

BAYESIAN DATA ANALYSIS AS A TOOL FOR BEHAVIOR ANALYSTS

MICHAEL E. YOUNG

2019

Los análisis bayesianos se están volviendo plausibles dado el incremento en poder computacional. Además, dan respuesta a problemas sobre el uso histórico de análisis estadísticos: las pruebas de hipótesis se enfocan en las decisiones y excluyen el juicio del investigador dado que éste es propenso a errores. En contraste, la aproximación bayesiana enfatiza la evidencia (y su grado) más que las decisiones.

Un problema importante de las pruebas de hipótesis nula es que con muestras lo suficientemente grandes, detectarán efectos increíblemente pequeños siempre que estos no sean exactamente de cero. El análisis Bayesiano, por el contrario, aborda el problema del rechazo de la hipótesis nula evaluando simultáneamente la verosimilitud de toda una distribución de valores hipotéticos del efecto, y no solamente de cero.

INFERENCIA BAYESIANA

El teorema de Bayes se trata simplemente de la forma matemáticamente correcta de integrar las creencias previas con la nueva evidencia. Este conocimiento previo se suele representar en forma de una probabilidad. La nueva evidencia de un resultado observado es proveída con dos probabilidades condicionales. Por ejemplo, en el caso de la diabetes, el prior sería la incidencia de diabetes en la población. La evidencia, es decir, una prueba positiva, sería integrada junto con la probabilidad de que una persona resulte positiva dado que sea de hecho diabética (es decir, un positivo verdadero, $P(\text{positivo}|\text{diabetico})$), y la probabilidad de resultar positivo dado que no se sea diabético (es decir, un falso positivo, $P(\text{positivo}|\text{no diabetico})$).

En el análisis de la conducta se pueden hacer análisis bayesianos de los datos obtenidos mediante la experimentación, o bien, se puede proponer que los organismos mismos son “Bayesianos ingenuos”.

El análisis bayesiano implica que un investigador comience con una hipótesis previa (como la pendiente de una función) y la actualice con base en la evidencia recolectada. La actualización se puede hacer de dos maneras: usando Factores de Bayes para comparar la verosimilitud relativa de dos o más hipótesis tras la recolección de datos; o usando un análisis Bayesiano de datos completo para proponer una distribución *prior* del parámetro, la cuál se actualiza con los datos experimentales para formar una nueva distribución posterior.

COMPARACIÓN DE MODELOS Y FACTORES DE BAYES

En pruebas de hipótesis nula las comparaciones de modelos implican estimados individuales de parámetros: si cada estimado es distinto de cero.

Un problema serio con esta aproximación frecuentista es que se está estimando la probabilidad P de los datos *cuando la hipótesis nula es correcta* ($P(\text{datos}|\text{hipotesis nula})$), mientras que el investigador piensa que está estimando la probabilidad P de que *la hipótesis nula sea correcta dados los datos observados* ($P(\text{hipotesis nula}|\text{datos})$). Esta aproximación ignora las probabilidades a priori y, aparentemente, tiende a sobreestimar las probabilidades de que cierta teoría sea correcta.

La verdadera P (“una teoría cualquiera”) es bastante pequeña usualmente debido a la enorme variedad de teorías similares que son posibles y casi indistinguibles entre sí dado un número finito de datos. Por ello, es conveniente hacer comparaciones con varias teorías alternativas (*the method of strong inference*, Platt, 1964).

De forma clásica, se puede usar el valor de R^2 para tomar decisiones con respecto a los modelos ajustando para la complejidad de cada uno. Pero esto requiere que un modelo esté anidado en el otro, es decir, que un modelo contenga todas las variables del otro y alguna más. Cuando esa condición no se cumple, como al comparar un modelo exponencial con uno hiperbólico, es necesario usar el juicio del investigador. El problema con ello es que el software no suele indicar el grado de incertidumbre, sino solamente la R^2 . Ignora intervalos de confianza y números de observaciones, por ejemplo.

Una aproximación alternativa compara la verosimilitud de que los datos hayan sido producidos por un modelo u otro (*likelihood ratio*). Así, una verosimilitud de 5×10^{-6} sería 25 veces superior a una de 2×10^{-7} . Sin embargo, no debe caerse en la trampa de creer que se debe exceder un cierto valor crítico de verosimilitud, pasando de una aproximación basada en evidencia a una categórica. La meta es acumular evidencia, no tomar decisiones discretas. Varios experimentos con *likelihood ratio* pequeños pero en la misma dirección forman un resultado deseable y meritorio de reporte que de otro modo podría ser ignorado.

Existen métricas para comparar los modelos ajustando para su complejidad. Por ejemplo, *Akaike Information Criterion* (AIC) y *Bayesian Information Criterion* (BIC). Estos criterios comparan la evidencia relativa de los modelos en cuestión: el modelo con el menor valor es más verosímil como origen de los datos. Estas métricas solo permiten comparar modelos basados en el mismo conjunto de datos.

El factor de Bayes (BF) es similar al *likelihood ratio* salvo porque incluye información acerca de los priors de los parámetros del modelo, es decir, aquello que se cree sobre los parámetros antes de la recolección de datos. El uso de priors, aunque parece sacrilegio, se justifica al pensar en que únicamente se están haciendo explícitas las presunciones que ya tienen los investigadores como producto de su experiencia, además del hecho de que pueden usarse priors no informativos que resultan en análisis funcionalmente idénticos a los frecuentistas.

El BF se puede aproximar utilizando valores de BIC, asumiendo un prior particular poco informativo llamado *unit information prior*. Sin embargo, esto ha sido criticado por su tendencia a favorecer demasiado a los modelos simples, por lo que algunos investigadores

prefieren usar AIC. Hay otros priors *por default* que se pueden utilizar, o se puede recurrir a la experiencia del investigador. Esta aproximación permite especificar un “rango nulo” en lugar de un único valor de hipótesis nula. Así, se busca evidencia de que el valor encontrado es mas que “irrelevantemente pequeño”, definiendo un rango de valores que son irrelevantes.

Esta aproximación cambia el enfoque del científico hacia la fuerza relativa de la evidencia, y evita el uso arbitrario de umbrales para tomar decisiones.

ANÁLISIS DE DATOS BAYESIANO

En la aproximación Bayesiana todos los parámetros del modelo tienen priors que se actualizan con los datos del experimento. Se evalúa todo un rango de valores posibles y sus verosimilitudes. Así, el análisis requiere de una distribución de priors, que se integra con los resultados del experimento y resulta en una distribución posterior.

El análisis Bayesiano puede ser especialmente útil con tamaños de muestra pequeños en los que test de significancia normales serían insuficientes.

Dado que el análisis Bayesiano se basa en simulaciones en una clase de aproximación de fuerza bruta, tiene gran flexibilidad y puede ajustar una gran variedad de modelos. Sin embargo, eso implica que su costo en poder y tiempo de computación es mucho mayor.

De particular interés para el análisis de la conducta es el hecho de que los análisis bayesianos permiten la construcción ordenada en incremental del conocimiento, con priors que se actualizan con el transcurso de los experimentos. En el AEC suelen existir numerosos estudios que permiten descartar ciertos priors y sugerir otros, por lo que no se necesitan utilizar los priors estándar no-informativos. Esto, sumado a los pocos sujetos que se suelen utilizar, resulta en una gran practicidad para los análisis bayesianos.

Muchos científicos se sienten incómodos con el uso de priors. Los bayesianos, sin embargo, deben ser explícitos en su elección de priors, que pueden y deben ser criticados por sus pares.

Una distribución de priors sumamente mala puede tener un efecto negativo sobre la estimación. Sin embargo, este no suele ser el caso si los priors son tomados de investigaciones previas y algo de sentido común (y en caso de no tener lo uno ni lo otro, siempre se puede recurrir a un prior poco informativo). Y aún en el peor de los casos, dado que el científico es explícito en el origen de sus priors, sus pares pueden detectar con facilidad su error.

El uso actual del conocimiento previo es bastante más informal, traducándose en puntos de corte arbitrarios para la inclusión de datos, por ejemplo. Esto es susceptible a sesgos de memoria, promedio no sistemático, muestreo idiosincrático de la literatura, y razonamiento motivado, entre otras cosas.