Modelos causales

Stanford Encyclopedia of Philosophy

June 11, 2019

Overview

1. Introducción
2. Conceptos básicos
3. Modelos deterministas de ecuaciones estructurales (SEM)
4. Modelos causales probabilísticos

Introducción

- Los modelos causales son modelos matemáticos que representan relaciones causales dentro de un sistema individual o un población.
- Un modelo causal hace predicciones sobre el comportamiento de un sistema.
- Un modelo causal implica el valor verdadero, o la probabilidad, de afirmaciones contrafactuales sobre un sistema; predice los efectos de las intervenciones; e implica la dependencia o independencia probabilística de variables incluidas en el modelo.
- Los modelos causales también facilitan la inversa de estas inferencias: si observamos correlación probabilística entre variables, o las salidas de intervenciones experimentales, podemos determinar cuales modelos causales son consistentes con estas observaciones.

Variables

- Las *variables* son los bloques de construcción más básicos de los modelos causales. Una variable es una función que puede tomar una amplia variedad de valores. Los valores de una variables pueden representar la ocurrencia o no ocurrencia de un evento, un rango de eventos incompatibles, una propiedad de un individuo o una población de individuos, o un valor cualitativo.
- El conjunto de valores posibles de una variables es el *rango* de la variable.
- Un *mundo* es una especificación completa de un modelo causal. Un mundo incluirá, entre otras, una asignación de valores a todas las variables del modelo.
- Si X es una variable en un modelo causal, y x es un valor particular en el rango de X, entonces $X = x$ es una *proposición atómica*.

Probabilidad

- Probabilidad condicional de A dado B

$$P(A|B) = \frac{P(A,B)}{P(B)}.$$

- A y B son independientes probabilísticamente sólo en el caso $P(A,B) = P(A)P(B)$.
- Las variables X y Y son independientes sólo en caso de que todas las proposiciones de la forma $X = x$ y $Y = y$ sean independientes.
- A y B son condicionalmente independientes dado C sólo si $P(A,B|C) = P(A|C)P(B|C)$. Si $P(B,C) > 0$, esto es equivalente a $P(A|B,C) = P(A|C)$.

Grafos

- Si **V** es un conjunto de variables incluidas en un modelo causal, una manera de representar las relaciones causales entre las variables en **V** es usando un grafo.
- Una arista dirigida de Y a Z en una DAG representa que Y es una causa directa de Z.
- Otro tipo de grafos además de los DAG, se conocen como grafo mixto acíclico dirigido (ADMG). Un ADMG contiene aristas con flechas por ambos lados, así como de un solo lado.
- Una arista de doble cabeza representa una *causa común latente*. Una causa común latente de las variables X y Y es una causa común que no está incluida en el conjunto de variables **V**. Un grafo sobre **V** debe contener un arista con doble cabeza entre X y Y cuando existe una variable L omitida en **V**, tal que si L fuera añadida a **V**, sería una causa directa de X y Y.

SEM I

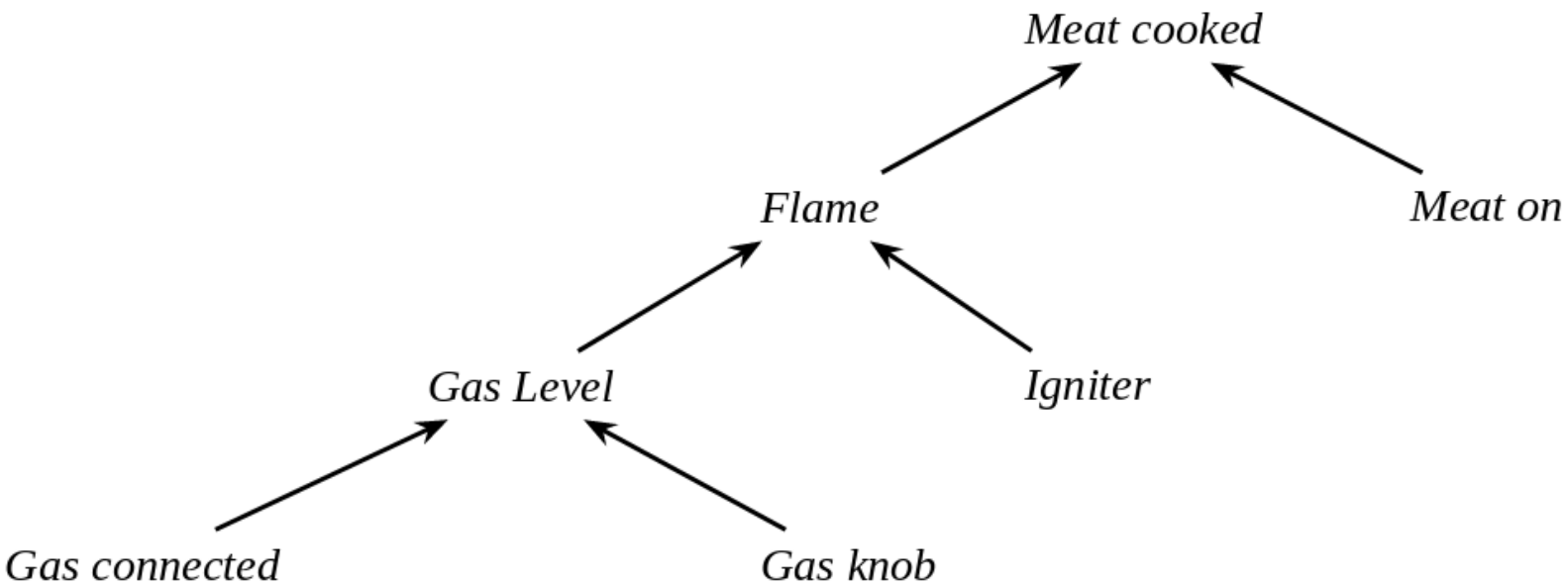
Un SEM caracteriza un sistema causal con un conjunto de variables, y un conjunto de ecuaciones describiendo como cada variable depende de sus predecesores causales inmediatos. Considera una parrilla de gas, utilizada para cocinas carne. Podemos describir las operaciones de la parrilla usando las siguientes variables:

- Gas conectado (1 si sí, 0 si no).
- Perilla de gas (0 apagada, 1 bajo, 2 medio, 3 alto).
- Nivel de gas (0 apagado, 1 bajo, 2 medio, 3 alto).
- Encendedor (1 presionado, 0 no).
- Flama (0 apagada, 1 baja, 2, media, 3 alta).
- Carne puesta (0 no, 1 si).
- Carne cocida (0 cruda, 1 rara, 2 medio, 3 bien).

Por ejemplo, las ecuaciones pueden ser:

SEM II

- Gas = Gas conectado x Perilla de gas
- Flama = nivel de gas x encendedor
- carne cocida = flama x carne puesta



SEM III

- Variables endógenas: sus valores están determinados por otras variables en el modelo.
- Variables exógenas: sus valores son determinados fuera del sistema.
- El contenido distintivo causal o estructural de un SEM deriva a partir de la manera en que las *intervenciones* son representadas. Intervenir una variable es establecer el valor de esa variable por un proceso que sobreescribe la estructura causal habitual, sin interferir con el proceso causal que gobierna a las otras variables.
- Una intervención sobre la variable X anula la ecuación normal para X, mientras no cambia las otras ecuaciones.
- Para representar una intervención sobre una variable, reemplazamos la ecuación para esa variable con una nueva ecuación indicando el valor con el cual la variable se define.

Contrafactuales estructurales

- Un contrafactual es una proposición en la forma de un condicional subjuntivo. El antecedente postula alguna circunstancia, a menudo una contraria al hecho.
- Los SEM deterministas naturalmente dan lugar a una lógica de contrafactuales. Estos contrafactuales son llamados *contrafactuales estructurales* o *contrafactuales intervencionistas*. Los contrafactuales estructurales son similares a los llamados *contrafactuales sin retroceso*. En un contrafactual sin retroceso, uno no razona hacia atrás desde una suposición contrafactual para sacar conclusiones sobre las causas de la situación hipotética.
- Se puede interpretar el contrafactual condicional $A\Box \rightarrow B$ como decir que B sería verdad, si A fuera verdad por una intervención.

Algunas características de los contrafactuales estructurales I

- Si el antecedente es una conjunción de proposiciones atómicas se reemplazan todas las ecuaciones relevantes.
- Un caso especial ocurre cuando el antecedente es una conjunción de proposiciones atómicas que asignan valores diferentes a la misma variable. En este caso, el antecedente es una contradicción, y el contrafactual se considera trivialmente verdad.
- Si el antecedente es una disyunción de proposiciones atómicas, o una disyunción de conjunciones de proposiciones atómicas, entonces la consecuencias debe ser verdad cuando se realizan todas las posibles intervenciones o conjuntos de intervenciones descritas por el antecedente.
- Algunas negaciones son tratadas como disyunciones.

Algunas características de los contrafactuales estructurales II

- El antecedente de un contrafactual siempre se considera realizado por una intervención, incluso si el antecedente es verdad en un mundo dado.
- Los valores verdaderos de contrafactuales son determinados solamente por las estructuras causales de los mundos, junto con las intervenciones especificadas en sus antecedentes. Ninguna otra consideración de similitud juega un papel.
- El análogo del modus ponens falla para el condicional estructural: i.e., a partir de A y $A\Box\rightarrow B$ no se puede inferir B.
- La sustitución de proposiciones lógicas equivalentes en el antecedente de un contrafactual no siempre preservan el valor verdadero.

Modelos causales probabilísticos

- La probabilidad puede ser utilizada para representar incertidumbre acerca del valor de variables no observadas en un caso particular, o la distribución de valores de variables en una población.
- Estamos interesados cuando alguna característica de la estructura causal de un sistema puede ser *identificada* a partir de la distribución de probabilidad sobre valores de variables, quizás en conjunción con supuestos de fondo y otras observaciones.
- Por ejemplo, podemos conocer la distribución de probabilidad sobre un conjunto de variables $\mathbf{V}$, y queremos saber cuales estructuras causales sobre las variables en $\mathbf{V}$ son compatibles con la distribución.

Modelos de ecuaciones estructurales con errores aleatorios I

- Se puede introducir probabilidad en un SEM mediante una distribución de probabilidad sobre las variables exógenas.
- Sea $\mathbf{V} = \{X_1, X_2, \ldots, X_n\}$ un conjunto de variables endógenas, y $\mathbf{U} = \{U_1, U_2, \ldots, U_n\}$ el conjunto correspondiente de variables exógenas. Supongamos que cada variable endógena X_i es una función de sus padres en $\mathbf{V}$ junto con U_i, esto es

$$X_i = f_i(\mathbf{PA}(X_i), U_i).$$

- La representación gráfica del SEM incluirá sólo las variables endógenas $\mathbf{V}$, y usaremos $\mathbf{PA}(X_i)$ para denotar el conjunto de padres endógenos de X_i.
- U_i es a veces llamado una variable *error* de X_i: es responsable de cualquier diferencia entre el valor actual de X_i y el valor predicho sobre la base de sólo $\mathbf{PA}(X_i)$.

Modelos de ecuaciones estructurales con errores aleatorios II

- U_i encapsula todas las causas de X_i que no están incluidas en $\mathbf{V}$.
- U_i puede ser un vector de variables. Además, las variables de error y no necesitan ser distintas o independientes entre ellas.
- Una asignación de valores a las variables exógenas $U_1, U_2, \ldots, U_n$ determina de forma única los valores de todas las variables en el modelo.
- Entonces, si tenemos un distribución de probabilidad P' sobre las variables en $\mathbf{U}$, esto inducirá una distribución de probabilidad única P en $\mathbf{V}$.

Condición de Markov I

Supongamos que tenemos un SEM con variables endógenas $\mathbf{V}$, variables exógenas $\mathbf{U}$, una distribución de probabilidad P sobre $\mathbf{U}$ y $\mathbf{V}$ y un DAG $\mathbf{G}$ que representa la estructura causal en $\mathbf{V}$. Si las variables de error U_i son probabilísticamente independientes en P, entonces la distribución de probabilidad en $\mathbf{V}$ satisface la *condición de Markov* (MC) con respecto a $\mathbf{G}$. La condición de Markov tiene varias formulaciones, las cuales son equivalentes cuando $\mathbf{G}$ es un DAG:

- MC Apagada: Para toda variable X en $\mathbf{V}$, y todo conjunto de variables $\mathbf{Y} \subseteq \mathbf{V} \setminus \mathbf{DE}(X)$, $P(X|\mathbf{PA}(X), \mathbf{Y}) = P(X|\mathbf{PA}(X))$.
- MC Factorización. Sea $\mathbf{V} = \{X_1, X_2, \ldots, X_n\}$. Entonces $P(X_1, X_2, \ldots, X_n) = \prod_i P(X_i|\mathbf{PA}(\mathbf{X_i}))$.
- MC d-separación. Sean $X, Y \in \mathbf{V}$, $\mathbf{Z} \subseteq \mathbf{V} \setminus \{X, Y\}$. Entonces $P(X, Y|\mathbf{Z}) = P(X|\mathbf{Z}) \times P(Y|\mathbf{Z})$ si $\mathbf{Z}$ d-separa X y Y en $\mathbf{G}$.

Condición de Markov II

- MC provee condiciones suficientes para que las variables sean probabilísticamente independientes, pero no es una condición necesaria.
- Sean V_i y V_j dos variables distintas en $\mathbf{V}$, con sus variables de error exógenas correspondientes U_i y U_j, representando las causas de V_i y V_j que son excluidas de $\mathbf{V}$. Supongamos V_i y V_j comparten al menos una causa común que excluida de $\mathbf{V}$. En este caso, no se espera que U_i y U_j sean independientes, y el teorema de Pearl no aplicaría. En este caso, la relación causal entre las variables en $\mathbf{V}$ no estará representada apropiadamente por una DAG, pero requiere un ADMG con una arista con doble flecha conectando V_i y V_j.
- Un modelo causal que está comprendido por un DAG y una distribución de probabilidad que satisface MC se llama *red bayesiana causal*.

Condiciones de minimalidad y fidelidad

- La MC establece una condición suficiente pero no necesaria para la independencia condicional. Como tal, la MC por sí misma nunca puede implicar que dos variables sean dependientes condicional o incondicionalmente.
- Las condiciones de minimalidad y fidelidad son dos condiciones que dan las condiciones necesarios para independencia probabilística.

Condición de minimalidad

- La condición de Minimalidad dice, intuitivamente, que cada arista en el grafo evita alguna relación de independencia condicional que se pueda obtener de otro modo.
- Si **G** es un grafo acíclico dirigido sobre **V** y P una distribución de probabilidad sobre **V**, ningún subgrafo de **G** satisface la condición de Markov con respecto a P.
- Si una distribución P satisface las condiciones de Markov y de minimalidad para un DAG **G**, se dice que **G** *representa* P. Para cualquier **G** y cualquier distribución P que satisfacen las dos condiciones, si las variables A y B son dependientes, entonces ya sea que:
  - existe un camino dirigido en **G** de A a B, o
  - existe un camino dirigido en **G** de B a A, o
  - existe una variable C y caminos dirigido en **G** de C a B y de C a A.

Condición de fidelidad

Si todas y sólo las relaciones de independencia condiciones verdaderas en P están implicadas por la MC aplicada a $\mathbf{G}$, se dice que P y G son fieles una de la otra. Además, se dice que una distribución P es fiel siempre que exista algún gráfico acíclico dirigido al que sea fiel.

Identificabilidad de la estructura causal I

- Si se tiene un conjunto de variables $\mathbf{V}$ y conocemos la distribución de probabilidad P sobre $\mathbf{V}$, ¿qué podemos inferir de la estructura causal sobre $\mathbf{V}$? Esta pregunta está relacionada con la cuestión de si es posible *reducir* causalidad a probabilidad.
- Pearl prueba el siguiente teorema (**Identificabilidad con orden de tiempo**): Si
  - las variables en $\mathbf{V}$ están indexadas por tiempo, tal que sólo las variables anteriores pueden causar las posteriores,
  - la probabilidad P asigna probabilidad positiva a cada posible asignación de valores de las variables en $\mathbf{V}$,
  - no existen variables latentes, así que el grafo causal correcto $\mathbf{G}$ es un DAG,
  - y P satisface las codiciones de minimalidad y de Markov con respecto a $\mathbf{G}$,

  entonces es posible identificar de manera única $\mathbf{G}$ con base en P.

Identificabilidad de la estructura causal II

- Si no se tiene información acerca del orden del tiempo, o algunas otras suposiciones que restringen las posibles estructuras causales entre las variables en $\mathbf{V}$, entonces no siempre será posible identificar la estructura causal sólo de la probabilidad.
- Dada una distribución de probabilidad P sobre $\mathbf{V}$, únicamente es posible identificar una *clase de equivalencia de Markov* de estructuras causales.
- Ésta es el conjunto de todos los DAG en $\mathbf{V}$ que implican todas y sólo las relaciones de independencia condicional contenidas en P. En otras palabras, es el conjunto de todos los DAG $\mathbf{G}$ tal que P satisface las condiciones de fidelidad y de Markov con respecto a $\mathbf{G}$.
- El algoritmo PC es un algoritmo que genera la clase de equivalencia de Markov para cualquier distribución de probabilidad con una clase de equivalencia de Markov no vacía.

Identificabilidad con suposiciones acerca de la forma funcional I

- Supongamos que se tiene un SEM con variables endógenas **V**, y variables exógenas **U**, donde cada variable en **V** está determinada por la ecuación de la forma:

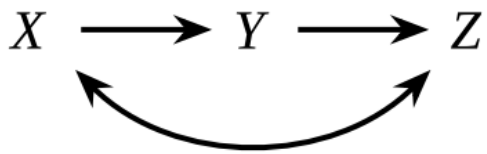
$$X_i = f_i(\mathbf{PA}(X_i), U_i).$$

  Supongamos además que se tiene una distribución P' sobre **U** en la cual todas las U_i son independientes. Esto induce una P sobre **V** que satistace a MC relativamente al DAG causal correcto sobre **V**. El SEM probabilístico genera una red bayesiana causal.
- Métodos mencionados anteriormente intentar inferir **G** a partir de las relaciones de dependencia e independencia probabilísticas.

Identificabilidad con suposiciones acerca de la forma funcional II

- ¿Podemos hacerlo mejor haciendo uso de información adicional sobre la distribución de probabilidad P, más allá de las relaciones de dependencia e independencia?
- Hay supuestos bastante generales que nos permiten inferir mucho más. Por ejemplo **LiNGaM**, **aditivo no lineal**, **post no lineal**, etc.
- Incluso cuando son suposiciones específicas, éstas implican, por ejemplo, que conocer la distribución de probabilidad de dos variables X y Y, se puede inferir si X causa Y o Y causa X.

Causas comunes latentes



- Un ADMG representa que las variables de error para X y Z no son probabilísticamente independientes.
- Un modelo causal que incorpora un ADMG y una distribución de probabilidad que satisface MC d-separación se llama un modelo causal semi-markoviano (SMCM).
- La clase de equivalencia de Markov es más grande que si no se permiten variables lantentes.

Intervenciones

- $P(Y = y|X = x)$ nos da la probabilidad de que $Y = y$, dado que se ha *observado* que $X = x$.
- Estamos interesados en predecir el valor de Y que resultará si intervenimos al fijar el valor de $X = x$. Esto se escribe como $P(Y = y|do(X = x))$.
- Cuando intervenimos, se anula la estructura causal normal, obligando a una variable a tomar un valor que no podría haber tomado el sistema por sí mismo.
- Gráficamente, se puede representar el efecto de la intervención eliminando aristas dirigidas hacia la variable que se interviene. Tal intervención a veces se describe como "romper" esas aristas.

Intervenciones vs. Contrafactuales I

- Existe una conexión entre las intervenciones y los contrafactuales.
- Las intervenciones son representadas por el operador *do*.
- Se entiende que las primeras están en modo indicativo, se refieren a intervenciones que realmente se realizaron. Los contrafactuales están en modo subjuntivo, y se refieren a intervenciones hipotéticas.
- En las intervenciones, nos interesa evaluar probabilidad como

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

- Asumimos que la intervención $do(\mathbf{Z} = \mathbf{z})$ se realiza en el mundo actual, y por lo tanto se están observando los valores que toman otras variables $\mathbf{X} = \mathbf{x}$ en el mismo mundo donde la intervención tomo lugar.

Intervenciones vs. Contrafactuales II

- En el caso de los contrafactuales, observamos el valor de varias variables en el mundo actual, en el que no hay intervención. Después preguntamos que hubiera pasado si se hubiera realizado una intervención. Las variables cuyos valores fueron observados pueden tomar diferentes valores en el mundo hipotético donde la intervención se lleva a cabo.

Intervenciones vs. Contrafactuales III

Intervención	Contrafactual
Una intervención se realiza sobre el tratamiento de un paciente con la droga, y observamos que no se recupera.	Se observa que un paciente no se recupera de la enfermedad.
¿Cuál es la probabilidad de que se recupere dada la intervención y la evidencia observada?	¿Cual es la probabilidad de que se hubiera recuperado si la hubieran tratado con la droga?
	No trivial. La respuesta no es es necesariamente cero, ni es necesariamente $P(recupera \mid tratamiento)$. Si sabemos que de hecho fue tratada, entonces podemos inferir que ella no se hubiera recuperado si se trató. Pero no sabemos si fue tratada. El hecho que no se recuperó, nos da información parcial: esto hace menos probable que en efecto haya sido tratada; y además hace más probable que tenga un sistema inmune débil, y así. Debemos hace uso de toda esa información para tratar de determinar la probabilidad de que ella se hubiera recuperado si se hubiera tratado.
Trivialmente, cero.	

Intervenciones I

- Supongamos un SEM con variables exógenas **U** y variables endógenas **V**. Tenemos ecuaciones de la forma

$$X_i = f_i(\mathbf{PA}(X_i), U_i)$$

y una distribución P' sobre **U**. P' induce una distribución de probabilidad P sobre **V**. Para representar una intervención que define X_k a x_k, se reemplaza la ecuación para X_k con $X_k = x_k$. Ahora, P' induce una nueva distribución de probabilidad $P*$ sobre **V** (porque el ajuste de **U** da lugar a diferentes valores de las variables en **V** después de la intervención). $P*$ es la nueva distribución de probabilidad $P(\bullet, do(X_k = x_k))$.

- Pearl desarrolló un sistema axiomática llamado el cálculo do para calcular probabilidad post-interveción que puede ser aplicado a sistemas con variables latentes, donde la estructura causal está representada por un ADMG.

Intervenciones II

El criterio de la puerta trasera. Sean X y Y variables en $\mathbf{V}$, y $\mathbf{Z} \subseteq \mathbf{V} \backslash \{X, Y\}$ tal que:

- ningún miembro de $\mathbf{Z}$ es descendiente de X, y
- cualquier camino entre X y Y que termina con una arista en X ya sea que: a) incluye un no colisionador en $\mathbf{Z}$,o b) incluye un colisionador que no tiene descendientes en $\mathbf{Z}$;

entonces $P(Y|do(X), \mathbf{Z}) = P(Y|X, \mathbf{Z})$. Esto es, si podemos encontrar un conjunto condicional apropiado, la probabilidad resultante de una intervención en X será la mimas que la probabilidad condicional.

Descubrimiento causal con intervenciones

- Se puede aprender más de la estructura causal si se pueden realizar intervenciones más que sólo haciendo observaciones pasivas. Sin embargo, cuánto podemos inferir depende de qué tipo de intervenciones podemos realizar y de qué suposiciones de fondo hacemos.
- Si no existen causas comunes latentes, tal que la estructura causal verdadera de $\mathbf{V}$ está representada por un DAG $\mathbf{G}$, entonces siempre será posible descubrir las estructura causal completa usando intervenciones.
- Si existen LCC, tal que la estructura causal está representada por un ADMG, entonces puede que no sea posible descubrir la estructura causal verdader con intervenciones de una sola variable. Si intervenimos múltiples variables al mismo tiempo, entonces es posible descubrir el grafo causal completo.
- Una intervención suave influye el valor de una variable sin romper las aristas hacia esa variable.

Contrafactuales I

- Si tenemos una SEM completo, podemos asignar probabilidades a los contrafactuales. Sea $\mathbf{V} = \{X_1, X_2, \ldots, X_n\}$ un conjunto de variables endógenas, y $\mathbf{U} = \{U_1, \ldots, U_n\}$ un conjunto de variables exógenas. Tenemos

$$X_i = f_i(\mathbf{PA}(X_i)U_i).$$

  Tenemos una P' sobre $\mathbf{U}$, que induce a P sobre $\mathbf{U} \cup \mathbf{V}$. Supongamos que se observa el valor de las variables: $X_j = x_j$ para todo $j \in \mathbf{S} \subseteq \{1, \ldots, n\}$. Ahora queremos evaluar el contrafactual "si X_k hubiera sido x_k, entonces X_l hubiera sido x_l", donde k y l pueden estar en $\mathbf{S}$ pero no necesariamente. Podemos evaluar la probabilidad del contrafactual usando el siguiente procedimiento de tres pasos:

  - Actualizar la probabilidad P condicionando sobre las observaciones, para obtener una nueva distribución $P(\bullet | \cap_{j \in \mathbf{S}} X_j = x_j)$. Llamemos a la restricción de esta función de probabilidad a $\mathbf{U}$, P''.
  - Reemplaza la ecuación para X_k con $X_k = x_k$.

Contrafactuales II

- Usar la distribución P'' sobre $\mathbf{U}$ junto con el conjunto de ecuaciones modificadas para inducir una nueva distribución de probabilidad $P*$ sobre $\mathbf{V}$. $P*(X_t = x_t)$ es la probabilidad del contrafactual.

Terminología estadística vs. causal