarrow X \rightarrow Z$(Figura 1.2b). Por último, un *efecto común* es una conexión del tipo $X \rightarrow Z \leftarrow Y$(Figura 1.2c).

A partir de estas definiciones es posible establecer tres reglas para la determinación de las independencias condicionales en un modelo gráfico probabilista:

Regla 1.5.1 (Independencia condicional en cadenas) *Dos variables X y Y son condicionalmente independientes dado $\boldsymbol{Z}$, si existe un solo camino unidireccional entre X y Y, y $\boldsymbol{Z}$ es cualquier conjunto de variables que intercepta ese camino.*

Regla 1.5.2 (Independencia condicional en causas comunes) *Si una variable X es un ancestro común de las variables Y y Z, y hay solo un camino entre Y y Z, entonces Y y Z son condicionalmente independientes dado X.*



(a) cadena

(b) causa común

(c) efecto común

Figura 1.2: Tipos de conexiones en un modelo gráfico probabilista

Regla 1.5.3 (Independencia condicional en efectos comunes) *Si una variable Z es descendiente de dos variables X y Y, y existe un solo camino entre X y Y, entonces X y Y son marginalmente independientes pero condicionalmente dependientes dado Z o cualquiera de los descendientes de Z*

En la figura 1.3a si tomamos $\mathbf{Z} = \{V, Z\}$ y aplicamos la Regla 1.5.1, se cumple que $(X \perp Y \mid \mathbf{Z})$. A su vez, en la figura 1.3b, donde existe la causa común $Y \leftarrow X \rightarrow W$ si conocemos el valor de X entonces se cumple que $(Y \perp Z \mid \mathbf{X})$. Por último, en la figura 1.3c, tenemos el efecto común $X \rightarrow Z \leftarrow Y$, por tanto las variables X y Y son marginalmente independientes pero condicionalmente dependientes dado Z o W.

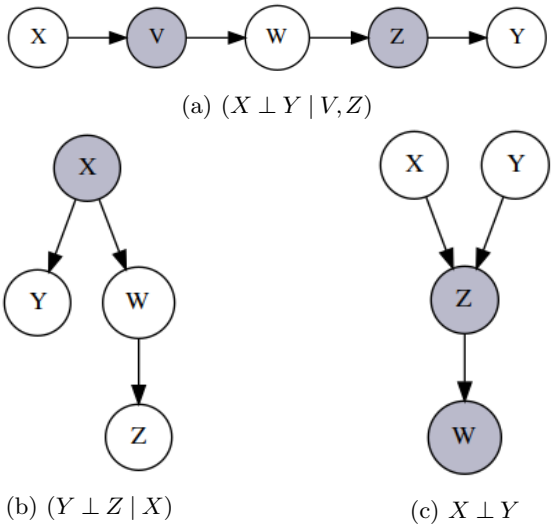
Definición 1.5.4 *Un camino p está bloqueado por un conjunto de nodos $\boldsymbol{Z}$ si y solo si:*

1. *p contiene una cadena de nodos $A \rightarrow B \rightarrow C$ o una causa común $A \leftarrow B \rightarrow C$ tal que $B \in \boldsymbol{Z}$.*
2. *p contiene un efecto común $A \rightarrow B \leftarrow C$ tal que ni B ni ninguno de sus descendientes pertenece a $\boldsymbol{Z}$.*

Definición 1.5.5 (d-separación) *Si Z bloquea todo camino entre X y Y entonces X y Y están d-separados condicionado a $\boldsymbol{Z}$, y por tanto independientes condicionado a Z. En caso contrario se dice que están d-conectados.*

En la figura 1.4, se cumple que A esta d-separado de F condicionado a $\mathbf{Z} = \{B, C\}$:

- El camino $A \rightarrow B \rightarrow F$ es una cadena y $B \in \mathbf{Z}$, por tanto está bloqueado.
- El camino $A \leftarrow C \rightarrow F$ es una causa común y $C \in \mathbf{Z}$, por tanto está bloqueado.



(a) $(X \perp Y \mid V, Z)$

(b) $(Y \perp Z \mid X)$

(c) $X \perp Y$

Figura 1.3: Relaciones de independencia en modelos gráficos probabilistas.

- El camino $A \rightarrow D \leftarrow F$ es un efecto común y no contiene ningún nodo que pertenezca a $\mathbf{B}$, por lo que también está bloqueado.

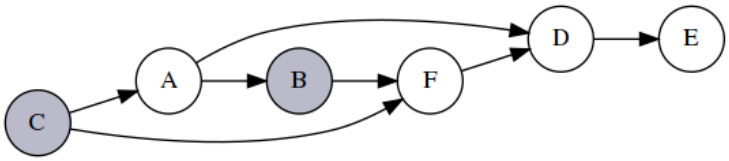


Figura 1.4: Ejemplo de d-separación

1.6. Redes bayesianas

El modelo probabilista que se deriva directamente de la teoría es la distribución de probabilidad conjunta (DPC). Sin embargo, el uso de la DPC conlleva un costo computacional que lo hace impracticable para los problemas reales. Por poner un ejemplo, si se dispone de n variables binarias, la DPC tendría un total de 2^n entradas. A partir de estas limitaciones se hace necesario encontrar un modelo alternativo que represente de manera compacta la DPC.

Definición 1.6.1 (Red bayesiana) *Una red bayesiana es un par ordenado* $\langle G, P \rangle$ *donde:*

- $G = \langle V, E \rangle$ *es un grafo acíclico y dirigido donde* V *representa a un conjunto de variables aleatorias* $X = \{X_1, X_2, ..., X_n\}$ *y* E *determina las relaciones*

de dependencia entre dichas variables. Sea Pa_{X_i} el conjunto de los nodos padres de X_i en G y $NoDesc_{X_i}$ el conjunto de los nodos que no descienden de X_i en G. Entonces para toda variable X_i de G se cumple que:

$$(X_i \perp NoDesc_{X_i} \mid Pa_{X_i})$$

Esta propiedad es llamada la condición local de Markov.

- *P es una distribución de probabilidad sobre las variables $X_1, X_2, ..., X_n$ que factoriza sobre G:*

$$P(X_1, X_2, ..., X_n) = \prod_{i=1}^{n} P(X_i \mid Pa_{X_i})$$

Las redes bayesianas [12] utilizan las relaciones de independencia condicional entre las variables para obtener una representación compacta de la DPC.

Consideremos el caso de una red bayesiana que se usa para modelar el comportamiento de la enfermedad en un paciente (figura 1.5). Las variables tenidas en cuenta son: la temperatura del paciente(T), si este presenta dolor de cabeza(D), la presencia de la enfermedad en el paciente(E), la edad (X) y el grado de gravedad de la enfermedad(G).

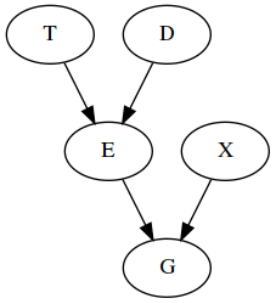


Figura 1.5: Una red bayesiana que modela el comportamiento de la enfermedad en un paciente

Cada nodo X_i debe almacenar la distribución de probabilidad $P(X_i \mid Pa_{X_i})$. Por ejemplo, si T toma los valores t_1 y t_0 que indican si tiene o no fiebre, D toma los valores d_1 y d_0 que indican si tiene o no dolor de cabeza y E toma los valores e_1 y e_0 que indican si el paciente tiene o no la enfermedad, entonces una posible distribución de probabilidad para $P(E \mid T, D)$ podría ser:

	e_0	e_1
t_0, d_0	0.85	0.15
t_0, d_1	0.7	0.3
t_1, d_0	0.4	0.6
t_1, d_1	0.2	0.8

1.6.1. Inferencia

Existen dos tipos principales de inferencia en redes bayesianas: la *actualización de creencias*, también llamada *inferencia probabilística* y la *revisión de creencias*, también llamada *explicación MAP*[14] [15]. La *actualización de creencias* consiste en determinar la distribución conjunta de un conjunto $\mathbf{X}$ de variables de la red a partir de un conjunto de evidencias $\mathbf{E}$, determinada por la expresión $P(\mathbf{X} \mid \mathbf{E})$. La *revisión de creencias* consiste en obtener la configuración más probable de las variables en un conjunto $\mathbf{X}$ de variables de la red dado un conjunto de observaciones $\mathbf{E}$. Es decir, se desea obtener el valor de la expresión: $arg\ max_{\mathbf{X}}\{P(\mathbf{X} \mid \mathbf{E} = \mathbf{e})\}$, que indica una asigación $\{X_1 = x_1, X_2 = x_2, ..., X_n = x_n\}$ tal que no exista ninguna otra asignación con mayor probabilidad.

A su vez los algoritmos de inferencia se clasifican en dos grandes categorías: inferencia exacta e inferencia aproximada. La inferencia probabilística exacta en general ha sido catalogada como un problema NP-duro por Cooper [16]. La inferencia probabilística aproximada también fue clasificada como NP-duro por Dagum y Luby [17]. En 1994, Shimony demostró que la revisión de creencias exacta era NP-duro [18] y posteriormente en 1998 Abdelbar y Hedetniemi demostraron que la revisón de creencias aproximada también era NP-duro [19].

Inferencia exacta

En los 80, Pearl publicó un algoritmo que resolvía la inferencia exacta en tiempo polinomial respecto al número de nodos en redes individualmente conectadas [20]. Pearl además publicó un algoritmo para redes con conexiones múltiples en el que transformaba la red en una con conexiones individuales y aplicaba el algoritmo anterior. Sin embargo, el proceso de transformación de la red es un problema NP-completo.

El algoritmo de propagación de inferencia exacta más famoso es el *algoritmo de propagación en árbol de cliques* propuesto por Lauritzen y Spiegelhalter [21]. También es llamado el *algoritmo de clustering* o *propagación de creencias*. El algoritmo construye el árbol de cliques mediante la triangulación del grafo moral correspondiente al grafo no dirigido subyacente en la red y luego realiza propagación de mensajes en el árbol de cliques. El algoritmo de propagación en árbol de cliques es eficiente en redes esparcidas pero puede ser extremadamente lento en redes densas. En general, se comporta exponencial con respecto al tamaño del clique más grande del grafo moral triangulado.

El algoritmo de *eliminación de variables*(VE)[22] consiste en ir eliminando otras variables una por una e ir sumándolas. La complejidad del algoritmo se mide por el número de sumas y multiplicaciones que se realizan. Una eliminación óptima de las variables conduce a la menor complejidad, pero el problema de hallar una eliminación óptima es NP-completo.

El algoritmo de *Inversión de arcos / reducción de nodos* de Shachter [23] [24] aplica una serie de operadores a la red, que invierten la orientación de las aristas usando la regla de Bayes y terminan reduciendo la red al conjunto de nodos que representan a las variables consultadas con los nodos de la evidencia

como sus predecesores.

La *inferencia probabilística simbólica*(SPI) ve a la inferencia probabilística como un problema de optimización combinatorio: encontrar una factorización óptima dado un conjunto de distribuciones de probabilidad [25].

Inferencia aproximada

Los algoritmos de simulación estocástica, también llamados muestreo estocástico o algoritmos de Monte Carlo, son los algoritmos de inferencia aproximada más conocidos. Generan un conjunto de muestras seleccionadas al azar o instanciaciones de la red de acuerdo a las tablas de probabilidad condicional del modelo y después estiman las probabilidades de las variables de la consulta por la frecuencia de apariciones en la muestra. La precisión depende de la cantidad de muestras independientemente de la estructura de la red. Pueden ser divididos en dos tipos: *algoritmos de muestreo por importancia* y *Métodos de Monte Carlo con Cadenas de Markov* [15].

Los *métodos de simplificación de modelos* simplifican el modelo hasta que sea factible utilizar un algoritmo de inferencia exacta. Algunos simplifican el modelo eliminado probabilidades pequeñas o consideradas irrelevantes [26]. Otros involucran la eliminación de arcos [27] o de dependencias débiles [28]. El *algoritmo de espacio de estados* reduce la cardinalidad de las tablas de probabilidad condicional para simplificar el modelo [29].

Los *métodos basados en búsqueda* asumen que una fracción relativamente pequeña del espacio de probabilidad conjunta contiene la mayoría de la masa de probabilidad. Estos algoritmos buscan las instanciaciones de mayor probabilidad y las usan para obtener una aproximación razonable. Entre los más relevantes se encuentran el método de búsqueda *Top-N* de Henrion [30] y la variante de búsqueda de Poole usando "conflictos"[31, 32].

1.7. Modelos causales estructurales

Las redes bayesianas modelan las dependencias entre las variables pero no nos dicen nada acerca de las relaciones de causalidad entre estas. Una arista de X hacia Y en el DAG de una red bayesiana no implica que X cause a Y, sino que existe correlación entre ambas. Para modelar estas relaciones de causalidad es necesario un modelo que defina explícitamente las mismas.

Definición 1.7.1 (Modelo Causal Estructural) *Un modelo causal estructural(SCM) consiste en una tupla $\langle U, V, F\rangle$ donde:*

- *U es un conjunto de variables llamadas exógenas, términos de error o factores omitidos y representan factores externos al modelo.*
- *V es un conjunto de variables llamadas endógenas cuyos valores están determinados a partir de factores dentro del modelo.*

- *F es un conjunto de funciones que determinan los valores de las variables de V a partir de los valores de un subconjunto de variables de $U \cup V$.*

A partir de una asignación completa a las variables de U es posible determinar perfectamente el valor de todas las variables de V mediante las funciones de F. Cada variable $U_i \in U$ tiene asignada una distribución de probabilidad $P(U_i)$. De esta forma se añade no determinismo al modelo.

En el presente trabajo nos limitaremos a los modelos completamente especificados, donde se conocen todos los valores que pueden tomar todas las variables del modelo, así como su función de definición en el caso de las variables endógenas o su distribución de probabilidad en el caso de las exógenas. Se asume además que las variables exógenas son independientes entre sí. Por último, una variable no puede estar definida en términos de sí misma y en general no pueden existir dependencias cíclicas entre las variables.

Cada SCM tiene asociado un grafo dirigido acíclico que representa las relaciones de causalidad entre las variables del modelo. Cada vértice representa una variable del modelo y para cualesquiera dos variables X, Y tal que Y depende de X, existe una arista en el grafo dirigida del nodo que representa a X hacia el de Y. A partir de la relación entre un SCM y su modelo gráfico es posible establecer una definición gráfica de causalidad:

Definición 1.7.2 *Sea $M = \langle U, V, F \rangle$ un SCM y sea G su grafo asociado. Sean las variables $X, Y \in U \cup V$:*

- *X es una causa directa de Y si existe una arista de X hacia Y en G.*
- *X es una causa potencial de Y si X es ancestro de Y en G.*

Supongamos que se tiene un circuito lógico de dos entradas(X, Y) unidas por una compuerta AND(Z) y la salida de esta es negada(W). Un modelo causal estructural para este ejemplo podría ser el siguiente:

Modelo 1.7.1

$$U = \{X, Y\} \qquad V = \{Z, W\} \qquad F = \{f_z, f_w\}$$

$$P(X = 0) = P(X = 1) = P(Y = 0) = P(Y = 1) = 0{,}5$$

$$\begin{aligned} f_z: &\quad Z = X \cdot Y \\ f_w: &\quad W = 1 - Z \end{aligned}$$

En la figura 1.6 se muestra el DAG asociado a este modelo, que representa las relaciones de causalidad entre las variables. Se cumple que X es una causa directa de Z, y a la vez es una causa potencial de W.

Si realizamos las asignaciones $X = 1$ y $Y = 0$ a las variables de U entonces, mediante las funciones de correspondientes, obtenemos que $Z = 0$ y $W = 1$.

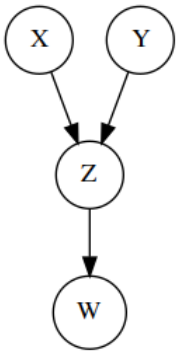


Figura 1.6: Grafo aociado a un modelo causal que representa el comportamiento de un circuito lógico.

1.7.1. Relación entre modelos estructurales causales y redes bayesianas

A partir de un modelo causal estructural es posible obtener una red bayesiana que codifique la información del modelo. De esta forma se preservarán las relaciones de causalidad entre las variables y será posible utilizar una red bayesiana para responder a preguntas causales.

Sea $M = \langle U, V, F \rangle$ un modelo causal estructural. Entonces es posible construir una red bayesiana $N = \langle G, P \rangle$ tal que:

- G es el modelo gráfico causal de M.
- P' es un conjunto de distribuciones de probabilidad condicional $P'(X \mid Pa_X)$, tal que $X \in U \cup V$ y Pa_X denota los padres de X en G. Si $X \in U$, entonces $Pa_X = \emptyset$ y $P'(U_i) = P(U_i)$. En caso de que $X \in V$ entonces, la distribución de probabilidad condicional se define como:

$$P'(X = x | Pa_X = x^*) = \left\{ \begin{array}{ll} 1 & \text{si } x = F_X(x^*) \\ 0 & eoc \end{array} \right\}.$$

  donde $F_X \in F$ es la función que define a la variable X.

Supongamos que se tiene el modelo:

Modelo 1.7.2

$$U = \{X, Y\} \qquad V = \{Z\} \qquad E = \{f_z\}$$

$$f_z : Z = X + Y$$

$$P(X = x) = \left\{ \begin{array}{ll} 0{,}7 & \text{si } x = 1 \\ 0{,}3 & e.o.c \end{array} \right\}.$$

$$P(Y=y)=\left\{\begin{array}{ll} 0{,}2 & \text{si } y=1 \\ 0{,}8 & e.o.c \end{array}\right\}.$$

Entonces es posible construir una red bayesiana cuya estructura sea la representada en el grafo de la figura 1.7.

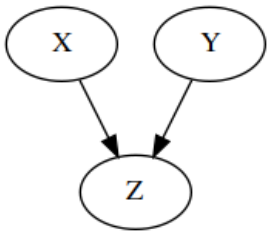


Figura 1.7: Grafo que representa la estructura de una red bayesiana obtenida a partir de un modelo estructural causal probabilístico

Por otro lado, asumiendo que Val(X) = $Val(Y)$ = {0,1} y Val(Z)={0,1,2}, las tablas de probabilidad asociadas a los nodos se pueden construir siguiendo las reglas descritas anteriormente:

X	$P(X)$
0	0.3
1	0.7

Y	$P(Y)$
0	0.8
1	0.2

X	Y	Z	$P(Z \mid X,Y)$
0	0	0	1
0	0	1	0
0	0	2	0
0	1	0	0
0	1	1	1
0	1	2	0
1	0	0	0
1	0	1	1
1	0	2	0
1	1	0	0
1	1	1	0
1	1	2	1

1.8. Intervenciones

Una *intervención*, en su variante más simple, consiste en determinar el comportamiento de una variable Y cuando otra variable X es forzada a tomar un valor x. El resultado de una intervención está determinado por la expresión $P(Y \mid do(X))$. El operador *do* indica que la variable X es intervenida.

Es importante recalcar la diferencia que existe entre $P(Y=y \mid X=x)$ y $P(Y=y \mid do(X=x))$. El primero calcula la probabilidad de $Y=y$ a partir de

la observación de $X = x$, mientras que el segundo lo hace a partir de una intervención donde X es forzada a tomar el valor x. En términos de distribuciones, $P(Y \mid X = x)$ refleja la distribución de Y en individuos cuyo valor de X es x, mientras que $P(Y \mid do(X = x))$ refleja la distribución de Y si todos los individuos de la población son forzados a tomar el valor $X = x$.

El operador *do* descarta los caminos espúreos entre X y Y que distorsionan el efecto real de X sobre Y. Un mundo donde $P(X \mid do(Y))$ fuera igual a $P(X \mid Y)$ estaría lleno de paradojas. Por ejemplo, las personas decidiendo no ir al médico reducirían la probabilidad de enfermarse o se eliminarían las estaciones de policía para reducir el número de crímenes.

El experimento aleatorio controlado ha sido la herramienta por excelencia de los estadísticos para calcular los resultados de una intervención. Sin embargo, estos experimentos suelen ser costosos y en ocasiones incluso imposibles en la práctica. Uno de los logros más notables de la teoría de la causalidad ha sido brindar herramientas para simular una intervención sin llegar a realizarla.

La figura 1.8a muestra el grafo de un modelo que define las relaciones de causalidad entre el género(Y), tomar un medicamento(X) y la recuperación del paciente(Y). Se sabe que el género del paciente afecta la recuperación, así como en la decisión de usar o no el medicamento. Supongamos que se quiere medir el efecto del medicamento en la recuperación de un paciente, o sea $P(Y \mid do(X = x))$.

Definición 1.8.1 (Submodelo) *Sea M un SCM, $X \subset V$ y x una asignación de valores a X. Se dice que M_x es submodelo de M si M_x contiene las mismas variables de U y V, pero el conjunto de funciones F es sustituido por F_x, donde:*

$$F_x = \{F_{V_i} \in F : V_i \notin X\} \cup \{X_i = x_i : X_i \in X\}$$

Es decir, el submodelo M_x de M se obtiene sustituyendo todas las funciones que definen a las variables de X por la asignación $X_i = x_i$ correspondiente. Gráficamente este procedimiento consiste en eliminar todas las aristas que inciden en las variables de X, puesto que ya no dependen de ninguna otra variable sino que se les asigna un valor en concreto. La figura 1.8b muestra el submodelo M_x resultante de intervenir la variable X en el modelo M de la figura 1.8a.

Se cumple entonces que $P(Y = y \mid do(X = x))$ es igual a la probabilidad condicional $P_m(Y = y \mid X = x)$ que prevalece en M_x. Para calcular esta última probabilidad, es necesario encontrar invariantes entre P y P_m. Por un lado, se sabe que $P_m(Z = z) = P(Z = z)$ dado que eliminar la arista $Z \to X$ en el grafo original no afecta la probabilidad de Z en el grafo resultante. Por otro lado, también se cumple que $P(Y = y \mid X = x, Z = z) = P_m(Y = y \mid X = x, Z = z)$ porque la forma en la que Y responde a X y Z es la misma independientemente de que una de ellas cambie naturalmente o se fije su valor. Por último, X y Z están d-separados y por tanto son independientes en M_x, por lo que $P_m(Z = z \mid X = x) = P_m(Z = z)$. Juntando todas estas consideraciones, se tiene que:

$$P(Y = y \,|do(X = x)) \tag{1.1}$$
$$= P_m(Y = y \mid X = x) \quad \text{(por definición)} \tag{1.2}$$
$$= \sum_z P_m(Y = y \mid X = x, Z = z) P_m(Z = z \mid X = x) \tag{1.3}$$
$$= \sum_z P_m(Y = y \mid X = x, Z = z) P_m(Z = z) \tag{1.4}$$
$$\tag{1.5}$$

La ecuación 1.3 es obtenida mediante la Fórmula de la Probabilidad Total. Utilizando las invariantes anteriores, es posible obtener una fórmula para el efecto causal de X en Y en términos de probabilidades preintervención:

$$P(Y = y \mid do(X = x)) = \sum_z P(Y = y \mid X = x, Z = z) P(Z = z) \tag{1.6}$$

La fórmula anterior es denominada *fórmula de ajuste* [13]. En este caso se dice también que se está "ajustando para Z".

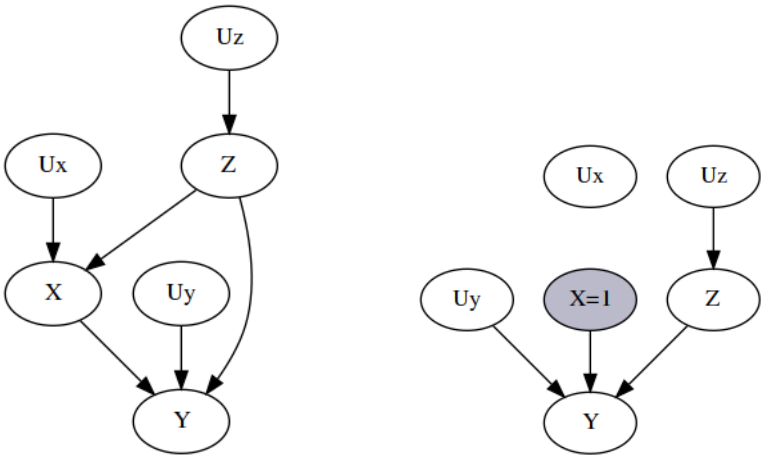


Figura 1.8: Obtención del submodelo resultante de una intervención

La fórmula de ajuste puede ser generalizada si se ajusta con $\mathbf{Z} = Pa_X$, el conjunto de los padres de X [13]:

Teorema 1.8.1 (Regla del efecto causal) *Sea G un DAG correspondiente a un SCM. El efecto causal de X en Y está determinado por la fórmula:*

$$P(Y = y \mid do(X = x)) = \sum_z P(Y = y \mid X = x, Pa_X = z) P(Pa_X = z)$$

donde z toma todos las posibles combinaciones de valores que pueden tomar los padres de X.

A veces interesa medir el efecto de intervenir varias variables. Para ello se utiliza la fórmula del *producto truncado* o *fórmula-g*:

$$P(x_1, x_2, ..., x_n \mid do(x)) = \prod_{i=1}^{n} P(x_i \mid Pa_{X_i}) \quad \forall X_i : X_i \notin X$$

La fórmula del producto truncado, como su nombre lo indica, resulta de eliminar los factores que corresponden a las variables intervenidas, de la fórmula de la distribución conjunta. Por ejemplo, supongamos que se tiene la siguiente fórmula de distribución conjunta para un modelo gráfico probabilista:

$$P(X, Y, Z, W) = P(X)P(Y \mid X)P(Z \mid X)P(W \mid Y, Z)$$

Entonces, si se quiere intervenir las variables Y y Z, la fórmula del producto truncado queda:

$$P(X = x, W = w \mid do(Y = y), do(Z = z)) = P(X)P(W = w \mid Y = y, Z = z)$$

1.8.1. El criterio de la puerta trasera

En ocasiones no es posible ajustar para los padres de una variable porque a pesar de estar representados en el grafo, no se cuenta con suficiente información sobre estos para medirlos. En tales circunstancias es necesario encontrar un conjunto alternativo de variables $\mathbf{Z}$ con el que ajustar [13].

Definición 1.8.2 (Criterio de la puerta trasera) *Dado un par ordenado de variables (X, Y) en un grafo dirigido acíclico G, un conjunto de variables $\boldsymbol{Z}$ satisface el criterio de la puerta trasera relativo a (X, Y) si ningún nodo en $\boldsymbol{Z}$ desciende de X y $\boldsymbol{Z}$ bloquea cada camino entre X y Y que contiene una arista que entra a X.*

Es fácil observar que $Z = Pa_X$, el conjunto de los padres de X, satisface el criterio de la puerta trasera. Los caminos de X hacia Y con una arista que incide en X son llamados caminos de puerta trasera de X hacia Y.

Teorema 1.8.2 *Si un conjunto de variables $\boldsymbol{Z}$ satisface el criterio de la puerta trasera relativo a (X, Y), entonces el efecto causal de X en Y está dado por la fórmula:*

$$P(Y = y|do(X = x)) = \sum_{z} P(Y = y|X = x, Z = z)P(Z = z)$$

Supongamos que queremos medir la recuperación de un paciente a partir de una intervención en la que se le suministra un medicamento. En la recuperación del paciente interviene también el peso, y aunque no conocemos su valor,

también se sabe que el estatus socio-económico es una causa tanto del medicamento que se le suministra, como del peso del paciente. Se quiere determinar cuán efectivo será el medicamento en la recuperación del paciente(figura 1.9). A pesar de que no conocemos el valor de Z, si escogemos $\mathbf{Z} = \{W\}$, vemos que este conjunto satisface el criterio de la puerta trasera puesto que bloquea el camino de puerta trasera: $X \leftarrow Z \rightarrow W \rightarrow Y$. Luego la fórmula de ajuste resulta:

$$P(Y = y|do(X = x)) = \sum_{w} P(Y = y|X = x, W = w)P(W = w)$$

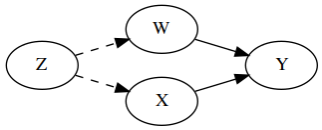


Figura 1.9: Aplicación del criterio de la puerta trasera

Puede que interese identificar el efecto de una intervención no en toda la población sino en un subconjunto de esta que reúna unas características $Z = z$ luego de la intervención. Este es llamado el *efecto z-epecífico*

Definición 1.8.3 *Sean X,Y variables aleatorias y $\boldsymbol{Z}$ un conjunto de variables aleatorias. Se denomina efecto z-específico de X en Y al resultado de la expresión: $P(Y = y \mid do(X = x), \boldsymbol{Z} = \boldsymbol{z})$.*

Teorema 1.8.3 *El efecto z-específico $P(Y = y \mid do(X = x), Z = z)$ es identificado siempre que se pueda encontrar un conjunto de variables S tal que $S \cup Z$ satisface el criterio de la puerta trasera, resultando en la siguiente fórmula de ajuste:*

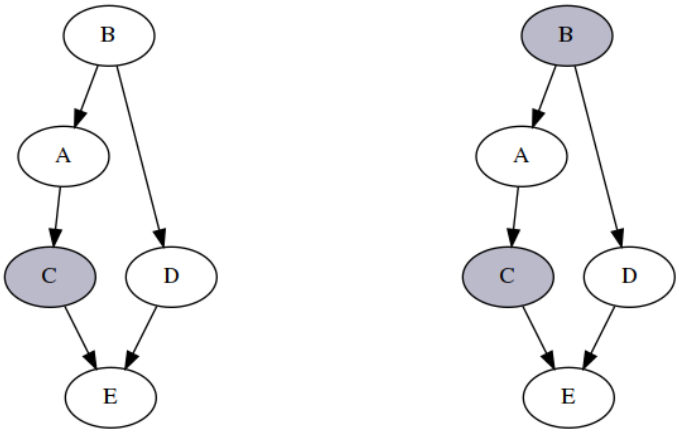
$$P(Y = y \mid do(X = x)) = \sum_{s} P(Y = y \mid X = x, S = s, Z = z)P(S = s)$$

En el modelo de la figura 1.10a se muestra que el efecto z-específico de A en E no es identificable si se escoge $S = \emptyset$ porque $\{C\}$ no satisface el criterio de la puerta trasera. Sin embargo, en la figura 1.10b se muestra como escogiendo $S = \{B\}$, entonces este si es identificable, ya que $\{B, C\}$ sí lo satisface. En este caso la fórmula de ajuste resultante sería:

$$P(E = e \mid do(A = a), C = c) = \sum_{b} P(E = e \mid A = a, B = b, C = c)P(B = b)$$

1.8.2. El criterio de la puerta principal

El criterio de la puerta trasera no cubre todos los escenarios donde es posible aplicar el operador *do*. Existe otro criterio que nos puede ayudar en tales casos [13].



(a) $S = \emptyset$ (b) $S = \{B\}$

Figura 1.10: Identificación del efecto Z-específico

Definición 1.8.4 (Criterio de la puerta principal) *Un conjunto de variables* $\boldsymbol{Z}$ *se dice que satisface el criterio de la puerta principal relativo a un par ordenado de variables* (X, Y) *si:*

1. $\boldsymbol{Z}$ *intercepta todos los caminos dirigidos de* X *hacia* Y*.*
2. *No hay ningún camino de puerta trasera de* X *hacia* $\boldsymbol{Z}$ *que esté desbloqueado.*
3. *Todos los caminos de puerta trasera de* $\boldsymbol{Z}$ *hacia* Y *son bloqueados por* X*.*

Teorema 1.8.4 *Si* $\boldsymbol{Z}$ *satisface el criterio de la puerta principal relativo a* (X, Y) *y si* $P(X, \boldsymbol{Z}) > 0$*, entonces el efecto causal de* X *en* Y *está dado por la fórmula:*

$$P(Y = y|do(X = x)) = \sum_{z} P(\boldsymbol{Z} = \boldsymbol{z}|X = x) \sum_{x'} P(Y = y|X = x', \boldsymbol{Z} = \boldsymbol{z}) P(X = x')$$

En el diagrama de la figura 1.11 se puede comprobar que $\{Z\}$ satisface el criterio de la puerta principal relativo a (X, Y) porque:

1. Z intercepta el único camino dirigido de X hacia Y: $X \rightarrow Z \rightarrow Y$.
2. El camino $X \leftarrow U \rightarrow Y \leftarrow Z$ está bloqueado por U.
3. El camino de puerta trasera $Z \leftarrow X \leftarrow U \rightarrow Y$ es bloqueado por X.

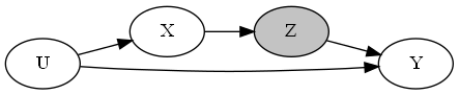


Figura 1.11: Aplicación del criterio de la puerta principal

1.8.3. El cálculo-do

El cálculo-do [2] es un conjunto de reglas que nos permite transformar una expresión en otra con el objetivo de obtener una expresión final libre del operador *do* que nos permita conocer el resultado de una intervención a partir de probabilidades preintervención:

Regla 1:

$$P(Y|do(X),Z,W) = P(Y|do(X),Z)$$

en el caso de que Z bloquee todos los caminos de W a Y en el submodelo M_x.

Regla 2:

$$P(Y|do(X),Z) = P(Y|X,Z)$$

si Z satisface el criterio de la puerta trasera.

Regla 3:

$$P(Y|do(X)) = P(Y)$$

si no existe un camino de X hacia Y con solo aristas hacia delante.

La Regla 1 nos dice que cuando observamos una variable W luego de la intervención de X y esta sea irrelevante a Y condicionado a una variable Z, la distribución de probabilidad de Y no cambiará. Por otro lado, la Regla 2 nos dice que una vez que ajustemos para cada variable de confusión, toda correlación que permanezca es un efecto causal genuino. Por último, Regla 3 nos dice que si intervenimos una variable que no afecta a Y, entonces su distribución de probabilidad no se ve afectada.

Cada regla tiene a su vez un significado sintáctico. La Regla 1 permite añadir o eliminar observaciones. La Regla 2 permite reemplazar una intervención por una observación y viceversa. La Regla 3 permite la adición o eliminación de intervenciones.

1.8.4. Intervenciones mediante inferencia bayesiana

Otra alternativa para calcular el efecto de una intervención es manipular el modelo M, obteniendo un submodelo M_x en el que calcular el efecto de la intervención $P(Y = y \mid do(X = x))$ se reduce a convertir la variable intervenida en una observación y calcular $P_m(Y = y \mid X = x)$. Una vez obtenido este modelo, se puede construir una red bayesiana a partir del mismo mediante el procedimiento descrito en la sección 1.7.1. Posteriormente se instancian las variables observadas

y se aplica un algoritmo de inferencia bayesiana para obtener la respuesta a la consulta deseada.

1.9. Contrafactuales

Los contrafactuales son utilizados para comparar dos posibles salidas de una variable, exactamente bajo las mismas condiciones, excepto en una que difiere. Por ejemplo, para aliviar un malestar un individuo decide tomar un medicamento, recuperándose en 6 horas. Este podría preguntarse: Dado que tomé el medicamento y me recuperé en 6 horas, ¿cuál sería el tiempo de recuperación si no lo hubiese tomado ? Para responder a la pregunta, existen tres variables que deben ser tenidas en cuenta:

- $Y_{x=1}$ es el tiempo de recuperación si se toma el medicamento.
- $Y_{x=0}$ es el tiempo de recuperación si no se toma.
- X representa la decisión de tomar o no el medicamento.

Luego, la expresión que responde a la pregunta es:

$$E(Y_{x=0}|X=1, Y_{x=1}=6)$$

Note que $Y_{x=1}$ y $Y_{x=0}$ corresponden a la misma variable pero con un antecedente distinto. La expresión anterior es distinta de $E[Y \mid do(X=0), Y=6]$ porque esta última no tiene en cuenta la diferencia entre el tiempo de recuperación dado que se tomó medicamento del tiempo de recuperación dado que no se tomó. Esta diferencia plantea la existencia de dos mundos: el mundo real, donde el individuo toma el medicamento y se recupera en 6 horas, y un mundo alternativo, donde el individuo no toma el medicamento y podría, por tanto, experimentar un tiempo de recuperación distinto al del mundo real.

1.9.1. Contrafactuales deterministas

Sea M un SCM completamente especificado, donde se conocen tanto las funciones (F) como los valores de todas las variables exógenas. En dicho modelo cada asignación $\mathbf{U} = \mathbf{u}$ a las variables exógenas identifica a un individuo de la población o situación particular en la naturaleza del problema. Consideremos la sentencia contrafactual: "Y hubiese tomado el valor y, si X hubiese tomado el valor x, en la situación $\mathbf{U} = \mathbf{u}$", denotada como $Y_x(\mathbf{u}) = y$ donde Y y X son dos variables en V. Este tipo de contrafactual se denomina *determinista*, porque corresponde a un individuo de la población del que conocemos el valor de cada variable relevante. Supongamos que se tiene información de un individuo mediante la evidencia $\mathbf{E} = \mathbf{e}$, que asigna a cada variable endógena un valor. Para calcular el contrafactual determinista $Y_x(\mathbf{u}) = y$ en base a la evidencia $\mathbf{e}$ se debe seguir el siguiente algoritmo [13]:

1. Abducción: Usar la evidencia $\mathbf{E} = \mathbf{e}$ para determinar el valor de $\mathbf{U}$. Este paso explica el pasado, ($\mathbf{u}$) en base a la evidencia actual $\mathbf{e}$.

2. Acción: Modificar el modelo M, reemplazando las ecuaciones que definen a X por la función apropiada $X = x$, para obtener un modelo modificado M_x. Este paso cambia el curso de la historia (mínimamente) para cumplir con la hipótesis del contrafactual.

3. Predicción: Usar M_x y el valor de $\mathbf{U}$ para computar el valor de Y, la consecuencia del contrafactual. Este paso predice el futuro (Y) basado en el nuevo entendimiento del pasado y la nueva condición establecida.

Ilustremos este procedimiento mediante un ejemplo. Un conductor debe moverse desde su hogar a su trabajo. Durante el recorrido, se encuentra con un desvío que igual conduce a su destino pero no decide tomarlo. Una vez llegado a su centro de trabajo, nota que le tomó media hora el recorrido habiéndose movido a una velocidad media de 50 km/h y se pregunta cuánto se hubiese demorado yendo a la misma velocidad pero tomando el desvío. Un modelo causal para esta situación podría ser el siguiente:

$$\begin{aligned} D &= U_d \\ V &= U_v \\ T &= \frac{25 - 5 \cdot D}{V} + U_t \end{aligned}$$

En este caso D es una variable binaria que representa la decisión de tomar el desvío y toma valor 1 en caso afirmativo y 0 en caso contrario, V la velocidad a la que se mueve el conductor (en km/h) y T el tiempo que este demora en llegar a su destino (en horas). Siguiendo los pasos listados en el algoritmo anterior:

1. Abducción: Primeramente obtenemos los valores de U_d, U_v y U_t a partir de la evidencia, sustituyendo los valores $D = 0$, $V = 50$ y $T = 0{,}5$ en las ecuaciones del modelo y despejando las variables deseadas.

$$\begin{aligned} U_d &= D = 0 \\ U_v &= V = 50 \\ U_t &= T - \frac{25 - 5 \cdot D}{V} = 0{,}5 - \frac{25 - 5(0)}{50} = 0 \end{aligned}$$

2. Acción: Luego corresponde modificar el modelo, sustituyendo la ecuación que define a D por $D = 1$:

$$\begin{aligned} D &= 1 \\ V &= U_v \\ T &= \frac{25 - 5 \cdot D}{V} + U_t \end{aligned}$$

3. Predicción: Por último, predecimos la variable deseada, haciendo uso de los valores de $\mathbf{U}$ y del modelo M_d computados anteriormente.

$$\begin{aligned} & T_{D=1}(U_d = 1, U_v = 50, U_t = 0) \\ & = \frac{25 - 5 \cdot 1}{50} + 0 \\ & = 0{,}4 \end{aligned}$$

Por tanto, tomar el desvío le hubiese hecho llegar antes al trabajo.

1.9.2. Contrafactuales no deterministas

Los contrafactuales pueden utilizarse también en el caso no determinista para estudiar el comportamiento de las variables en una clase o rango de la población. La expresión $P(Y_x = y \mid \mathbf{E} = \mathbf{e})$ expresa la probabilidad de que la variable Y tomase el valor y si se fijara el valor de X a x en individuos donde se observa la evidencia $\mathbf{E} = \mathbf{e}$.

El procedimiento para calcular contrafactuales deterministas puede modificarse para el caso no determinista como se muestra a continuación [13]:

1. Abducción: Actualizar $P(\mathbf{U})$ a partir de la evidencia $\mathbf{E} = \mathbf{e}$ para obtener $P(\mathbf{U} \mid \mathbf{E} = \mathbf{e})$.

2. Acción: Modificar el modelo M, reemplazando las ecuaciones que definen a las variables en X por la apropiada asignación $X = x$ para obtener el modelo modificado M_x.

3. Predicción: Usar el modelo modificado M_x, y las probabilidades actualizadas sobre las variables de $\mathbf{U}$, $P(\mathbf{U} \mid \mathbf{E} = \mathbf{e})$, para calcular la probabilidad de Y_x, la respuesta al contrafactual.

1.9.3. Contrafactuales mediante inferencia bayesiana

Al igual que en las intervenciones, la inferencia bayesiana puede utilizarse en la resolución de contrafactuales, una vez que estos sean definidos en términos de datos recogidos de observaciones. A partir del procedimiento para computar contrafactuales no deterministas mostrado en la sección anterior, es posible diseñar un algoritmo que calcule contrafactuales utilizando la inferencia bayesiana. Los pasos a seguir por el algoritmo se describen a continuación:

1. Construir una red bayesiana N a partir del modelo M.

2. Realizar inferencia en la red para obtener las creencias actualizadas de las variables exógenas a partir de la introducción de la evidencia $E = e$(abducción).

3. Modificar la red de tal forma que se refleje la transformación del modelo original M en el submodelo M_x, eliminando todas las aristas que inciden en el nodo de la variable X pero preservando las probabilidades que se actualizaron en el paso anterior(acción)

4. Instanciar tanto la variable intervenida X como las variables observadas observaciones y aplicar inferencia en la red para obtener la respuesta al contrafactual(predicción).

Este método ofrece una alternativa intuitiva para calcular contrafactuales. Su inconveniente es la gran cantidad de recursos que demanda al tener que realizarse dos veces un algoritmo de inferencia cuya complejidad es exponencial.

1.9.4. Método de las redes gemelas

El método de las redes gemelas [33, 34] ofrece una alternativa al algoritmo anterior para reducir el costo computacional. Este se basa igualmente en la inferencia bayesiana. Para ello se expande el modelo original M, construyendo un nuevo modelo M' cuyo grafo causal consiste en dos redes idénticas en estructura interconectadas. Una representa el mundo real y otro el alternativo. Formalmente $M' = \langle U', V', F' \rangle$ donde:

- $U' = U$. El conjunto de las variables exógenas se mantiene intacto.
- $V' = V \cup V^*$ donde $V^* = \{V_i^* : V_i \in V\}$. Se crea una variable endógena idéntica a cada variable del modelo original, pero etiquetada con un nombre distinto. Las variables de V^* corresponden al mundo alternativo.
- $F' = F \cup F^*$ donde $F^* = \{F_{V_i}^* : F_{V_i} \in F\}$. Cada variable V_i^* poseerá una función $F_{V_i}^*$ que resultará semejante a la de su gemela pero cuyos argumentos corresponderán a las variables del mundo alternativo.

Una vez construidas las redes gemelas se representa en el mismo modelo tanto el mundo real como el mundo alternativo. Supongamos que se quiere calcular el contrafactual $P(Y_x = y \mid Z = z)$. En el modelo de las redes gemelas M' se cumple que:

$$P(Y_x = y \mid Z = z) = P'(Y^* = y \mid do(X^* = x), Z = z)$$

De esta forma, calcular un contrafactual en el modelo original se reduce a calcular una intervención en el modelo de las redes gemelas. Podemos ir más allá y utilizando el algoritmo visto en la sección de intervención en redes bayesianas, obtener un modelo M'_x que simule la intervención de X y donde se cumpla que:

$$P'(Y^* = y \mid do(X^* = x), Z = z) = P'_m(Y^* = y \mid X^* = x, Z = z)$$

. Por tanto calcular un contrafactual en el modelo original M puede reducirse a predecir el valor de una variable a partir de observaciones en el modelo resultante de intervenir el modelo de las redes gemelas M'_x. Para calcular esta predicción

puede transformarse el modelo en una red bayesiana y aplicar un algoritmo de inferencia.

Para disminuir el costo computacional de la inferencia en una red de tamaño igual al doble de la original se puede utilizar el método de mezcla de nodos [35]. Por cada nodo X^* del mundo alternativo, si este no desciende de un nodo intervenido ni de un nodo que corresponde a una de las variables que queremos estimar, se puede mezclar con su correspondiente nodo del mundo real X, conformando un único nodo X' donde $Pa'_X = Pa_X \cup Pa^*_X$ y $Ch'_X = Ch_X \cup Ch^*_X$. Este nodo toma los mismos valores que X y X^* y la misma función de definición que X.

En resumen, calcular un contrafactual en un SCM puede resumirse en el siguiente algoritmo:

1. Construir el modelo de redes gemelas M' a partir de M.
2. Obtener el modelo M'_x a partir de M', resultante de la intervención $X = x$.
3. Reducir el tamaño del modelo M'_x mediante la mezcla de nodos, obteniendo un modelo M''_x
4. Construir una red bayesiana N a partir de M''_x.
5. Calcular $P'_m(Y* = y \mid X^* = x, Z = z)$ en N, aplicando un algoritmo de evidencia bayesiana.

La figura 1.12 muestra el proceso de transformaciones que va sufriendo la red a lo largo de la ejecución del algoritmo para calcular el contrafactual $P(W_z = w \mid Y = y)$. La figura 1.12a muestra el modelo original, 1.12b el modelo de las redes gemelas, 1.12c el submodelo resultante de intervenir las redes gemelas, 1.12d el modelo resultante de aplicar mezcla de nodos al submodelo y 1.12e la red bayesiana construida y con las correspondientes variables instanciadas.

1.10. Atribución

Consideremos el ejemplo de una persona que contrae una enfermedad y para eliminarla decide tomar un medicamento. Un tiempo después, dicha persona se recupera y se hace la pregunta: ¿Realmente fue tomar el medicamento lo que me hizo recuperarme de la enfermedad ?. La probabilidad de necesidad (PN) mide el grado en que tomar el medicamento $X = 1$ es necesario para recuperarse de la enfermedad($Y = 1$). El valor de PN lo determina la probabilidad de que no exista recuperación si no se toma el medicamento, dado que se tomó el medicamento y hubo recuperación:

$$PN = P(Y_0 = 0|Y = 1, X = 1)$$

Por otro lado, puede darse el caso de un individuo que no tome el medicamento y no se recupere de la enfermedad. En tal caso, dicho individuo podría

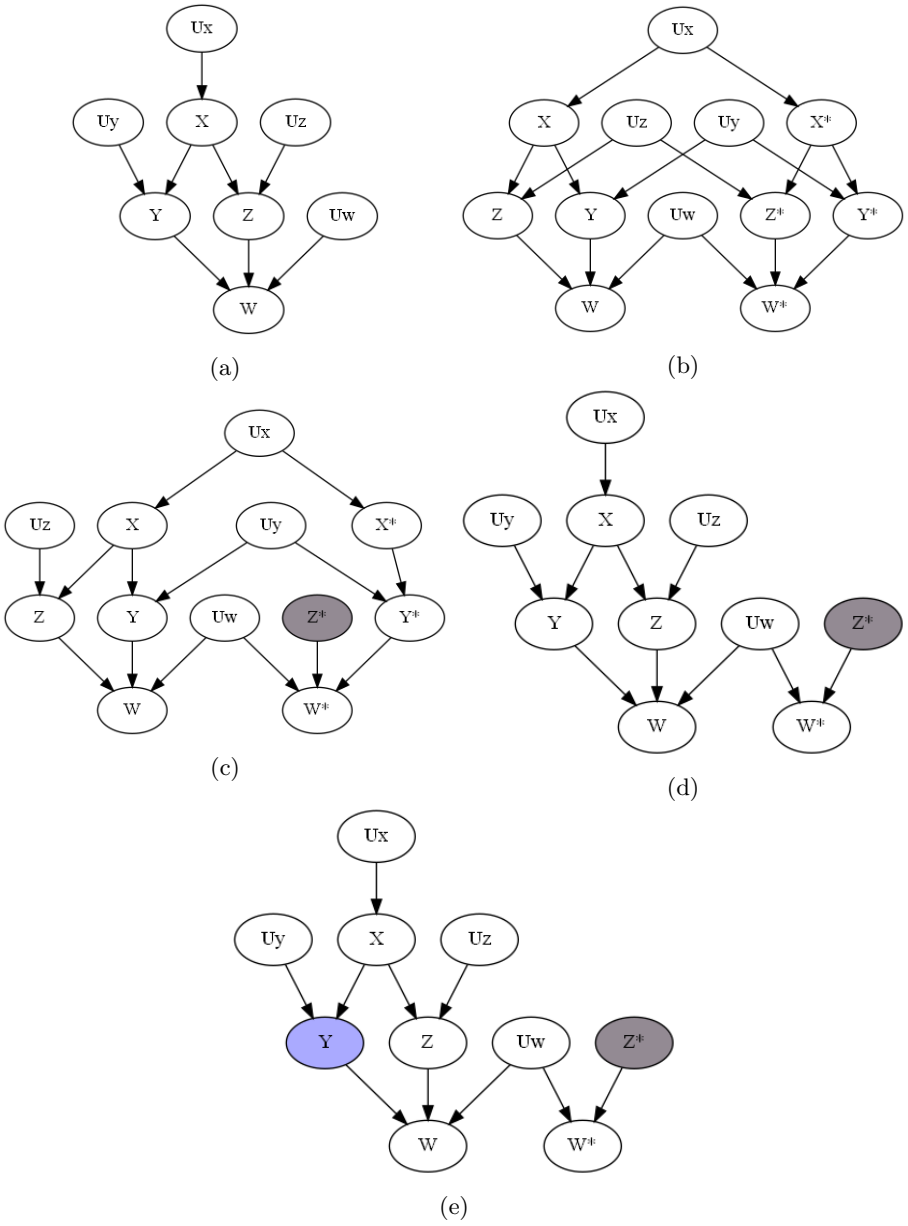


Figura 1.12: Aplicación del método de las redes gemelas para calcular un contrafactual

hacerse la pregunta: ¿Si hubiese tomado el medicamento me hubiese recuperado ? La probabilidad de suficiencia (PS) indica el grado en que la acción no realizada (tomar el medicamento) hubiese sido suficiente para lograr el efecto contrario, es decir, recuperarse ($Y = 1$). Matemáticamente, es equivalente a la probabilidad de que el individuo se hubiese recuperado si hubiese tomado el medicamento, dado que no lo tomó y no hubo recuperación.

$$PS = P(Y_1 = 1|X = 0, Y = 0)$$

Imaginemos un tercer individuo que padece la enfermedad. ¿Y si resulta que la enfermedad está en una fase terminal, y tomar o no el medicamento no influirá en la recuperación del mismo? ¿Y si en realidad no necesita tomar el medicamento puesto que su sistema inmunológico por si solo puede combatir la enfermedad ? La única razón para que dicho individuo tome el medicamento es que haciéndolo se recuperará y si no lo toma entonces no se recuperará. En otras palabras, lo que quiere determinar el individuo es la probabilidad de que tomar el medicamento sea una condición necesaria y suficiente para su recuperación:

$$PNS = P(Y_1 = 1, Y_0 = 0)$$

Las tres probabilidades anteriores, PN, PS y PNS son ejemplos del uso de la causalidad para atribuir un efecto (recuperarse) a una causa (tomar un medicamento). A continuación se definen formalmente [13]:

Definición 1.10.1 *Sean X y Y variables binarias en un modelo causal M. Sean x, y los valores que corresponden a las asignaciones X=verdadero y Y=verdadero respectivamente. A su vez, sean x', y' los valores de sus respectivos complementos. Se definen las siguientes probabilidades:*

- ***Probabilidad de necesidad (PN):***

$$PN = P(Y_{x'} = y'|Y = y, X = x)$$

- ***Probabilidad de suficiencia (PS):***

$$PS = P(Y_x = y|X = x', Y = y')$$

- ***Probabilidad de necesidad y suficiencia (PNS):***

$$PNS = P(Y_x = y, Y_{x'} = y')$$

En el caso de la PNS, resulta imposible calcularla experimentalmente a partir de la definición provista. Sin embargo, es posible definirla en términos de PN y PS [36]:

Proposición 1.10.1 *Las probabilidades PN, PS y PNS satisfacen la siguiente relación :*

$$PNS = P(x, y)PN + P(x', y')PS$$

1.11. Mediación

El modelo canónico para un problema de mediación toma la siguiente forma:

$$x = f_X(U_X) \qquad m = f_M(x, u_M) \qquad y = f_Y(x, m, u_Y)$$

donde X(origen), M(mediador) y Y(destino) son variables aleatorias, f_X, f_M y f_Y son funciones y U_X, U_M y U_Y son los factores omitidos que actúan sobre las variables X, M y Y respectivamente (Figura 1.13). Se asume además que los factores omitidos son independientes entre sí.

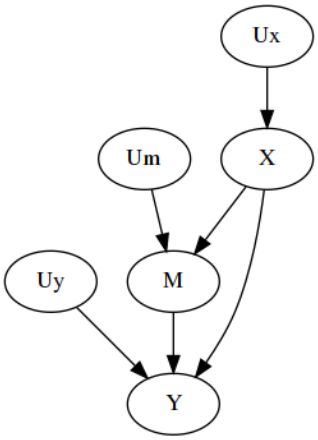


Figura 1.13: Modelo canónico para análisis de mediación

Teniendo el cuenta el modelo anterior, existen cuatro tipos de efectos diferentes que describen la transición de $X = x$ hacia $X = x'$ [13]. Sin perder generalidad, supongamos que $x = 0$ y $x' = 1$.

1. **Efecto total**

$$\begin{aligned} TE &= E[Y_1 - Y_0] \\ &= E[Y \mid do(X = 1)] - E[Y \mid do(X = 0)] \end{aligned}$$

   El efecto total mide el incremento esperado en Y cuando X cambia de $X = 0$ hacia $X = 1$, mientras que M toma su valor natural, dictado por la función f_M

2. **Efecto directo controlado**

$$\begin{aligned} CDE(m) &= E[Y_{1,m} - Y_{0,m}] \\ &= E[Y \mid do(X = 1, M = m)] - E[Y \mid do(X = 0, M = m)] \end{aligned}$$

   El efecto directo controlado mide el incremento esperado en Y cuando X cambia de $X = 0$ hacia $X = 1$ y el mediador M es forzado a tomar un valor m.

3. **Efecto directo natural**

$$\begin{aligned} NDE &= E[Y_{1,M_0} - Y_{0,M_0}] \\ &= E[Y \mid do(X = 1, M = M_0)] - E[Y \mid do(X = 0, M = M_0)] \end{aligned}$$

   El efecto directo natural mide el incremento esperado en Y cuando X cambia de $X = 0$ a $X = 1$ y el mediador M obtiene el valor que hubiese tomado en caso de que X fuese forzado a tomar el valor $X = 0$.

4. **Efecto indirecto natural**

$$\begin{aligned} NIE &= E[Y_{0,M_1} - Y_{0,M_0}] \\ &= E[Y \mid do(X = 0, M = M_1)] - E[Y \mid do(X = 0, M = M_0)] \end{aligned}$$

   El efecto indirecto natural mide el incremento esperado en Y cuando X se mantiene con el valor fijo $X = 0$ y el mediador cambia de M_0 a M_1.

Supongamos que se quiere determinar si existe discriminación por género en las entrevistas de contratación para un trabajo. Se dispone de un modelo como el de la figura 1.13 que representa la relación entre el género del entrevistado (X), su calificación en la entrevista (M) y si este es contratado (Y). Para determinar si existe discriminación se puede utilizar el efecto directo controlado, asignándoles un valor de calificación fijo m a la población de estudio y utilizando la fórmula vista para el CDE, donde $X = 1$ representa el sexo femenino y $X = 0$ el masculino y por otro lado $Y = 1$ indica que el entrevistado fue contratado y $Y = 0$ indica el caso contrario. Si se obtiene un valor negativo, ello indica que las mujeres son más propensas a no ser aceptadas en una entrevista de trabajo por discriminación.

Capítulo 2

Propuesta de software

Se desarrolló un programa capaz de brindar al usuario la posibilidad de realizar operaciones de inferencia causal. Las características que reúne el programa son las siguientes:

1. Ofrecer una interfaz visual que permita a usuarios no expertos en programación realizar operaciones de inferencia causal. Dicha interfaz debe asistir al usuario tanto en la creación de los modelos como en las operaciones de inferencia causal, notificándolo en el caso de que existan errores en la entrada provista.

2. Permitir la construcción y visualización de SCMs. Debe permitir además exportar los modelos a ficheros externos para posteriormente ser cargados en el programa.

3. Permitir realizar operaciones de inferencia causal sobre modelos causales estructurales. Ofrecer en un menú los distintos tipos de inferencia causal disponibles. Permitir al usuario seleccionar los parámetros que el programa tendrá en cuenta durante la inferencia.

4. Ofrecer los resultados correspondientes a la consulta causal especificada por el usuario.

2.1. Implementación

El lenguaje de programación escogido para la implementación fue **Python** en su versión 3,8. Su elección se debe a que es un lenguaje de alto nivel, multiparadigma y con una filosofía que apuesta por un código legible y sencillo.

La implementación de la parte lógica del proyecto consiste de 3 módulos contenidos en un directorio llamado *causality*. Estos módulos son:

- *main.py*: Contiene la implementación del SCM.

- *tests.py*: Conjunto de tests unitarios destinados a probar las componentes fundamentales de la implementación.
- util.py: Conjunto de métodos auxiliares que serán utilizados desde otros módulos.

2.1.1. Implementación del modelo causal estructural

En el módulo *main.py* se trabaja con un conjunto de clases para la implementación del modelo:

```
class Variable:
   def __init__(self, name, values):...

class ExogenousVariable(Variable):
   def __init__(self, name, values, distribution):...

class EndogenousVariable(Variable):
   def __init__(self, name, values, parents, expression):...

class SCM:
   def __init__(self, exogenous_variables, endogenous_variables):...
```

La clase *Variable* reúne las características comunes de las variables tanto exógenas como endógenas. Debe proveerse un nombre para la variable que la identifique en el modelo, por lo que este debe ser único. Además debe proveerse una lista de los valores que la variable puede tomar. La naturaleza de las variables es discreta y estas solo pueden tomar valores numéricos. Esta clase no se utiliza directamente sino que es heredada por las clases *ExognousVariable* y *EndogenousVariable*.

La clase *ExogenousVariable* representa a las variables exógenas de un SCM. Para crear una variable exógena se debe especificar su nombre, los valores que toma y su distribución de probabilidad. Dicha distribución consiste en una lista P donde cada valor $P[i]$ se corresponde con la probabilidad de que la variable tome el valor $V[i]$, donde V es la lista de valores provista. La lista debe cumplir con todas las propiedades de una distribución de probabilidad:

- La suma de sus valores debe ser igual a 1
- Cada valor debe ser encontrarse en el intervalo $[0, 1]$

En el caso de las variables endógenas, representadas por la clase *EndogenousVariable* debe especificarse el nombre, los valores que toma, los padres de la variable (variables de las que depende) y la expresión que define a la misma. Es importante especificar todos los posibles valores que puede tomar la variable a partir de las variables de las que depende. Toda variable endógena debe depender de al menos otra variable y una de estas variables debe ser exógena. Esto último es necesario para el cálculo correcto de los contrafactuales, ya que

el método utilizado (redes gemelas) utiliza las variables exógenas para conectar las variables del mundo real con las del mundo alternativo.

La expresión que define a la variable será una cadena con una sintaxis similar a las expresiones de Python. Además se podrán usar funciones de la clase *math*, que viene integrada a dicho lenguaje. Ejemplo de posibles expresiones válidas son:

- X + Y - 1
- (X and Y) if Z == 0 else (X or Y)
- max(X, Y) - min(X,Y)

Las variables que aparecen en las expresiones deben corresponderse con los nombres de las variables de las que depende la variable que se está definiendo.

Una vez creadas las variables exógenas y endógenas, estas se agrupan en dos listas y se llama al constructor de la clase SCM, la cual representa a los modelos causales estructurales.

Por ejemplo, si se tiene un modelo como el siguiente:

Modelo 2.1.1

$$U = \{Ux, Uy, Uz\} \qquad V = \{X, Y, Z\} \qquad E = \{f_x, f_y, f_z\}$$

$$Val(Ux) = Val(Uy) = Val(X) = Val(Y) = \{0,1,2\}$$
$$Val(Uz) = \{0,1\}$$
$$Val(Z) = \{0,1,2,3,4\}$$

$$P(X = x) = \left\{ \begin{array}{ll} 0{,}1 & \text{si } x = 0 \\ 0{,}2 & \text{si } x = 1 \\ 0{,}7 & \text{si } x = 2 \end{array} \right\}.$$

$$P(Y = y) = \left\{ \begin{array}{ll} 0{,}2 & \text{si } y = 0 \\ 0{,}2 & \text{si } y = 1 \\ 0{,}6 & \text{si } y = 2 \end{array} \right\}.$$

$$P(Z = z) = \left\{ \begin{array}{ll} 0{,}8 & \text{si } z = 0 \\ 0{,}2 & \text{si } z = 1 \end{array} \right\}.$$

$$\begin{array}{ll} f_x: & X = Ux \\ f_y: & Y = Uy \end{array}$$

$$f_z : Z = \left\{ \begin{array}{ll} X + Y & \text{si } Uz = 0 \\ X * Y & \text{si } Uz = 1 \end{array} \right\}.$$

Entonces para construir dicho modelo se usaría el código:

```
Ux = ExogenousVariable("Ux", [0, 1, 2], [0.1, 0.2, 0.7])
Uy = ExogenousVariable("Uy", [0, 1, 2], [0.2, 0.2, 0.6])
Uz = ExogenousVariable("Uz", [0, 1], [0.8, 0.2])

X = EndogenousVariable("X", [0, 1, 2], [Ux], "X")
Y = EndogenousVariable("Y", [0, 1, 2], [Uy], "Y")
Z = EndogenousVariable("Z", [0, 1, 2, 3, 4], [X, Y, Uz], "X + Y if Uz
 == 0 else X * Y")

model = SCM([Ux, Uy, Uz], [X, Y, Z])
```

Los modelos también tienen una representación en formato JSON (Javascript Object Notation) con el objetivo de persisitirlos en ficheros y así poder usarlos posteriormente. El método *export_to_json* convierte un objeto de tipo SCM a su representación en JSON, mientras que el método *import_from_json* crea un objeto de tipo SCM a partir de la representación en JSON correspondiente. A su vez, los métodos *export_to_json_file* e *import_from_json_file* permiten guardar y cargar de memoria externa respectivamente.

2.1.2. Algoritmos de inferencia

La implementación de los algoritmos de inferencia causal se basa en la inferencia bayesiana. A la hora de responder a una pregunta causal se aplica un algoritmo que en algún punto pasa por la construcción de una red bayesiana y la ejecución de un algoritmo de inferencia para obtener la respuesta deseada.

Para el trabajo con redes bayesianas se utilizó la librería **pgmpy** [37]. Esta es una librería de código abierto para el trabajo con redes bayesianas, enfocada en la modularidad y la extensibilidad. Contiene implementaciones de algoritmos de estimación de parámetros, aprendizaje de estructura, inferencia exacta, aproximada y además inferencia causal.

El método *build_bayesian_network* construye una red bayesiana a partir de un modelo causal estructural, mediante un proceso similar al descrito en la sección 1.7.1. La parte más compleja de dicho algoritmo consiste en la creación de las tablas de probabilidad de las variables endógenas, donde es necesario computar todas las combinaciones posibles de valores que pueden tomar los padres más la variable. Así, si una variable X depende de un conjunto $Pa_X = \{P_1, P_2, ..., P_n\}$ de variables, su tabla de probabilidad contaría con $|X| \cdot \prod_{i=1}^{n} |P_i|$ entradas, donde $|P_i|$ denota la cardinalidad de la variable P_i y $|X|$ la cardinalidad de X. Para determinar el valor de cada probabilidad $P(X_i = x_i \mid Pa_X = x^*)$ se evalúa la expresión que define a X en la correspondiente asignación x^* a las variables padres de X. Si el resultado es igual a x_i, entonces la probabilidad toma valor 1. En caso contrario toma valor 0. Para evaluar la expresión se utiliza la función *eval* integrada a Python, que toma una expresión en forma de cadena de caracteres y los valores que se le asigna a las variables, y devuelve el resultado de evaluar dicha expresión.

Existen 3 tipos principales de inferencia en la implementación: predicciones,

intervenciones y contrafactuales. Las signaturas de los métodos correspondientes a dichas operaciones se definen a continuación:

```
def predict(self, variables, observations={}, map=False, joint=False,
 algorithm="BP"):...

def do_bayesian(self, variables, do_variables, observations={},
 map=False, joint=False, algorithm="BP"):...

def counterfactual(self, variables, do_variables, observations={},
 map=False, joint=False, algorithm="BP"):...
```

El método *predict* permite realizar una predicción en el modelo, devolviendo las probabilidades actualizadas de un conjunto de variables solicitadas (parámetro *variables*) a partir de un conjunto de variables observadas (*observations*). Por ejemplo, si se quiere determinar la probabilidad $P(Y \mid X = 1)$ se hece la llamada siguiente al método:

```
result = model.predict(["Y"], observations={"X": 1})
```

El método *do_bayesian* calcula el resultado de una intervención, incorporando el parámetro *do_variables* que indica las variables que son intervenidas y los valores que son forzados a tomar. Suponiendo que se quiere calcular $P(Y \mid do(X = 1), Z = 2)$ la llamada al método sería:

```
result = model.do_bayesian(["Y"], {"X": 1}, observations={"Z": 2})
```

Para calcular el resultado de la intervención se usa el procedimiento descrito en la sección 1.8.4: se construye el submodelo correspondiente a la intervención mediante el método *get_submodel*, y convirtiéndolo en red bayesiana con el método *build_bayesian_network* se aplica inferencia para obtener la respuesta deseada.

El método *counterfactual* calcula el resultado de un contrafactual, recibiendo los mismos parámetros que el método *do_bayesian* pero con un significado ligeramente distinto: las variables de la respuesta deben corresponder al mundo alternativo, para lo cual se debe añadir el caracter '*'al final del nombre de la variable, y lo mismo ocurre con las variables especificadas en *do_variables* que corresponden a las intervenciones realizadas en el mundo alternativo. Para calcular el contrafactual $P(Y_{X=1} \mid Z = z)$ la llamada al método *counterfactual* sería:

```
result = model.counterfactual(["Y*"], {"X*": 1},observations={"Z": 2})
```

Para calcular el contrafactual se utiliza el método de las redes gemelas descrito en la sección 1.9.4. Se construye la estructura mediante el método *build_twin_network* que a partir de un modelo construye el modelo de redes gemelas correspondiente. La respuesta al contrafactual se obtiene llamando al método *do_bayesian*

sobre el modelo de redes gemelas resultante.

Los métodos anteriores disponen de un conjunto de parámetros opcionales . El parámetro *map* especifica si en vez de realizar inferencia probabiliística se debe realizar inferencia MAP. El parámetro *joint* indica si se desea obtener en la respuesta la distribución de probabilidad conjunta o las distribuciones por separado. El parámetro *algorithm* indica el algoritmo de inferencia que será usado y puede tomar dos valores: "BP "(propagación de creencias) y "VE"(eliminación de variables).

En caso de que la inferencia especificada sea probabilística, los métodos *predict*, *do_bayesian* y *counterfactual* devuelven una función. Si es solicitada la distribución conjunta, esta recibe como parámetro una asignación en forma de diccionario de los valores que toma cada variable y devuelve la probabilidad correspondiente. Si no se pide la distribución conjunta, entonces la función devuelta recibe como parámetro el nombre de una variable de la respuesta y el valor que toma, devolviendo la probabilidad correspondiente. Por otro lado, si se especifica que la inferencia es de tipo MAP, entonces se devuelve un diccionario con la asignación más probable a las variables.

Por ejemplo, para el modelo 2.1.1, si deseamos obtener el valor de $P(Z_{X=1} = 3|Z = 3)$, la instrucción de código quedaría:

```
model.counterfactual(["Z*"], {"X*": 1}, {"Z": 3})("Z*", 3)
```

En cambio la probabilidad conjunta $P(Z = 3, Y = 2|do(X = 1))$ se obtendría así:

```
model.do_bayesian(["Z", "Y"], {"X": 1})({"Z": 3, "Y": 2})
```

Por otro lado la asignación más probable $\text{MAP}(Z, Y|do(X = 1))$ se obtendría leyendo el diccionario que devuelve el llamado al método *do_bayesian_map*:

```
result = model.do_bayesian(["Y", "Z"], {"X": 1}, map=True)
for variable, value in result.items():
   print(variable, value)
```

Como se vio en las secciones 1.10 y 1.11, las fórmulas de mediación y atribución se definen en términos de intervenciones y contrafactuales y por tanto su implementación es inmediata. En el caso de la atribución se reciben dos parámetros correspondientes al tratamiento(causa) y la respuesta(efecto). Cada uno de estos parámetros consiste en una tupla de tres elementos de la forma (nombre de la variable, valor verdadero, valor falso):

```
def prob_of_neccesity(self, cause, effect):...

def prob_of_sufficiency(self, cause, effect):...

def prob_of_neccesity_and_sufficiency(self, cause, effect):...
```

Por ejemplo si se tienen dos variables binarias X, Y, y se quiere calcular la

necesidad de tomar un medicamento($X = 1$) para recuperarse ($Y = 1$), entonces el llamado al método *prob_of_neccesity* sería:

```
model.prob_of_neccesity(self, ("X", 1, 0), ("Y", 1, 0))
```

Las fórmulas de mediación implementadas son el efecto total y el efecto directo controlado, que aplican para dos pares de variables cualesquiera:

```
def total_effect(self, cause, effect_variable):...

def controlled_direct_effect(self, cause, effect_variable,
 mediator):...
```

Ambas reciben un parámetro *cause* que consiste en una tupla (nombre de la causa, valor inicial, valor final) y otro parámetro *effect_variable* que indica el nombre de la variable en la que se mide el efecto. Si quisiésemos medir el efecto total de X en Y cuando X cambia de 0 a 1, entonces llamaríamos a la primera función con los parámteros:

```
model.total_effect(("X", 0, 1), "Y")
```

El efecto directo controlado recibe otro parámetro que consiste en una tupla (nombre, valor) correspondiente al mediador que se está analizando. Una posible llamada al método sería:

```
model.controlled_direct_effect(("X", 0, 1), "Y", ("Z", 0))
```

Aquí se calcula el efecto de X en Y cuando el mediador Z es obligado a tomar el valor 0.

2.1.3. Complejidad

La complejidad de los algoritmos de inferencia causal es exponencial ya que todos utilizan un algoritmo de inferencia bayesiana exacta que puede ser propagación de creencias [21], cuya complejidad es exponencial con respecto al tamaño del mayor clique del grafo moral triangulado, o eliminación de variables [22], cuya solución pasa por encontrar un conjunto de variables de eliminación óptimo, siendo este un problema NP-completo.

Definición 2.1.1 *Sea $G = \langle V, E \rangle$ un grafo dirigido donde $N = |V|$ y $M = |E|$. El coeficiente de densidad de G es:*

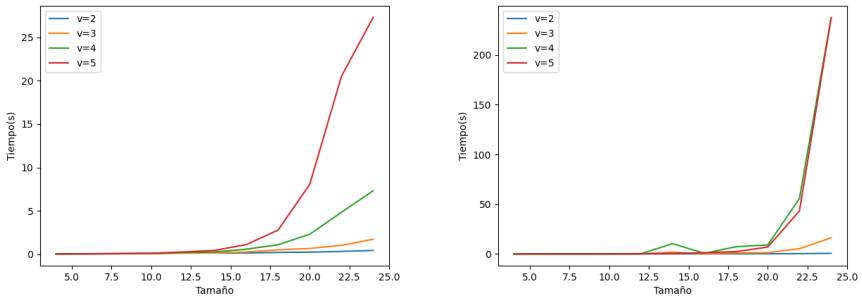
$$D_G = \frac{M}{N^2}$$

Definición 2.1.2 *Un grafo G es esparcido si $D_G \leq 0{,}08$. En caso contrario se dice que es denso.*

El tiempo de ejecución de los algoritmos depende de varios factores: el número de nodos o tamaño del grafo, la densidad del grafo y la cantidad máxima de valores que puede tomar cada variable. Se obtuvieron resultados satisfactorios para grafos esparcidos y con variables que toman un número pequeño de valores.

Los tiempos de ejecución obtenidos para las predicciones fueron muy similares a los de las intervenciones. Esto tiene sentido, puesto que una predicción es una intervención sobre un conjunto vacío de variables.

En las predicciones se comportó ligeramente mejor el algoritmo de propagación de creencias que el de eliminación de variables. La Figura 2.1 compara el comportamiento de los dos, tiendo en cuenta distintas cardinalidades máximas para las variables del modelo (valor de v). La Figura 2.2 muestra el comportamiento de las predicciones usando el algoritmo de propagación de creencias en modelos de hasta 60 variables donde todas son binarias.



(a) BP (b) VE

Figura 2.1: Comparación entre los algoritmos de propagación de creencias y eliminación de variables

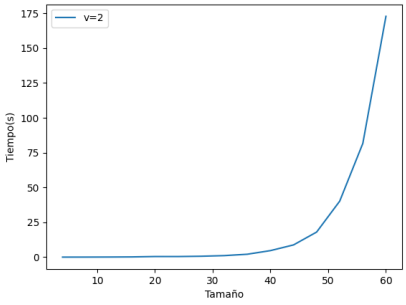


Figura 2.2: Comportamiento del algoritmo de propagación de creencias para modelos con sólo variables binarias

Los contrafactuales resultaron ser viables en redes más pequeñas que las de

las predicciones e intervenciones, pues el método de las redes gemelas transforma la red en una de casi el doble de su tamaño y realiza inferencia sobre ella. En la Figura 2.3 se muestra el comportamiento de los contrafactuales en dependencia del algoritmo que se use durante la fase de inferencia bayesiana en modelos con solo variables binarias.

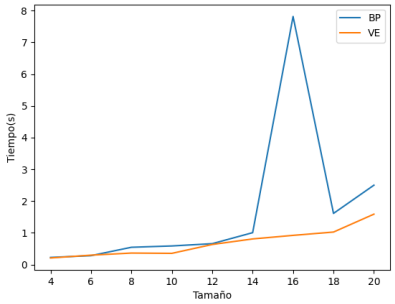


Figura 2.3: propagación de creencias vs eliminación de variables en contrafactuales

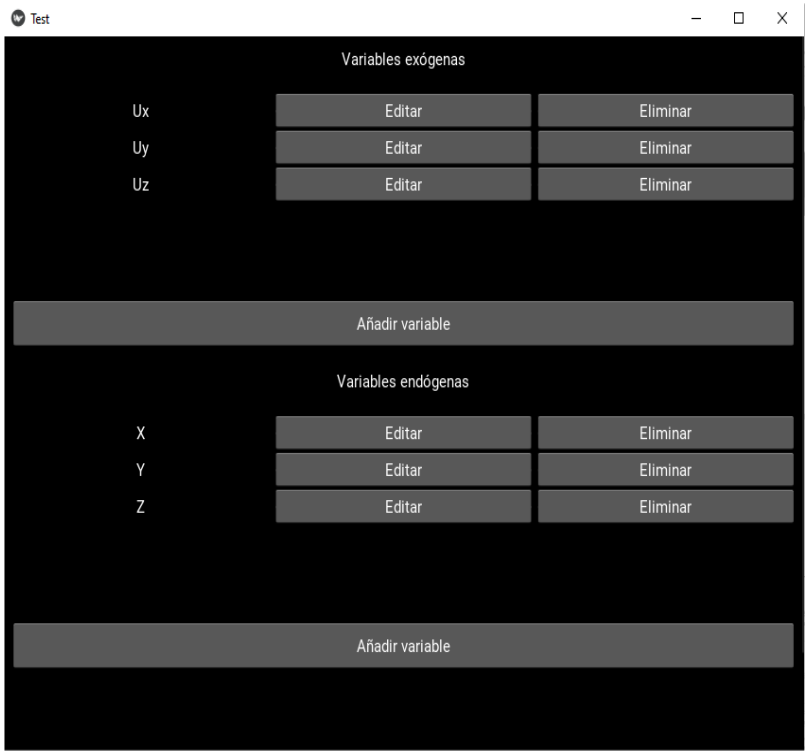
2.2. Interfaz de usuario

La interfaz de usuario está concebida para facilitar el uso de la inferencia causal en problemas que lo requieran, abstrayendo al usuario del ambiente de la programación. Se brinda al usuario la posibilidad de crear, editar, cargar y guardar modelos estructurales causales. Además es posible visualizar el grafo causal asociado. Se brindan 5 modalidades de inferencia causal: predicción, intervención, contrafactual, atribución y mediación. Para cada tipo de inferencia se provee un formulario donde el usuario especifica los parámetros de la consulta a realizar. Para asegurar que la entrada provista por los usuarios al programa es la correcta, en todos los formularios existen métodos de validación que notifican al usuario en caso de que la entrada sea incorrecta.

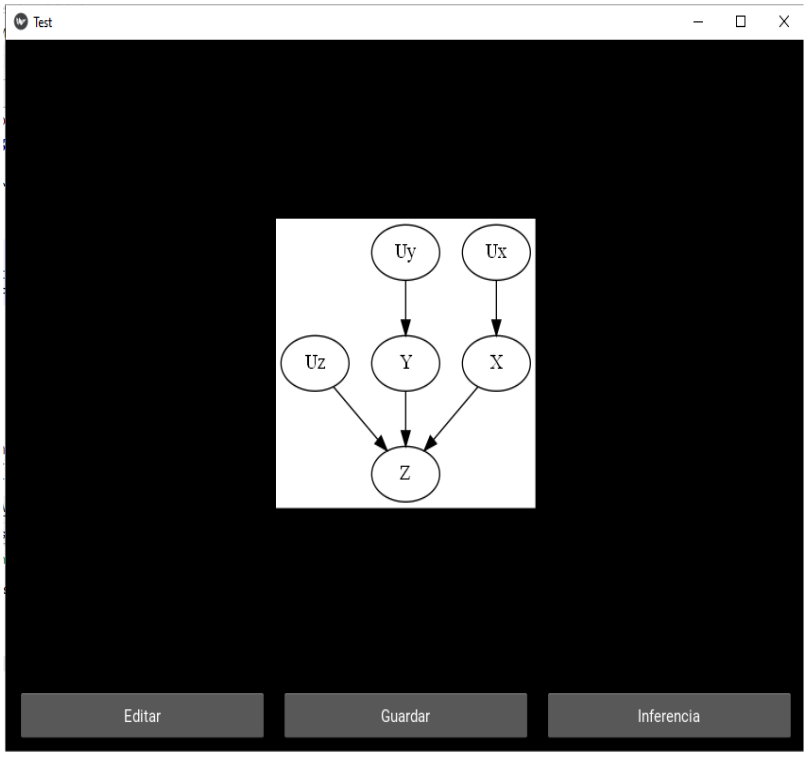
La interfaz visual fue desarrollada con la librería **kivy**[38]. Entre sus principales características se encuentran que es de código abierto, multiplataforma y eficiente. Permite la creación de aplicaciones de manera rápida y a diferencia de otras librerías para interfaces visuales como PyQt, ofrece una sintaxis más elegante, al estilo de Python.

Para la visualización de los grafos causales se utilizaron dos librerías: **pydot**[39] y **networkx**[40]. Ello se debe a que en algunas arquitecturas de sistemas operativos la primera, que es la preferida, requiere de la instalación de dependencias adicionales, y no se desea que esa tarea recaiga en manos del usuario. Por ello cuando se solicita la imagen de un grafo causal, se da prioridad a *pydot*, y en caso de que no sea posible utilizarla, se usa la segunda, que es más flexible.

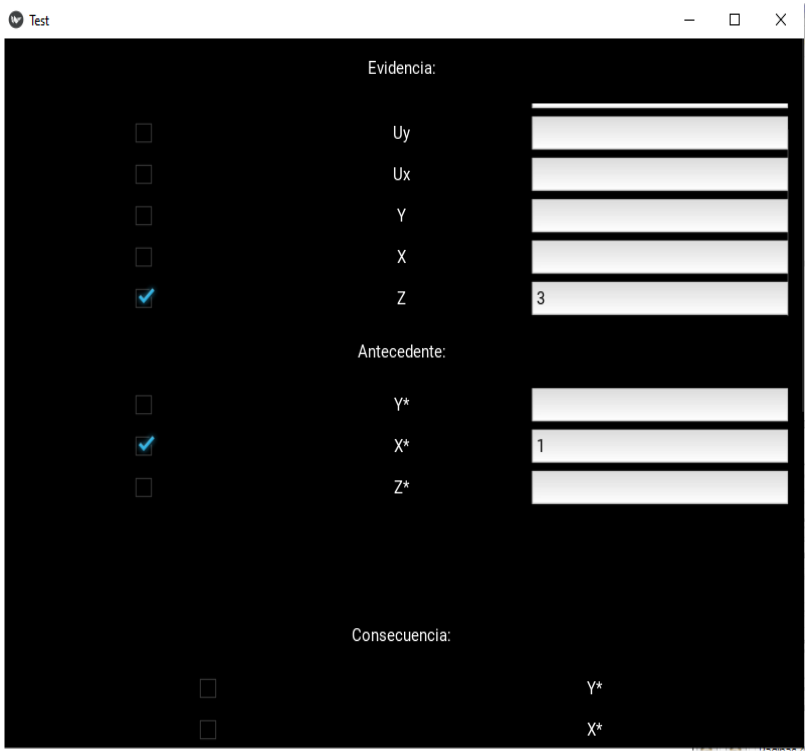
Por último, para el proceso de empaquetado del software se utilizó **PyInstaller**[41]. Esta librería agrupa el código y las dependencias en un paquete que puede ser usado en otras computadoras sin la necesidad de instalar un intérprete de Python ni las dependencias, siempre y cuando sea en un sistema operativo similar a donde se construyo el paquete.



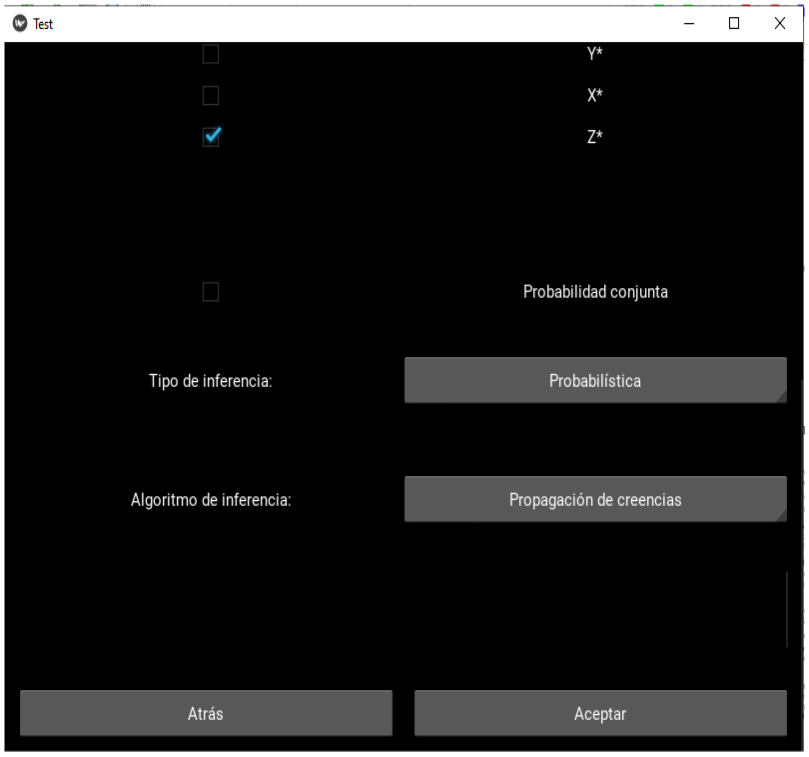
(a) Ventana de creación de un modelo causal estructural



(b) Ventana principal del modelo



(a)



(b)

(c) Formulario para calcular un contrafactual

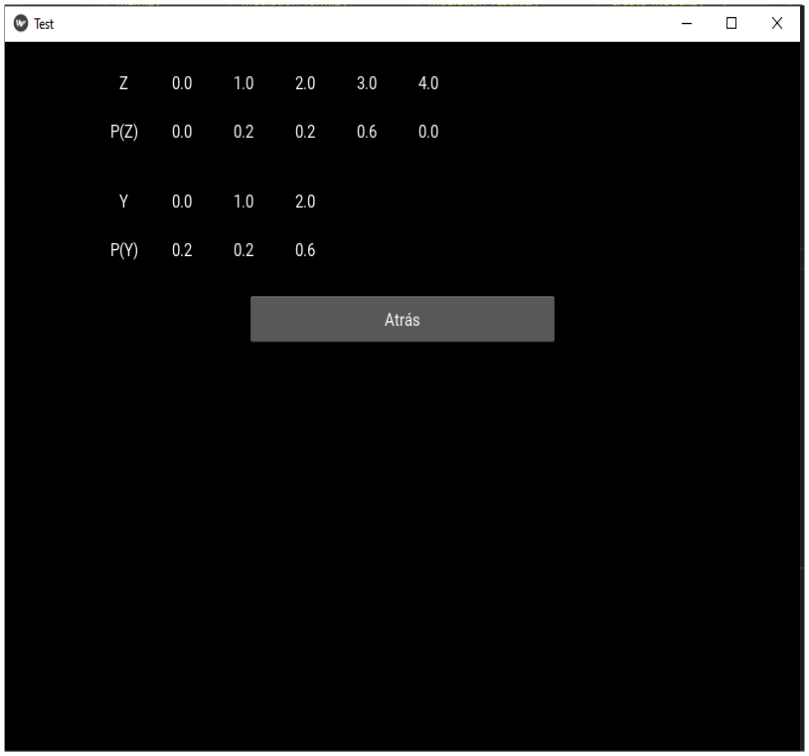


Figura 2.6: Resultados de una intervención

Capítulo 3

Aplicaciones

Con el objetivo de ejemplificar las posibles aplicaciones del programa desarrollado se proponen un conjunto de aplicaciones del análisis causal en diversas áreas de la ciencia. Los modelos usados se basan en los propuestos en un conjunto de artículos académicos donde se emplea el anális causal para proponer soluciones a problemas relacionados con la medicina, la epidemiología y la psicología ([42, 43, 44]). Estos fueron simplificados y adaptados para satisfacer las restricciones del programa implementado. Todos los resultados mostrados en las siguientes secciones se corresponden con las salidas provistas por el programa.

3.1. Control de la COVID-19

La causalidad ha resultado ser especialmente útil en el área epidemiológica. Desde la aparición de la COVID-19 se han presentado trabajos donde se utiliza la inferencia causal para determinar qué factores disminuyen el riesgo de contagio y cuáles lo incrementan. [45, 44].

Supongamos que se quiere poner en práctica un conjunto de medidas en una ciudad destinadas a frenar un brote de COVID-19. Dado que dichas medidas demandan recursos y no se conoce cuán efectivas serán, parece razonable utilizar una herramienta para estimar el resultado de su aplicación.

Primeramente es necesario construir un modelo que declare explícitamente las relaciones de causalidad entre las variables de estudio. Este modelo puede obtenerse de dos maneras. La primera consiste en, utilizando un algoritmo, encontrar relaciones de causalidad subyacentes en los datos. Esta variante es conocida como *descubrimiento causal*. La segunda vía consiste en construir el modelo a partir de un conjunto de presuposiciones producto de la experiencia o la intuición. Se optará por la segunda variante.

En la Tabla 3.1 se muestran las variables tenidas en cuenta para la construcción del modelo. Todas las variables toman valores numéricos. En los casos en que la variable sea binaria, el valor 1 indica que la variable toma valor verdadero. En las variables categóricas, cada valor tiene asociado una categoría de

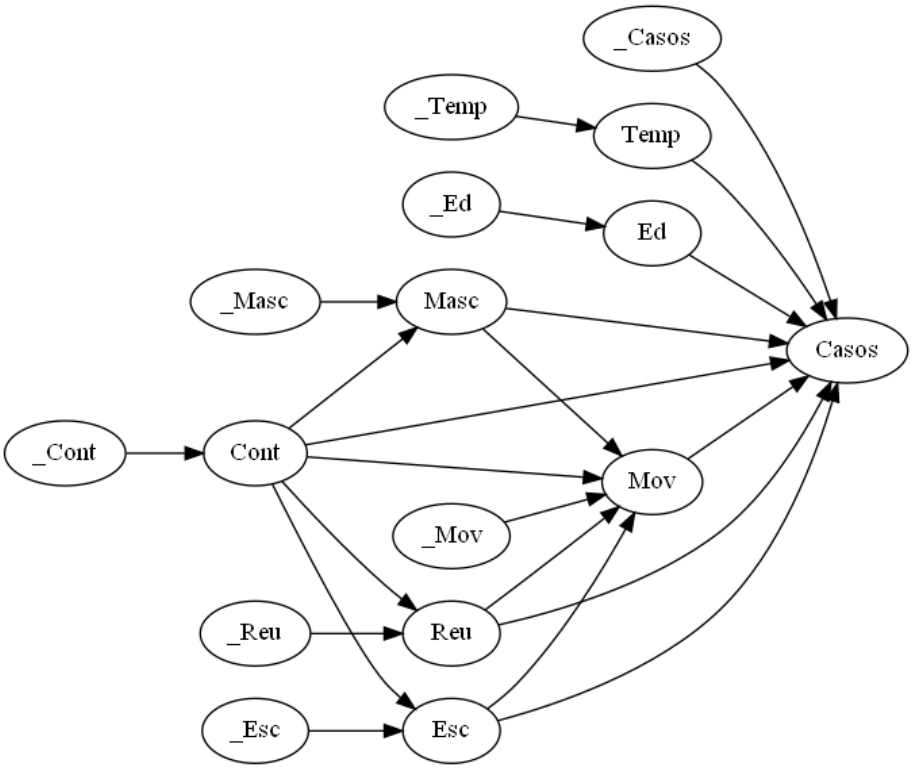


Figura 3.1: Grafo causal asociado al modelo sobre la COVID-19

la variable. Cada variable endógena tiene asociada un término de error o variable exógena cuyo nombre es el mismo pero con el prefijo "_"añadido. Estas variables no son incluidas en la tabla por simplicidad pero forman parte de la función de definición de las variables y añaden no-determinismo al modelo.

A partir de los datos extraídos se pueden obtener las marginales de las variables. La tabla 3.2 muestra la correspondiente al número de casos reportados. Como se observa existe una tendencia a diagnosticar un número alto de casos en la ciudad. Para contrarrestar este comportamiento se debe tomar un conjunto de medidas. Por ejemplo, establecer el uso obligatorio de mascarillas puede reducir el riesgo de brotes. El resultado de esta medida se puede calcular a partir de la intervención $P(Casos \mid do(Masc = 1))$ y se muestra en la tabla 3.3. Sin embargo, a pesar de que su aplicación disminuye el riesgo de un número alto de casos, no es suficiente. En el caso de que se aplicase solo esta medida y el número de casos siga siendo alto, cabría preguntarse que hubiera ocurrido si además se hubiesen tomado medidas adicionales, como cerrar las escuelas, restringir la movilidad y prohibir las reuniones en masa. El contrafactual $P(Casos_{Masc=1,Esc=1,Reu=1,Mov=0} \mid Masc = 1, Casos = 2)$ da respuesta a esta interrogante y se muestra en la tabla 3.4.

Nombre	Tipo	Valores	Descripción
Temp	Categórica	0(baja), (1) media, (2) alta	Temperatura
Ed	Categórica	0 (niño), 1(joven), 2(adulto), 3(anciano)	Edad
Gen	Categórica	0(Hombre), 1(Mujer)	Género
DP	Categórica	0(baja), 1(media), 2(alta)	Densidad de población
Cont	Categórica	0(bajo), 1(medio), 2(alto)	Nivel de contagio
Masc	Binaria	0,1	Uso obligatorio de mascarilla
Reu	Binaria	0,1	Prohibición de reuniones masivas
Esc	Binaria	0,1	Cierre de las escuelas
Mov	Categórica	0(bajo), 1(medio), 2(alto)	Nivel de movilidad
Casos	Categórica	pocos, normal, muchos	Número de casos reportados

Tabla 3.1: Variables endógenas del modelo de la COVID-19

v	pocos	normal	muchos
P(v)	0.1372	0.2683	0.5945

Tabla 3.2: Marginal correspondiente a la variable Casos

v	bajo	medio	alto
P(v)	0.196	0.3046	0.4994

Tabla 3.3: Marginal correspondiente a la variable Casos resultante de calcular la intervención $P(Casos \mid do(Masc = 1))$

v	bajo	medio	alto
P(v)	0.6432	0.1167	0.2402

Tabla 3.4: Marginal correspondiente a la variable Casos resultante de calcular el contrafactual $P(Y_{Masc=1,Esc=1,Reu=1,Mov=0} \mid Masc = 1, Casos = 2)$

3.2. Comportamiento antisocial en adolescentes

Con el objetivo de entender el comportamiento antisocial en los adolescentes se estudian un conjunto de factores que se sospecha que pueden estar relacionados a este. Las variables tenidas en cuenta son: la autoestima del individuo (SE), situaciones adversas durante la niñez (ACE), síntomas de hiperactividad y déficit de atención (ADHD) y el nivel de agresividad en el individuo (AGR). Para medir el valor de cada variable en un individuo se le aplican una serie de

tests, que devuelven una puntuación entre 0 y 4 para cada variable. Con estos datos, y con hipótesis a partir de estudios previos se construye un modelo cuyo diagrama corresponde al de la figura 3.2. A partir de los datos recopilados se determina la distribución que corresponde a los términos de error, así como la función que determina a cada variable.

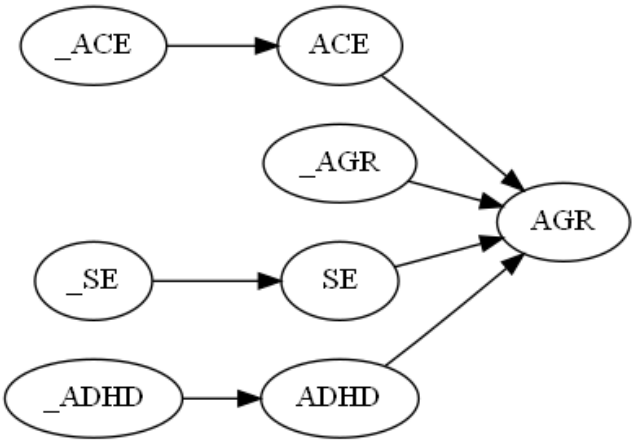


Figura 3.2: Grafo causal que explica el comportamiento agresivo en los adolescentes

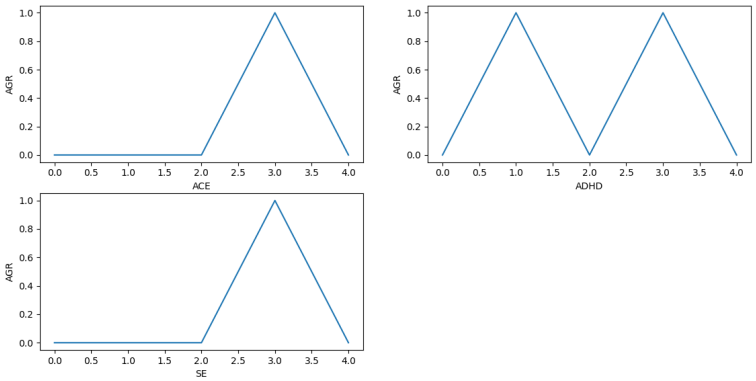


Figura 3.3: Efecto total de las variables sobre el comportamiento agresivo del individuo

Para determinar cuánto afecta cada variable al comportamiento del individuo se calcula el efecto total que ejerce cada variable sobre el comportamiento agresivo del individuo. En la figura 3.3 se muestra para cada variable V, por

cada valor v_i, el efecto total V sobre AGR cuando V cambia de v_i hacia v_{i+1}. Como se puede observar, los valores extremos de las variables no aumentan el comportamiento agresivo del individuo y sí los intermedios.

3.3. Factores de riesgo de arteriosclerosis

Con el objetivo de aclarar las relaciones de causalidad entre los factores de riesgo de arteriosclerosis, se construye el modelo representado en la figura 3.4. Las variables de interés son edad(Ed), consumo de cigarro(Sm), alcohol(Al), realización de ejercicio físico(Ex), obesidad(Ob), hiperlipidemia(Lipid) e hipertensión(Hyper). Todas son binarias excepto la edad, que toma cuatro valores posibles: niño, joven, adulto y anciano.

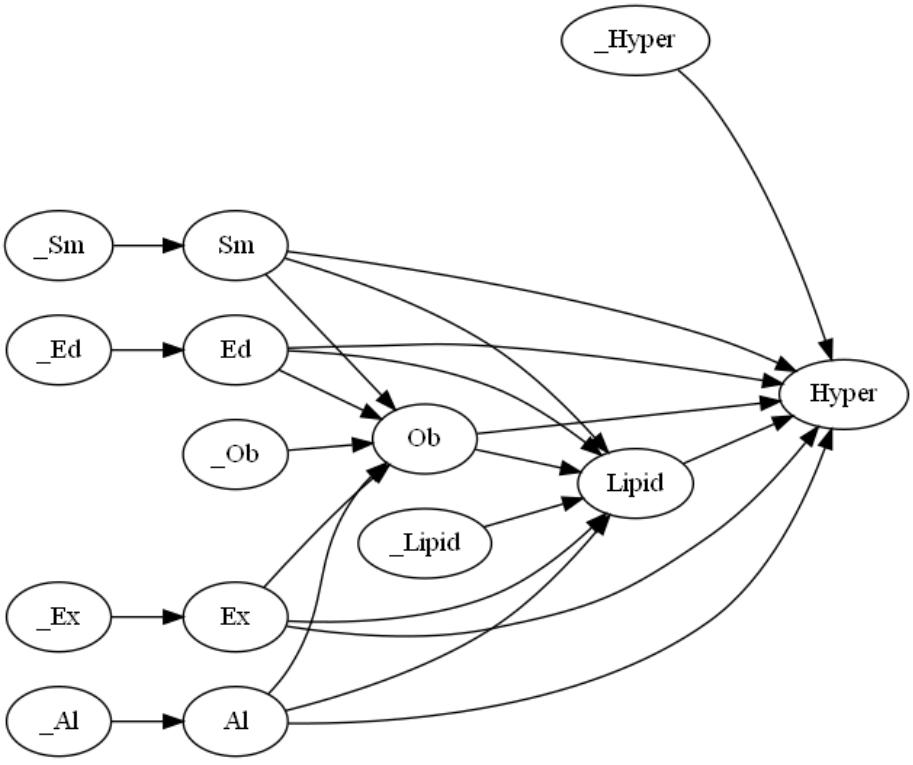


Figura 3.4: Grafo causal correspondiente al modelo que explica los factores de riesgo de arteriosclerosis.

Los valores que toman las marginales de las variables de obesidad, hiperlipidemia e hipertensión a partir de los datos recogidos y sin realizar ninguna observación sobre el modelo se recogen en la tabla 3.5.

V	P(V=0)	P(V=1)
Ob	0.574	0.426
Lipid	0.664	0.336
Hyper	0.5092	0.4908

Tabla 3.5: Marginales correspondientes a las variables Ob(Obesidad), Hyper(Hipertensión) y Lipid(Hiperlipidemia), obtenidas a partir de los datos recopilados

Alternativamente al efecto total, se puede utilizar la fórmula del *Efecto Causal Promedio*(ACE):

$$ACE = P(Y = 1 \mid do(X = 1)) - P(Y = 1 \mid do(X = 0))$$

El Efecto Causal Promedio se aplica entre dos variables binarias y compara el porciento de individuos que responden positivamente a un tratamiento con los que lo hacen si no se les aplica dicho tratamiento. Permite medir en términos de probabilidades cuánto influye una variable en otra. En dependencia del signo del resultado, se determina si este efecto es positivo o negativo y su módulo nos dice cuán relevante es este efecto. En la tabla 3.6 se muestra el valor de ACE entre cada una de las variables de Hipertensión, Hiperlipidemia y Obesidad y las variables de Ejercicio, Alcohol y Fumar. Como se puede apreciar, el consumo de cigarro y alcohol incrementa el riesgo de contraer Hipertensión, Hiperlipidemia y Obesidad mientras que la práctica de ejercicio físico lo disminuye.

	Ob	Lipid	Hyper
Al	0.524	0.3872	0.4952
Sm	0.365	0.32	0.422
Ex	-0.115	-0.265	-0.172

Tabla 3.6: Efecto Causal Promedio (ACE) que ejercen las variables Al, SM y Ex sobre los factores de riesgo.

Por último, aprovechando que la mayoría de las variables son binarias, se puede realizar un análisis de atribución para determinar cuán necesarias y suficientes son las variables causa (Al, Ex, Sm) para determinar los valores de las variables efecto (Ob, Hyper, Lipid). En las tablas 3.7 y 3.8 se muestran los valores de PN(probabilidad de necesidad), PS(probabilidad de suficiencia) y PNS(probabilidad de necesidad y suficiencia) para las causas Al y Sm respectivamente. En el caso de la variable Ex, todas las probabilidades tienen valor cero, lo cual tiene sentido teniendo en cuenta que el ACE dio negativo en todos los casos. Esto significa que la práctica de ejercicio físico no nos permite afirmar que existirán factores de riesgo y que la existencia de factores de riesgo no implica que el individuo realice ejercicio físico.

	PN	PS	PNS
Ob	0.7616	0.6268	0.5240
Lipid	0.7311	0.4515	0.3872
Hyper	0.6706	0.6543	0.4952

Tabla 3.7: Análisis de atribución sobre la variable Al

	PN	PS	PNS
Ob	0.5659	0.5069	0.3650
Lipid	0.6061	0.404	0.3200
Hyper	0.5672	0.6224	0.4220

Tabla 3.8: Análisis de atribución sobre la variable Sm

Conclusiones

Se realizó un recorrido general por la historia del desarrollo matemático y filosófico de la causalidad. El problema recayó primeramente en manos de los filósofos y comenzó a despertar el interés de algunos matemáticos a principios de la época moderna. Sin embargo fue abandonado tempranamente dado que estos desviaron su atención hacia la estadística, área de la matemática que también daba sus primeros pasos por aquel entonces. La evasión de la causalidad por parte de los estadísticos condujo al surgimiento de numerosas paradojas y problemas intratables por la falta de las herramientas adecuadas para encararlos, herramientas que no podía proporcionar la visión frecuentista de la estadística. La historia de la concepción matemática de la causalidad tiene su punto de inflexión con la publicación de Sewall Wright de los diagramas de caminos. Posteriormente comenzaron a desarrollarse diversas teorías y modelos que explican la causalidad, con Reichenbach, Suppes, Granger y Pearl entre sus principales exponentes.

Se expusieron los desarrollos de Judea Pearl en la teoría de la causalidad y los modelos gráficos probabilistas. Se repasó la teoría de la independencia condicional como base para el desarrollo de los modelos gráficos probabilistas. Se presentaron las redes bayesianas como un primer modelo capaz de realizar predicciones en las variables a partir de observaciones. Posteriormente se introdujo el modelo causal estructural de Pearl como modelo gráfico probabilista capaz de representar explícitamente las relaciones de causalidad entre variables y su relación con las redes bayesianas. A continuación se pasó a introducir las distintas modalidades de inferencia causal propias de un SCM. Primeramente se formalizó el concepto de intervención, definiendo el operador *do* y resaltando sus diferencias con la observación pasiva. A continuación se vieron los principales métodos de inferencia causal para el cálculo de intervenciones: el criterio de la puerta trasera, el criterio de la puerta principal, el cálculo-do y la inferencia bayesiana. Luego se expuso la teoría relacionada con los contrafactuales, sus variantes determinista y no determinista y los principales métodos para calcularlos, haciendo especial hincapié en el método de las redes gemelas. Por último se expusieron dos aplicaciones de las intervenciones y los contrafactuales conocidos como atribución y mediación. La primera permite explicar el comportamiento de una variable a partir del de otras y la segunda, el análisis del efecto que ejerce una variable sobre otra a través de un mediador.

Se desarrolló una implementación de los algoritmos de inferencia causal,

utilizando la idea de la cirugía en el grafo para intervenir el modelo y el método de las redes gemelas para calcular los contrafactuales, conjuntamente con la idea de transformar el modelo a una red bayesiana para aplicar inferencia bayesiana y obtener la respuesta a la consulta deseada.

Los algoritmos de inferencia causal se comportaron satisfactoriamente para modelos cuyos grafos son esparcidos y las variables toman una cantidad relativamente pequeña de valores. Sin embargo, cuando el número de nodos, estados y aristas del grafo causal crece notablemente, el tiempo de ejecución se ve comprometido, especialmente en el caso de los contrafactuales, en los cuales la inferencia es aplicada en un grafo que en el caso peor ocupa el doble de tamaño que el grafo original. La técnica de mezcla de nodos resultó ser útil para disminuir el costo computacional de calcular los contrafactuales.

Además se desarrolló una interfaz visual destinada a usuarios no expertos en programación para la resolución de problemas de naturaleza causal.

Por último se mostraron algunas aplicaciones de la teoría de la causalidad en las que puede ser utilizado el programa implementado.

La causalidad ha demostrado ser una herramienta útil para complementar el procesamiento estadístico de los datos. Su uso permite la modelación y el entendimiento de procesos que la asociación no captura. Simular intervenciones a partir de datos recopilados, en vez de llevar a cabo la intervención en la práctica, permite no solo el ahorro de recursos, sino también conocer el resultado de intervenciones que serían imposibles o inviables de determinar experimentalmente.

Para concluir, se debe resaltar que la simulación del pensamiento causal nos colocará un paso más cerca de simular el pensamiento humano. Dotar de razonamiento causal a las máquinas permitirá que estas adquieran capacidades distintivas del ser humano, como son planear acciones con antelación o aprender de los propios errores. Conforme surjan nuevas teorías que mejoren, extiendan y complementen las actuales y se desarrollen modelos más completos y realistas capaces de razonar casualmente, seremos capaces no solo de automatizar tareas complejas, sino además de entendernos mejor a nosotros mismos.

Referencias

[1] Yuval Noah Harari. Sapiens: A brief history of humankind by yuval noah harari. *The Guardian*, 2014. 1

[2] Judea Pearl and Dana Mackenzie. *The Book of Why*. Basic Books, New York, 2018. 1, 3, 4, 6, 25

[3] Metaphysics Aristotle and H Book. 1045a 8–10 aristotle in 23 volumes, vols. 17, 18, translated by tredennick hugh. *Lond William Heine Mann Ltd*, 1933. 2

[4] David Hume. A treatise of human nature by david hume, reprinted from the original edition in three volumes and edited, with an analytical index, by la selby-bigge, ma, revised by ph nidditch in 1978, 1896. 2

[5] Francis Galton. Regression towards mediocrity in hereditary stature. *The Journal of the Anthropological Institute of Great Britain and Ireland*, 15:246–263, 1886. 2

[6] Sewall Wright. The relative importance of heredity and environment in determining the piebald pattern of guinea-pigs. *Proceedings of the National Academy of Sciences of the United States of America*, 6(6):320, 1920. 3

[7] Christopher Hitchcock and Miklós Rédei. Reichenbach's Common Cause Principle. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Summer 2021 edition, 2021. 4

[8] Patrick Suppes. *A Probabilistic Theory of Causality*. Amsterdam: North-Holland Pub. Co., 1968. 4

[9] Richard Otte. A critique of suppes' theory of probabilistic causality. *Synthese*, pages 167–189, 1981. 5

[10] C. W. J. Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37(3):424–438, 1969. 5

[11] Wikipedia contributors. Granger causality — Wikipedia, the free encyclopedia. `https://en.wikipedia.org/w/index.php?title=Granger_`

causality&oldid=1049120222, 2021. [Online; accessed 19-October-2021]. 5

[12] Judea Pearl. Bayesian netwcrks: A model cf self-activated memory for evidential reasoning. In *Proceedings of the 7th conference of the Cognitive Science Society, University of California, Irvine, CA, USA*, pages 15–17, 1985. 6, 14

[13] Judea Pearl, Madelyn Glymour, and Nicholas P Jewell. *Causal inference in statistics: A primer*. John Wiley & Sons, 2016. 6, 21, 22, 23, 26, 28, 32, 33

[14] Judea Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1988. 15

[15] Haipeng Guo and William Hsu. A survey of algorithms for real-time bayesian network inference. In *Join Workshop on Real Time Decision Support and Diagnosis Systems*, 2002. 15, 16

[16] Gregory F. Cooper. The computational complexity of probabilistic inference using bayesian belief networks. *Artificial Intelligence*, 42(2):393–405, 1990. 15

[17] Paul Dagum and Michael Luby. Approximating probabilistic inference in bayesian belief networks is np-hard. *Artificial Intelligence*, 60(1):141–153, 1993. 15

[18] Solomon Eyal Shimony. Finding maps for belief networks is np-hard. *Artificial Intelligence*, 68(2):399–410, 1994. 15

[19] Ashraf M. Abdelbar and Sandra M. Hedetniemi. Approximating maps for belief networks is np-hard and other theorems. *Artificial Intelligence*, 102(1):21–38, 1998. 15

[20] Judea Pearl. Fusion, propagation, and structuring in belief networks. *Artificial Intelligence*, 29(3):241–288, 1986. 15

[21] Steffen L Lauritzen and David J Spiegelhalter. Local computations with probabilities on graphical structures and their application to expert systems. *Journal of the Royal Statistical Society: Series B (Methodological)*, 50(2):157–194, 1988. 15, 41

[22] Nevin L Zhang and David Poole. A simple approach to bayesian network computations. In *Proc. of the Tenth Canadian Conference on Artificial Intelligence*, 1994. 15, 41

[23] Ross D. Shachter. Evidence absorption and propagation through evidence reversals. In Max HENRION, Ross D. SHACHTER, Laveen N. KANAL,

and John F. LEMMER, editors, *Uncertainty in Artificial Intelligence*, volume 10 of *Machine Intelligence and Pattern Recognition*, pages 173–190. North-Holland, 1990. 15

[24] Max Henrion. An introduction to algorithms for inference in belief nets. In Max HENRION, Ross D. SHACHTER, Laveen N. KANAL, and John F. LEMMER, editors, *Uncertainty in Artificial Intelligence*, volume 10 of *Machine Intelligence and Pattern Recognition*, pages 129–138. North-Holland, 1990. 15

[25] Zhaoyu Li and Bruce D'Ambrosio. Efficient inference in bayes networks as a combinatorial optimization problem. *International Journal of Approximate Reasoning*, 11(1):55–81, 1994. 16

[26] Frank Jensen and SK Anderson. Approximations in bayesian belief universe for knowledge based systems. *arXiv preprint arXiv:1304.1101*, 2013. 16

[27] Robert A Van Engelen. Approximating bayesian belief networks by arc removal. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(8):916–920, 1997. 16

[28] Uffe Kjærulff. Reduction of computational complexity in bayesian networks through removal of weak dependences. In *Uncertainty Proceedings 1994*, pages 374–382. Elsevier, 1994. 16

[29] Michael P Wellman and Chao-Lin Liu. State-space abstraction for anytime evaluation of probabilistic networks. In *Uncertainty Proceedings 1994*, pages 567–574. Elsevier, 1994. 16

[30] Max Henrion. Search-based methods to bound diagnostic probabilities in very large belief nets. In *Uncertainty Proceedings 1991*, pages 142–150. Elsevier, 1991. 16

[31] David Poole. The use of conflicts in searching bayesian networks. In *Uncertainty in Artificial Intelligence*, pages 359–367. Elsevier, 1993. 16

[32] David Poole. Probabilistic conflicts in a search algorithm for estimating posterior probabilities in bayesian networks. *Artificial Intelligence*, 88(1-2):69–100, 1996. 16

[33] Logan Graham, Ciarán M Lee, and Yura Perov. Copy, paste, infer: A robust analysis of twin networks for counterfactual inference. 29

[34] A Balke and J Pearl. Probabilistic counterfactuals: Semantics, computation. Technical report, and applications, 1995. 29

[35] Ilya Shpitser and Judea Pearl. What counterfactuals can be tested. *arXiv preprint arXiv:1206.5294*, 2012. 30

[36] Judea Pearl. *Causality*. Cambridge University Press, 2 edition, 2009. 32

[37] Ankur Ankan and Abinash Panda. pgmpy: Probabilistic graphical models using python. In *Proceedings of the 14th Python in Science Conference (SCIPY 2015)*. Citeseer, 2015. 38

[38] `https://kivy.org/doc/stable/`. 43

[39] `https://github.com/pydot/pydot`. 43

[40] `https://networkx.org/`. 43

[41] `http://www.pyinstaller.org/`. 44

[42] Naomi Matsuura, Toshiaki Hashimoto, and Motomi Toichi. A structural model of causal influence between aggression and psychological traits: Survey of female correctional facility in japan. *Children and Youth Services Review*, 31(5):577–583, 2009. 48

[43] HyunSoo Oh and WhaSook Seo. Development of a structural equation model for causal relationships among arteriosclerosis risk factors. *Public Health Nursing*, 18(6):409–417, 2001. 48

[44] Edgar Steiger, Tobias Mussgnug, and Lars Eric Kroll. Causal graph analysis of covid-19 observational data in german districts reveals effects of determining factors on reported case numbers. *PloS one*, 16(5):e0237277, 2021. 48

[45] Matteo Bonvini, Edward Kennedy, Valerie Ventura, and Larry Wasserman. Causal inference in the time of covid-19. *arXiv preprint arXiv:2103.04472*, 2021. 48