

INFERENCIA CAUSAL E INFERENCIA ESTADÍSTICA

El primer puente construido entre la inferencia causal y la inferencia estadística, según Pearl (Pearl & Mackenzie, 2018, # 75) es el “path diagram” (diagrama de trayectorias) un mecanismo que permite subir y bajar entre la primera y la segunda gradas de la escalera de la causalidad mencionada anteriormente. Este es un mecanismo que deja en evidencia por primera vez que el análisis causal no trata solo con los datos, sino que debe incorporar alguna comprensión sobre el proceso de producción de los datos y de esta forma se obtiene algo que “no está” en los datos, algo que está en la comprensión de las relaciones de las variables de las que se obtienen los datos. La elaboración de un diagrama de trayectorias (path diagram) expone las correlaciones entre variables de un estudio medidas a través de datos, al tiempo que saca a la luz las cifras resultantes de las relaciones causales o asociativas entre variables mediante ecuaciones algebraicas, con lo queda en claro que “algunas correlaciones si implican causalidad”. En la actualidad (casi 100 años después del trabajo original de Sewall Wright) los coeficientes del path analysis (path coefficients/coeficientes de trayectoria) se interpretan como el resultado de una intervención hipotética sobre la primera correlación establecida con datos reales, que es la fuente del diagrama de trayectorias (path diagram). Estos diagramas en la actualidad son utilizados por pocas estudiosas interesadas en el tema y han evolucionado a otras formas de representación gráfica de las relaciones causales.

Otro hito en la evolución de la probabilidad y la estadística en la identificación de las relaciones causales en la ciencia fue el redescubrimiento relativamente reciente del teorema de Bayes (probabilidad condicional) por las ciencias de la computación, que ha permitido el desarrollo de aparatos “inteligentes”. El teorema de Bayes ofrece la oportunidad de estimar un valor al combinar las probabilidades de dos eventos relacionados entre sí, la probabilidad condicional y su reverso la independencia condicional. Tal vez el papel más importante del teorema de Bayes en relación con los cálculos estadísticos es que:

“Podemos estimar la probabilidad condicional directamente en una dirección, para la cual nuestro juicio es más confiable, y usar las matemáticas para derivar la probabilidad condicional en la otra dirección, para la cual nuestro juicio es un tanto brumoso” (Pearl & Mackenzie, 2018, # 90).

Pearl pone un ejemplo, parecido al que examinó Talbot más arriba, en el que si se observa a un chico lanzar una pelota hacia una ventana, la mayoría podrá predecir el efecto (probablemente la pelota romperá la ventana), mientras que se necesita mucha más información para deducir la causa cuando se observa el efecto, una ventana rota por ejemplo (¿cuál de los chicos lanzó la pelota que rompió la ventana? o incluso, antes que nada, si la ventana fue rota por una pelota). En la práctica médica un ejemplo de esta doble direccionalidad se puede graficar así: una dirección sería la probabilidad de que el evento enfermedad esté presente dado que se tiene un resultado positivo de una prueba de tamizaje en la población general o de una prueba de confirmación/rechazo diagnóstico en personas con conocidos factores de riesgo y/o un cuadro clínico específico. Mientras que la otra dirección sería la probabilidad de que se obtenga un resultado positivo en la prueba diagnóstica en el caso de que se realice como un mecanismo de tamizaje poblacional o en el caso de que se tenga una sospecha diagnóstica.

El teorema de Bayes tiene la propiedad de actuar como una regla matemática que consigue actualizar las apreciaciones de probabilidades ya establecidas, usando la nueva información de las probabilidades observadas en un nuevo estudio o recolectando nuevos datos. Pearl (Pearl & Mackenzie, 2018, # 93) utiliza un ejemplo sobre la actualización de la probabilidad inicial (prevalencia) de que una mujer presente un cáncer de seno una vez que reciba un resultado positivo para un examen de mamografía de tamizaje. El procedimiento aritmético es fascinante

y todo estudiante de ciencias de la salud debería revisar este ejemplo. A manera de resumen aquí se relata el postulado matemático derivado del teorema de Bayes que resuelve este problema. Una probabilidad actualizada de que la paciente tenga cáncer de seno, una vez que recibe el resultado positivo de una mamografía de tamizaje resulta de la multiplicación de la razón de verosimilitud por la probabilidad de que una mujer tenga cáncer de seno en la población (prevalencia). Es decir, se calcula primero la razón de verosimilitud, que representa la relación entre la probabilidad de que una mujer tenga un resultado positivo teniendo la enfermedad versus la probabilidad de que tenga un resultado positivo en la población general. Una vez realizada la mamografía, con la evidencia de un resultado positivo para una mamografía de tamizaje en sus manos, la probabilidad de cada mujer en particular de tener la enfermedad aumentará en una tasa calculable, esta tasa será diferente de la prevalencia de la enfermedad en la población que sería la probabilidad de cualquier persona de presentar esa enfermedad. La historia será diferente si la paciente conoce de algún factor de riesgo de padecer cáncer de seno presente en su caso, la probabilidad de que el resultado positivo corresponda a un verdadero positivo se incrementa significativamente, el teorema de Bayes tiene la propiedad de permitir incorporar esta nueva información en los cálculos y de esta forma “actualizar” el conocimiento sobre la probabilidad de ocurrencia de un evento dado que (Teorema de Bayes) se tiene evidencia de otro evento relacionado.

Creo que la mayoría de las mujeres de cuarenta años que tienen una mamografía positiva se sorprenderían al saber que todavía tienen menos del 1 por ciento de posibilidades de tener cáncer de mama. La figura 3.3 podría hacer que la razón sea más fácil de entender. La pequeña cantidad de verdaderos positivos (es decir, mujeres con cáncer de mama que dieron positivo) se ve abrumada por la cantidad de falsos positivos. Nuestra sensación de sorpresa ante este resultado proviene de la confusión cognitiva común entre la probabilidad directa y la probabilidad inversa. La probabilidad directa, que está bien estudiada y documentada a fondo, es la probabilidad de una prueba positiva si tiene cáncer: 73 por ciento. Pero la probabilidad que se necesita para la toma de decisiones individuales es la probabilidad inversa: la probabilidad de que tenga cáncer si la prueba es positiva. En nuestro ejemplo, esta probabilidad es inferior al 1 por ciento. (Pearl & Mackenzie, 2018, # Fe de erratas)

En el primer caso, cuando la situación en análisis es que se tiene un resultado positivo de una prueba de tamizaje y la probabilidad que se calcula es la de que la persona presente la enfermedad, esta situación es dependiente del contexto, si existen factores de riesgo o de protección específicos de cada paciente, estos pueden ser incorporados en los cálculos. Esta probabilidad condicional (o inversa) es sensible a las condiciones del contexto.

Mientras que en el segundo caso, cuando la situación en análisis es que se sospecha la enfermedad dado un cuadro clínico o epidemiológico y la probabilidad que se calcula es la de obtener un resultado positivo en una prueba diagnóstica, Pearl señala que esta situación es “robusta” frente al contexto, se modifica escasamente por factores de riesgo o protección.

“Esto explica porqué las profesionales de salud organizan y comunican su conocimiento con probabilidades previamente establecidas. Estas son propiedades de la enfermedad en sí mismo, el estadio de progresión o el valor predictivo de las pruebas diagnósticas, dado que estas permanecen relativamente estables ante las causas de la enfermedad (epidemiología, dieta, sanidad, estado socioeconómico, historia familiar)” (Pearl & Mackenzie, 2018, # 108).

Las profesionales de la salud suelen explicar a sus pacientes que en personas de las mismas características que la paciente y que provienen de poblaciones y estratos sociales similares la posibilidad de que un cierto tipo de enfermedad esté presente es alta o baja. De la misma manera, cuando una paciente acude con un resultado de exámenes auxiliares de diagnóstico,

los profesionales de la salud los analizan incorporando información sobre los antecedentes patológicos (familiares y personales), los factores de riesgo identificables (y conocidos), la sintomatología y los hallazgos del examen físico. Este procedimiento adiciona información sobre la experiencia del profesional enfrentando pacientes con similares características y contextualiza todo este razonamiento en función de la mejor evidencia científica aplicable al problema concreto de salud de la paciente. Las profesionales hacen razonamientos causales, a partir de probabilidades condicionales en ambos sentidos (hacia adelante y hacia atrás), puesto que están dotadas de este mecanismo de razonamiento y además los conocimientos adquiridos les permiten organizar la información en secuencias mentales parecidas a los DAGs mediante las cuales se efectúan ejercicios de “selección odiosa” de la variable causal, basados en parámetros estadísticos y en síntesis de conocimientos.

Las propiedades del teorema de Bayes hasta aquí permiten la actualización de las probabilidades conocidas en las dos direcciones, sin diferenciar si las relaciones son causales o no, por lo que Pearl se declara solo “medio Bayesiano”. Esto pone en evidencia la relación de dos eventos de probabilidad condicional, si se añade una o varias probabilidades más a estos dos eventos relacionados se tiene una “confluencia” (Pearl & Mackenzie, 2018, # 113).

Pearl describe tres tipos básicos de “confluencia”, recalando en el fenómeno que describe como que “unas variables escuchan a otras” para modificar sus valores en cada unidad de observación:

- Mediación (causa directa o indirecta): $A \rightarrow B \rightarrow C$. El ejemplo más básico es una “cadena”, que es frecuente postular en un trabajo científico. El mediador B actúa como un intermediario entre los datos que proporcionan A y C, una vez que se establece una regularidad de la presencia de B antes de que aparezca C, entonces la presencia de A no es suficiente para explicar el efecto C, es necesario que se observe que C no puede aparecer solo con la presencia de A, pero que A debe estar presente para que aparezcan tanto B como C. Esta estabilidad en la relación causal entre estas tres variables se puede inferir de los datos y este tipo de inferencia sucede de forma natural en el cerebro humano. Existe la posibilidad de que una parte de la acción de A se efectúe directamente sin la intervención de B y que la otra parte sea ejecutada a través de B.
- “Tenedor” (causa común): $A \leftarrow B \rightarrow C$. En este caso B actúa como causa compartida o confusor de A y C. El confusor hará que A y C eventualmente se correlacionen estadísticamente incluso si no existe una relación causal directa entre ellas. Esta es la fuente de las llamadas “correlaciones espurias” de las que se habla en estadística. La tradición epidemiológica ha definido la confusión en la relación entre variables según los siguientes criterios: que si se separa al grupo de estudio en las categorías de la confusora B, la correlación estadística entre A y C debería debilitarse considerablemente y no habrá correlación estadística entre A y C en cada uno de los subgrupos de la confusora B pero sí estará presente en el grupo total de individuos del estudio. Si la asociación estadística entre A y C no aparece en ausencia de B, entonces se habla de “independencia condicional”.
- “Colisionador” (efecto común): $A \rightarrow C \leftarrow B$. En este caso se tiene un efecto compartido. Si solo se observan los individuos que presentan la variable C (es decir, el efecto que se encuentra en estudio), entonces se encontrará una correlación de A y B en estos individuos, que está presente solo en este tipo de individuos, pero no en la población general. Este patrón de asociación funciona en forma contraria a las cadenas y los tenedores. La asociación estadística entre A y B aparecerá sólo cuando se realicen pruebas estadísticas estratificadas por la variable de resultado, mientras que no estarán presentes en la población general. De esta forma se producirá el llamado

“sesgo de colisión” o el “efecto soslayar” (explain-away effect), que consiste en que si las dos variables explicativas de la ocurrencia de C son demostrables estadísticamente, entonces será lícito suponer que entre ellas existe algún tipo de relación causal, a pesar de que en la realidad A y B no estén ni teórica ni estadísticamente relacionadas e incluso no se encuentren en la cadena causal. Corresponde a otra forma de “independencia condicional”.

Pearl advierte que en esta taxonomía se trata de identificar estas “confluencias” causales aisladas con fines de exposición, pero que en la práctica estas confluencias ocurren en un contexto en el que múltiples trayectorias causales se organizan en un modelo MULTICAUSAL de explicación de un efecto, las mismas que tienen que estar encadenadas en diferentes posibles combinaciones; con esto en mente, para un análisis de causalidad deben ser tomadas en consideración todas las trayectorias que redondean las principales relaciones en estudio. Estas relaciones se establecen en forma secuencial, dando la imagen de que las relaciones iniciales engendran a las subsiguientes a manera de “nodos padres” y “nodos hijos”, con flechas que señalan la secuencia de esta relación.

Una red Bayesiana es literalmente nada más que una representación compacta de una enorme tabla de probabilidades. (Pearl & Mackenzie, 2018, # 128)

La construcción de las cadenas de causalidad, depende de varios criterios como: el tema en cuestión, las relaciones conocidas entre las variables, la selección “odiosa” de “la causa” (variable de exposición a ser estudiada), la decisión de incluir variables de las que se desconoce su probabilidad en la población en estudio, la metodología de recolección de datos (i.e. si se incluye o no una intervención, se comparan grupos con o sin asignación aleatoria), la ausencia de relación entre dos variables (evidente por la ausencia de una flecha desde la exposición a la consecuencia), el cumplimiento de los requisitos para el criterio de la “puerta de atrás” y de la puerta frontal, la separación-d, la utilización del “do-operator”, etc.

Cabe imaginar que los tres tipos de confluencias se van a producir múltiples combinaciones con solo mirarlas nuevamente (las variables se escuchan unas a otras).

$A \rightarrow B \rightarrow C$, $A \leftarrow B \rightarrow C$, $A \rightarrow C \leftarrow B$
 $A \leftarrow B \rightarrow C$, $A \rightarrow C \leftarrow B$, $A \rightarrow B \rightarrow C$
 $A \rightarrow C \leftarrow B$, $A \rightarrow B \rightarrow C$, $A \leftarrow B \rightarrow C$, etc

Estas combinaciones de confluencias que conectan las reglas matemáticas de la inferencia derivadas del teorema de Bayes con una colección de hechos generales y específicos se organizan en las denominadas “Redes Bayesianas”. Según Pearl (Pearl & Mackenzie, 2018, # 111) le tomó un poco de tiempo darse cuenta de que las relaciones matemáticas o los “mensajes” entre los nodos “son probabilidades condicionales en una dirección y razones de verosimilitud en la otra”..... “Aplicar estas dos reglas repetidamente para cada nodo en una red se denomina propagación de convicciones (propagación probabilística)”.

Pero las redes bayesianas no necesariamente resuelven preguntas causales, las relaciones entre variables en una red Bayesiana pueden ser de causalidad o de asociación. Para resolver preguntas causales mediante la inferencia causal es indispensable utilizar un lenguaje gráfico con reglas bien establecidas, los diagramas causales, para los que siguen siendo válidas las propiedades de las redes bayesianas; las propiedades probabilísticas y los algoritmos de propagación de las “convicciones” que se describieron más arriba (Pearl & Mackenzie, 2018, # 114). Con los tres tipos de confluencias antes descritas, se pueden construir secuencias de diferentes grados de complejidad y en múltiples combinaciones, encadenadas en trayectorias más o menos largas según el avance del conocimiento científico. Pearl menciona que

probablemente no sea indispensable elaborar el modelo gráfico para cada ejercicio de investigación, pero que es absolutamente necesario hacerlo para identificar relaciones causales axiomáticas.

En una confluencia tipo tenedor el grave peligro que enfrenta la causalidad es que la relación que existe entre A y C, que están asociadas sólo por el hecho de que tienen a B como causa común, resulte en una correlación estadística significativa que lleve a la conclusión de que su relación es causal. El conocido aforismo “asociación no significa necesariamente causalidad” es ampliamente aceptado, este aforismo implica desde luego que toda relación de causalidad requiere necesariamente algún tipo de asociación. Ahora bien, si al contrario de lo que tenemos en una confusión de tipo tenedor, en realidad existe una relación causal entre A y B y adicionalmente se presenta una tercera variable C que tiene relación de asociación tanto con A como con B, se produce una situación compleja de causalidad que ha sido denominada de confusión (en inglés *confusion*, que puede entenderse como entrometimiento de una tercera variable en la relación verdadera de causalidad entre Causa y Efecto). La variable (o paquete de variables) de confusión debe cumplir ciertos criterios para ser considerada como tal y esta discusión continúa latente en la actualidad, por lo que será abordada *inextenso* más adelante.

Según la definición clásica de la epidemiología, la variable de confusión (B) se encuentra fuera de la trayectoria causal desde la variable causal seleccionada para el estudio (A) y la variable de resultado (C). La situación más común es que la variable de confusión actúe sobre la variable de resultado en una cierta proporción en forma directa y en otra proporción en forma indirecta, a través de la variable causal con la que también tiene una asociación. Esto abre una trayectoria “desde atrás” de la variable de confusión que utiliza la variable causal para modificar su influjo sobre la variable de resultado, creando una situación anómala en la que el resultado va a ser diferente en cada categoría de la variable de confusión y que podría ser paradójal en comparación con el resultado en el grupo total sin estratificar por las categorías de la variable de confusión. La estadística del siglo pasado se ha dedicado en gran medida a resolver este escenario indeseable y ha desarrollado un arsenal de pruebas estadísticas amplio para realizar el denominado control de confusión. Sin embargo, ha surgido una corriente importante de pensamiento que cuestiona que el control estadístico de la confusión sea suficiente y adecuado para resolver el problema de la confusión, si no se cambia la perspectiva total de la noción de causalidad (y por consiguiente de confusión) mediante el desarrollo de una notación matemática correspondiente y de un lenguaje gráfico que permita reconocer los diferentes tipos de confluencias que se presentan en el proceso de desentrañar el verdadero papel de una variable considerada “la causa” (selección odiosa) en el marco de las demás variables y cadenas de causalidad relacionadas entre sí y con el resultado.

En cambio, si la confluencia es de tipo mediación, queda en claro que la variable de mediación se encuentra en la trayectoria de causalidad y que es necesaria para poder establecerla. La discusión sobre la transitividad de la causalidad que se expuso en el capítulo de talante filosófico es adecuada para la comprensión del papel de este tipo de confluencia en la definición heurística de causalidad. La transitividad significa que una parte de la “carga” causal se transmitirá directamente desde la causa antecedente a la variable de mediación y en consecuencia desde esta a la variable de resultado, pero además que una parte de la causalidad podría ser ejercida directamente desde la causa antecedente hacia el resultado, sin tener que pasar por la acción de la variable de mediación y otra parte será transmitida “indirectamente”, a través de la variable de mediación. Esto plantea una condición relativamente reciente en la discusión de la causalidad, en la que no solo no se requiere un control de la confusión, sino también un adecuado manejo de la mediación tanto matemático como en su representación gráfica. El realizar un procedimiento de ajuste de la probabilidad de que suceda el efecto DADO que ha sucedido la causa, en presencia de una tercera variable que

se encuentra en la trayectoria causal entre A y C, produce el efecto de realizar el ajuste “por lo que se quería medir”, anulando la relación causal originalmente planteada.

Es una idea ampliamente aceptada que los estudios observacionales tienen una capacidad limitada para dar certera cuenta de las relaciones causales, incluso si están bien diseñados y si se ha logrado controlar la mayor parte de las variables de confusión. Esta noción actualmente se encuentra fuertemente cuestionada y existen posiciones científicamente fundadas que demuestran que esta noción no es exacta en todos los casos, sino en ciertas circunstancias específicas. Por el contrario, en este momento se dispone de suficientes argumentos tanto para obtener evidencia causal de estudios observacionales, como para combinar estudios observacionales con estudios experimentales y mejorar las capacidades de identificación causal de los estudios observacionales, así como redimensionar la dependencia científica de la determinación causal en base casi exclusivamente a estudios experimentales.