
Acerca de la defensa de la racionalidad bayesiana y la obra de Mario Bunge

About the defense of bayesian rationality and the work of Mario
Bunge

Andrés Gutiérrez Rojas^a
hugogutierrez@usantotomas.edu.co

La crítica de Silva a la obra de Mario Bunge está sustentada en la ligereza de Bunge al momento de utilizar el teorema de Bayes, con el cual descalifica abiertamente el uso del paradigma bayesiano en las ciencias. En este documento se comenta la definición correcta de los espacios muestrales al usar probabilidades y se recalca que el uso de los métodos bayesianos y de la estadística en general, no deben estar supeditados a un razonamiento puramente filosófico.

1. Sobre la probabilidad condicional y los espacios muestrales

Al leer la introducción que Mario Bunge hace al problema de calcular la probabilidad de que una persona con VIH tenga sida, es posible advertir que el autor hace una escala por un problema de probabilidad aplicada en donde se dice que un evento causa a otro. En estos términos, el autor concluye que ‘si C y E son dos sucesos aleatorios y C causa a E , entonces $P(E|C) \leq P(E)$ ’. Es aquí en donde creo que el susodicho autor debería haber razonado un poco más acerca de cuándo se tiene la igualdad en esta regla de probabilidades, puesto que en el ejemplo que suscita esta discusión, en el que el suceso tener VIH causa al suceso desarrollar sida, Bunge concluye que $P(sida|VIH) = P(sida)$ y sin ninguna precaución concluye que las probabilidades bayesianas son arbitrarias y no tienen cabida en la ciencia auténtica.

Desarrollando el anterior planteamiento, al suponer que C causa a E , nótese que

$$P(E|C) = \frac{P(E \cap C)}{P(C)} = \frac{P(E)}{P(C)} \geq P(E) \quad (1)$$

con igualdad si y solamente si $P(C) = 1$. En el ejemplo que nos atañe sobre el hallazgo de Bunge, lo anterior redundaría en que $P(VIH) = 1$, que significaría

^aDecano, Facultad de Estadística, Universidad Santo Tomás (Colombia). Colombia.

que la probabilidad de que ese cierto individuo, seleccionado al azar, tenga VIH es la unidad y eso solo sería posible si, como Silva muy bien lo explica, el espacio muestral estuviese definido por las personas que tienen VIH, puesto que en este ejemplo el evento VIH causa al evento sida. No quiero desarrollar más este punto, pues el autor del artículo hace muy bien en develar estos detalles, pero sí quiero recurrir al siguiente teorema que examina las implicaciones de la igualdad en la expresión (1).

Teorema 1. *Siendo $(\Omega, \mathfrak{F}, P)$ un espacio de probabilidad en donde los elementos de la sigma-álgebra \mathfrak{F} se denotan como A_1, A_2, \dots y son tales que $P(A_i) \neq 0$ para todo $A_i \neq \emptyset$, entonces*

$$P(A_k) = 1, \text{ para algún } A_k \in \mathfrak{F}, \text{ si y sólo si } A_k = \Omega.$$

Demostración. La implicación se prueba por contradicción al asumir que $P(A_k) = 1$. Al suponer que $A_k \neq \Omega$, entonces naturalmente $A_k^c \neq \emptyset$. Por tanto, acudiendo al enunciado del teorema, se tendría que $P(A_k^c) > 0$, y si tenemos en cuenta que $A_k \cup A_k^c = \Omega$, entonces

$$P(\Omega) = P(A_k^c) + P(A_k) > 1$$

Lo cual implica una contradicción a los axiomas básicos de probabilidad. Por otra parte, la demostración de la converso es trivial. \square

El anterior teorema se puede aplicar al ejemplo de Bunge en donde solo hay dos eventos de interés VIH y sida, cada uno con probabilidad no nula, y además, $P(\text{VIH}) = 1$, por tanto, el espacio muestral al cual Bunge se refiere en su ejemplo es sin duda el condicionado a VIH.

2. Razonamiento antibayesiano

En términos pragmáticos, y teniendo en cuenta los grandes avances que han surgido de la estadística bayesiana, ningún científico contemporáneo debería negar las bondades de este tipo de pensamiento. Desde las ciencias de la administración y el mercadeo, pasando por la teoría actuarial y ensayos clínicos, hasta las ciencias de la computación y la epidemiología, la estadística bayesiana ha jugado y jugará un papel fundamental en el intento humano de extracción de información con base en un conjunto de datos. Sin embargo, como se evidencia en la obra de Bunge, hay todavía pensadores que van en contra del paradigma bayesiano.

Entre los argumentos de Bunge, se encuentran frases como las siguientes

Por ser subjetivas las probabilidades bayesianas son arbitrarias y no deberían ser usadas en la ciencia...

Basado en lo anterior, cabría recordar que, todas las ciencias, y en particular la estadística, están compuestas de procedimientos subjetivos que guían a resultados medibles objetivamente (Bernardo 2008, Kadane 2008). Sin embargo, el uso del teorema de Bayes, que afirma que la distribución posterior de los parámetros condicionada a los datos observados es proporcional al producto de la verosimilitud con la distribución previa de los parámetros de interés, no debe ser usado a la ligera como una fórmula automática que brinda inferencias rápidas. En estos términos, la asignación de las probabilidades previas no puede ser arbitraria. La anterior afirmación parecería un punto a favor de los argumentos de Bunge, pero cualquier estadístico, clásico o bayesiano, con experiencia en la creación de modelos estocásticos sabrá que la realidad de la profesión consiste en la asignación de diferentes modelos de probabilidad a los datos observados y en la respectiva escogencia del mejor modelo. Esta tarea debe ser concienzuda y responsable.

Ya sea que el estadístico opte por un enfoque clásico o bayesiano, se debe notar que la asignación o la escogencia de las probabilidades (o distribuciones de probabilidad) no está sujeta a ningún principio objetivo: desde el punto de vista bayesiano, es menester asignar una distribución de probabilidad previa a los parámetros de interés y esta tarea es exactamente la misma que el estadístico clásico enfrenta al asignar una verosimilitud a sus datos. En la estadística aplicada, la asignación de verosimilitudes y distribuciones previas está determinada por la escogencia de un modelo que represente bien la realidad de los datos observados. En estos términos, el estadístico clásico podrá optar por un modelo o por otro y, de la misma manera, el estadístico bayesiano podrá seleccionar una distribución previa u otra. Esta libertad de escogencia, que ocurre bajo los dos paradigmas, no induce ninguna arbitrariedad. De hecho, como lo mencionan Andrews & Baguley (2013), las distribuciones previas son supuestos en el modelo bayesiano y como todos los supuestos, pueden ser buenos o malos, pueden extenderse, revisarse o posiblemente abandonarse al no ser idóneos con los datos observados.

Con base en lo anteriormente mencionado, el estadístico clásico podrá escoger un nuevo modelo a medida que tenga un conocimiento más profundo de la realidad a la cual intenta acercarse. De hecho, desde mi punto de vista, considero que esa escogencia ha definido la historia de la estadística hasta hoy. Por ejemplo, es posible desviarse de un modelo paramétrico Gaussiano y preferir un modelo no paramétrico, el cual si bien podría ser más complejo en su formulación matemática, tiene mejores propiedades en algunos contextos. De la misma manera, el estadístico bayesiano deberá escoger distribuciones previas plausibles. Esto, en términos prácticos, implica que si la distribución posterior se desvía significativamente de la distribución previa, entonces el estadístico deberá reconsiderar y reevaluar la escogencia de la distribución previa. De esta forma, es posible afirmar que la escogencia de distribuciones previas es tan objetiva como la escogencia de verosimilitudes en la estadística clásica.

En esta discusión es posible que afloren otro tipo de razonamientos filosóficos; por ejemplo que los parámetros de una distribución de probabilidad son fijos y no aleatorios. Salsburg (2001) afirma que, tal como lo mostró Pearson, el propósito de

la investigación estadística es estimar los parámetros (fijos pero desconocidos) de las distribuciones de probabilidad cuyos valores definen dicha distribución. Por lo anterior, si los parámetros se consideraran aleatorios, el enfoque de la investigación científica se desviaría de su propósito. Sin embargo, muchos autores bayesianos han fallado en argumentar que, lo que se considera aleatorio no es el parámetro en sí, sino la incertidumbre que se genera ante el desconocimiento del parámetro y es eso exactamente lo que debería ser modelado, por medio de una distribución previa.

3. Argumentos de ambas partes

A simple vista, es posible evidenciar que Bunge y Silva, y los comentaristas invitados en este número de esta Revista hacemos parte de las animadas, pero interesantes y necesarias, discusiones que durante siglos han caracterizado el desarrollo de la estadística y sus diferentes paradigmas, entre ellos el bayesiano y el clásico.

Los estadísticos bayesianos acuden a una vieja contradicción, basada en el principio de verosimilitud, para impulsar el abandono de los métodos clásicos en pro del paradigma bayesiano y, a su vez, los estadísticos clásicos han utilizado un problema atribuido a Laplace para ridiculizar el punto de vista bayesiano. Detengámonos en estos ejemplos por un instante.

El ejemplo del principio de verosimilitud sugerido por Lindley & Philips (1976) reza de la siguiente forma:

Suponga que se realizan 12 lanzamientos independientes de una moneda y se observan 9 caras y 3 sellos. Además, se desea cotejar el siguiente conjunto de hipótesis

$$H_0 : \theta = 0.5 \quad \text{vs.} \quad H_a : \theta > 0.5$$

En donde θ representa la probabilidad de obtener una cara. Si solamente tuviésemos acceso a esta información, surgirían dos opciones de verosimilitud que podrían ser candidatas para realizar el cotejo de las hipótesis de interés. En primer lugar la distribución binomial, que asume que X es la variable aleatoria que determina el número de caras en los 12 lanzamientos, se tiene que

$$\alpha_1 = P_{H_0}(X \geq 9) = \sum_{x=9}^{12} \binom{12}{x} \theta^x (1 - \theta)^{12-x} = 0.075$$

Y la otra verosimilitud candidata sería la distribución Binomial Negativa, que supone que X representa el número de caras necesarias para lograr 3 sellos, para lo cual se tiene que

$$\alpha_2 = P_{H_0}(X \geq 9) = \sum_{x=9}^{\infty} \binom{2+x}{x} \theta^x (1-\theta)^3 = 0.0325$$

De esta forma, asumiendo un nivel de significación del 5 %, el estadístico rechazará la hipótesis nula si la variable aleatoria X tuviese distribución binomial negativa y no la rechazaría si X tuviese distribución Binomial. Lo anterior constituye una violación al principio de verosimilitud (Birnbbaum 1962) que afirma que, si los datos han sido observados, la función de verosimilitud debe contener toda la información experimental relevante acerca de los parámetros de interés. Nótese que estas dos verosimilitudes son proporcionales entre sí como funciones de θ , pero aunque guían al mismo estimador de máxima verosimilitud, esta equivalencia lleva a dos conclusiones diferentes en términos del sistema de hipótesis de interés.

El segundo ejemplo, atribuido a Laplace y con el cual los estadísticos clásicos se mofan de los bayesianos es el siguiente:

Teniendo en cuenta que el sol ha salido n veces consecutivas en los pasados n días. Cabe cuestionarse acerca de ¿cuál es la probabilidad de que el sol salga mañana, dado que ha salido en los pasados n días? Para resolver esta pregunta, Laplace utilizó la regla de la sucesión (Wikipedia 2013) para demostrar que, siendo p la probabilidad de que el sol salga mañana, la esperanza posterior de p es $\frac{n+1}{n+2}$. Esta conjetura se logra al asumir que la distribución previa de p es uniforme continua, como una opción probabilista de la total incertidumbre sobre este parámetro, y una verosimilitud binomial, describiendo el número de éxitos en n ensayos. Lo anterior, siguiendo la regla de Bayes, redundante en una distribución posterior con kernel Beta de parámetros ($\alpha = n + 1$, $\beta = 1$).

Algunos autores (Popper 1957, Schay 2007, Gorroochurn 2011) han ridiculizado la escogencia de la distribución previa en el anterior problema, utilizando sus argumentos para desvirtuar el uso de la estadística bayesiana. En particular cabría preguntarse ¿por qué se justifica la asignación de una distribución uniforme continua a la probabilidad previa de que el sol salga? Esta escogencia implica que el conocimiento previo de que se tiene del problema le otorga la misma densidad a cualquier valor en el intervalo $(0, 1)$ y este supuesto no es científicamente razonable.

Los anteriores dos ejemplos son una muestra de una gran cantidad de argumentaciones que intentan desvirtuar los enfoques clásico y bayesiano y que, en mi apreciación, alimentan el desarrollo mismo de la estadística.

Sin embargo, en defensa del estadístico clásico se puede afirmar que el diseño del experimento debería ser incluido en el contraste de las hipótesis, porque en términos experimentales, no es lo mismo lanzar una moneda hasta obtener cierto número de caras, que lanzarla un número predefinido de veces y, por consiguiente, la manera como se llevó a cabo el experimento¹ es información relevante (Carlin & Louis 2009). Por otro lado, en defensa del estadístico bayesiano, podría argumentarse que el ejemplo está desenfocado y que el problema de Laplace no consiste en la asignación de la distribución previa sino en la asignación de la verosimilitud como una distribución binomial, que considera eventos independientes, a un problema que, a todas luces, no debería asumir una probabilidad de éxito constante cada día, ni tampoco que los días en los cuales el sol ha salido son eventos independientes unos de otros (Gelman & Robert 2013).

Por último, en mi interpretación de la práctica estadística y del ejercicio de modelación que se realiza en esta profesión, suscribo completamente la frase de Gelman & Cosma (2013) cuando afirman que un método estadístico puede ser útil aún cuando sea filosóficamente errado.

4. Conclusión

A pesar de que en esta discusión se ha demostrado que las conclusiones de Bunge son, por lo menos, desviadas de las reglas de probabilidad y por tanto no científicas, es de suponer que sin importar la falsedad de sus argumentos, el potencial de malinterpretación en una obra escrita difundida de forma masiva es grande. Máxime cuando un gremio, como el de los médicos, no tiene porqué conocer los axiomas fundamentales de la probabilidad. Por tanto, creo yo que este tipo de literatura desvirtúa el quehacer del estadístico, clásico o bayesiano, y crea en el profesional de la medicina una gran barrera hacia el uso de nuevas técnicas estadísticas.

Como bien lo afirman Cox & Donnelly (2011), no todos los análisis estadísticos que utilicen la regla de Bayes deberían ser llamados bayesianos. En este sentido se debería garantizar por lo menos que: 1) la distribución previa represente una evidencia del conocimiento previo del problema; 2) esta evidencia debe ser consistente con los datos observados; 3) realizar un análisis de sensibilidad² sobre las distribuciones previas candidatas cuando hay total incertidumbre del problema (Gustafson et al. 2010).

Con estas indicaciones en mente, es posible concluir que el ejercicio de probabilidades de Bunge, solamente constituye un esfuerzo filosófico mal planteado en contra de una corriente de análisis estadístico posicionada en el mundo científico, en donde más de la mitad de los artículos científicos publicados en las más impor-

¹El diseño en la recolección de los datos es informativo (y no ignorable), no solo en diseño experimental, sino en el análisis de encuestas probabilísticas, para las cuales las propiedades de los estimadores están supeditadas a la medida de probabilidad discreta inducida por el diseño muestral (Gutiérrez 2009).

²En el planteamiento de ensayos clínicos, este enfoque ha sido estudiado por Carlin & Louis (1996) y recientemente por Fúquene et al. (2009) y Cook et al. (2011).

tantes revistas estadísticas utilizan métodos bayesianos (Andrews & Baguley 2013, Figura 1).

Parafraseando a Gelman (2008) mi conclusión final sobre esta interesante discusión redundante en que hoy en día es posible ser no bayesiano –al utilizar en la práctica metodologías clásicas– pero ser antibayesiano –al desconocer el desarrollo científico que la estadística bayesiana ha traído consigo– ya no es una opción.

Recibido: 23 de julio de 2013

Aceptado: 14 de agosto de 2013

Referencias

- Andrews, M. & Baguley, T. (2013), ‘Prior approval: the growth of Bayesian methods in psychology’, *British Journal of Mathematical and Statistical Psychology* **66**, 1–7.
- Bernardo, J. M. (2008), ‘Comment on article by gelman’, *Bayesian Analysis* **3**(3), 443–658.
- Birnbaum, A. (1962), ‘On the foundations of statistical inference’, *Journal of American Statistical Association* **57**, 269–326.
- Carlin, B. P. & Louis, T. (1996), *Identifying prior distributions that produce specific decisions, with application to monitoring clinical trials*, Wiley, chapter Bayesian Analysis in Statistics and Econometrics: Essays in Honor of Arnold Zellner.
- Carlin, B. P. & Louis, T. A. (2009), *Bayesian methods for data analysis*, 3 edn, CRC.
- Cook, J., Fúquene, J. & Pericchi, L. (2011), ‘Skeptical and optimistic robust priors for clinical trials’, *Revista Colombiana de Estadística* **34**, 333–345.
- Cox, D. & Donnelly, C. (2011), *Principles of applied statistics*, Cambridge University Press.
- Fúquene, J., Cook, J. & Pericchi, L. (2009), ‘A case for robust bayesian priors with applications to clinical trials’, *Bayesian Analysis* **4**, 817–846.
- Gelman, A. (2008), ‘Objections to bayesian statistics’, *Bayesian Analysis* **3**(3), 445–450.
- Gelman, A. & Cosma, R. (2013), ‘Philosophy and the practice of bayesian statistics’, *British Journal of Mathematical and Statistical Psychology* **66**, 8–38.
- Gelman, A. & Robert, C. P. (2013), ‘Not only defended but also applied”: The perceived absurdity of bayesian inference’, *The American Statistician* **67**, 1–5.

- Gorroochurn, P. (2011), 'Errors of probability in historical context', *American Statistician* **65**, 246–254.
- Gustafson, P., McCandless, L., Levy, A. & Richardson, S. (2010), 'Simplified bayesian sensitivity analysis for mismeasured and unobserved confounders', *Biometrics* **66**, 1129–1137.
- Gutiérrez, A. (2009), *Estrategias de muestreo: diseño de encuestas y estimación de parámetros*, Editorial de la Universidad Santo Tomás.
- Kadane, J. (2008), 'Comment on article by Gelman', *Bayesian Analysis* **3**(3), 455–458.
- Lindley, D. & Philips, L. (1976), 'Inference from a Bernoulli process (a bayesian view)', *Journal of American Statistical Association* **30**, 112–119.
- Popper, K. (1957), 'Probability magic or knowledge out of ignorance', *Dialectica* **11**(354 - 374).
- Salsburg, D. (2001), *The lady tasting tea*, Henry Hold and Company.
- Schay, G. (2007), *Introduction to probability with statistical applications*, Birkhauser.
- Wikipedia (2013), 'Rule of succession'. Fecha de acceso: 17 de julio de 2013.
*http://en.wikipedia.org/wiki/Rule_of_succession