

# Enseñando la utilidad del Teorema de Bayes (y del Nomograma de Fagan) en medicina diagnóstica

ANDREA CANALS CIFUENTES  
PROGRAMA DE BIOESTADÍSTICA, ESCUELA DE SALUD PÚBLICA, FACULTAD DE MEDICINA  
UNIVERSIDAD DE CHILE  
DIRECCIÓN ACADÉMICA, CLÍNICA SANTA MARÍA. SANTIAGO, CHILE.

MAURICIO CANALS LAMBARRI  
PROGRAMA DE SALUD AMBIENTAL, ESCUELA DE SALUD PÚBLICA, FACULTAD DE MEDICINA  
UNIVERSIDAD DE CHILE  
DEPARTAMENTO DE MEDICINA (O), UNIVERSIDAD DE CHILE  
SERVICIO DE RADIOLOGÍA, HOSPITAL DEL SALVADOR. SANTIAGO, CHILE.

## Resumen

La inferencia bayesiana puede ser útil para resolver diferentes tipos de problemas. En el análisis de resultados de pruebas diagnósticas es especialmente útil ya que nos permite obtener la probabilidad de un diagnóstico dados los resultados de una prueba. Los objetivos de este estudio son presentar las diferentes formas de expresión del Teorema de Bayes, su aplicación a la medicina basada en la evidencia, y mostrar con un ejemplo pertinente y relevante su utilidad. Se revisó literatura sobre inferencia bayesiana, enseñanza del Teorema de Bayes, y su aplicación al análisis de pruebas diagnósticas. A partir de datos sobre PCR en tiempo real (RT-PCR), radiografía de tórax, y TAC de tórax, se muestra una aplicación de este teorema y la utilidad del nomograma de Fagan. El ejemplo presentado en este artículo nos muestra la gran aplicabilidad del Teorema de Bayes y la inferencia bayesiana en problemas de gran relevancia como la pandemia del COVID-19, y la importancia de su enseñanza.

## 1. Introducción

En el Teorema de Bayes convergen el pensamiento inductivo y el deductivo, exigiendo un proceso de pensamiento inverso al de la probabilidad frecuentista. Es por esto que, en la enseñanza de este teorema, las y los docentes necesitan no solo un buen dominio del contenido, sino también una clara comprensión de las probabilidades condicionales, de la probabilidad total, del teorema mismo, y de las implicaciones que tiene en la toma de decisiones [1].

Se han realizado esfuerzos para obtener los mejores métodos de enseñanza de este teorema y se han estudiado posibles inconvenientes para su enseñanza en cursos de pregrado [2, 3]. Por ejemplo, Berry [3] discute algunas posibles razones que dificultan su enseñanza, entre las que menciona: 1) la estadística bayesiana es inherentemente demasiado difícil de enseñar en un nivel elemental, 2) el enfoque bayesiano es subjetivo y no proporciona métodos que cumplan con los estándares de objetividad requeridos por la ciencia, 3) los bayesianos no pueden ponerse de acuerdo sobre distribuciones previas apropiadas, 4) los métodos basados en el paradigma frecuentista dominan en las ciencias.

Hoy en día, con el avance en el desarrollo de la

inferencia bayesiana y su alcance en la toma de decisiones, y en disciplinas como la medicina, se hace cada vez más necesaria una adecuada enseñanza del Teorema de Bayes en los cursos de pregrado. Por ejemplo, la medicina basada en la evidencia y la medicina diagnóstica necesitan una comprensión adecuada de este [4, 5, 6, 7, 8].

El propósito de este trabajo es resaltar los aspectos conceptuales más importantes del Teorema de Bayes, y mostrar, a través de una aplicación a un tema de gran actualidad como es la pandemia del COVID-19, la importancia de su conocimiento, alcance y enseñanza. La Sección 2 aborda los conceptos elementales del teorema, la Sección 3 considera su relación con la medicina basada en evidencia y la Sección 4 muestra una aplicación a las imágenes para el diagnóstico de COVID-19.

## 2. Teorema de Bayes

La forma clásica de este teorema se puede presentar de la siguiente manera. Llamaremos partición del espacio muestral a un conjunto de eventos  $B_i$ , mutuamente excluyentes, y que juntos conforman el espacio muestral de un experimento aleatorio ( $\cup B_i = \Omega$ ). Si consideramos cualquier evento  $A$ , que puede ocurrir simultáneamente con los eventos  $B_i$ , podemos obtener

la probabilidad de  $B_i$  dado  $A$  a partir de la siguiente expresión:

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{P(A)}$$

Las expresiones  $P(B_i)$  y  $P(B_i|A)$  corresponden a las probabilidades *a priori* y *a posteriori* de  $B_i$ , es decir,  $P(B_i)$  corresponde a la probabilidad de  $B_i$  desconociendo si  $A$  ha ocurrido o no, y  $P(B_i|A)$  es la probabilidad de  $B_i$  cuando ha ocurrido  $A$ . Si ahora suponemos que podemos usar una distribución de probabilidades para modelar lo que se conoce acerca del valor de un parámetro  $\theta$ , podemos usar una segunda forma de presentación del Teorema de Bayes, que da lugar a la inferencia bayesiana:

$$\pi(\theta|Y) = \frac{P(Y|\theta)\pi(\theta)}{P(Y)}$$

En esta relación,  $\pi(\theta)$  y  $\pi(\theta|Y)$  corresponden a las probabilidades *a priori* y *a posteriori* de un parámetro  $\theta$ , cada una asociada a una distribución de probabilidad de este parámetro.  $Y$  corresponde a los datos observados en una muestra, la probabilidad condicional  $P(Y|\theta)$  coincide con la función de verosimilitud de los datos, dado el parámetro  $\theta$ , y  $P(Y)$  es la probabilidad marginal de los datos. La probabilidad  $P(Y|\theta)$  puede ser escrita como  $l(\theta; Y)$  conduciendo a la expresión proporcional del teorema de Bayes:

$$\pi(\theta|Y) \propto l(\theta; Y)\pi(\theta)$$

Este teorema permite medir directamente la probabilidad de una hipótesis de investigación. Analicemos el siguiente desarrollo. Si solo son posibles dos hipótesis complementarias,  $H_0$  y  $H_1$ , y queremos medir directamente la probabilidad de que  $H_1$  sea verdadera habiendo obtenido un conjunto de datos  $X_i$ , aplicando directamente el Teorema de Bayes se obtiene:

$$\begin{aligned} P(H_1|X_i) &= \frac{P(X_i|H_1)P(H_1)}{P(X_i|H_1)P(H_1) + P(X_i|H_0)P(H_0)} \\ &= \frac{P(X_i|H_1)P(H_1)}{P(X_i)} \end{aligned}$$

y

$$\begin{aligned} P(H_0|X_i) &= \frac{P(X_i|H_0)P(H_0)}{P(X_i|H_1)P(H_1) + P(X_i|H_0)P(H_0)} \\ &= \frac{P(X_i|H_0)P(H_0)}{P(X_i)} \end{aligned}$$

Dividiendo ambas expresiones obtenemos:

$$\frac{P(H_1|X_i)}{P(H_0|X_i)} = \frac{P(X_i|H_1)P(H_1)}{P(X_i|H_0)P(H_0)}$$

Esta razón mide cuánto más probable es  $H_1$  que  $H_0$ , dados los datos. Esto es, las “chances” u *odds* a

*posteriori* de  $H_1$  ( $O_p H_1$ ) [9, 10]. Recordemos que las *odds* o chances de un evento corresponden al cociente entre las probabilidades del evento y su complemento (*odds* =  $p/(1-p)$ ), expresión de la cual podemos obtener también que  $p = \text{odds}/(\text{odds} + 1)$ .

Entonces, conociendo las *odds a posteriori* de  $H_1$  ( $O_p H_1$ ) podemos calcular su probabilidad como

$$P(H_1|X_i) = \frac{O_p H_1}{O_p H_1 + 1}$$

### 3. Teorema de Bayes en medicina basada en la evidencia

El teorema de Bayes tiene varias aplicaciones y es especialmente útil cuando se analizan datos relacionados con pruebas de diagnóstico. En medicina es común el uso de pruebas para diagnosticar enfermedades, sin embargo, las pruebas no son 100% exactas para hacer un diagnóstico. A veces un individuo tiene una enfermedad y la prueba es negativa (falso negativo) y otras veces la prueba es positiva para la enfermedad y el paciente no la tiene (falso positivo) (Tabla 1).

**Tabla 1:** Posibles resultados de una prueba diagnóstica en pacientes enfermos y no enfermos.

	Resultado prueba		Total
	Positiva (+)	Negativa (-)	
Enfermos $E$	$a$	$b$	$a + b$
No enfermos $E^C$	$c$	$d$	$c + d$
<b>Total</b>	$a + c$	$b + d$	$n$

En este caso, la cantidad de falsos positivos sería  $c$  y la cantidad de falsos negativos sería  $b$ . A partir de esta tabla se pueden definir ciertos conceptos:

- La sensibilidad ( $S$ ) de una prueba corresponde a la probabilidad de que la prueba sea positiva dado que el paciente está enfermo:

$$S = P(+|E) = \frac{a}{a + b}$$

La sensibilidad representa la capacidad de la prueba para detectar una enfermedad.

- La especificidad ( $S_p$ ) de la prueba corresponde a la probabilidad de que la prueba sea negativa dado que el paciente no tiene esta enfermedad (aunque tenga otra enfermedad):

$$S_p = P(-|E^C) = \frac{d}{c + d}$$

Es decir, mide la capacidad de la prueba para responder a esa y solo esa enfermedad.

Otra medida interesante es la razón de verosimilitud ( $LR$ ) que se define como el cociente entre la probabilidad de un determinado resultado de la prueba dado que el paciente está enfermo, dividida por la probabilidad de ese mismo resultado, dado que no lo está. Por ejemplo, para un resultado positivo:

$$LR(+)=\frac{P(+|E)}{P(+|E^C)}=\frac{P(+|E)}{1-P(-|E^C)}=\frac{S}{1-S_p}$$

y, análogamente, para un resultado negativo:

$$LR(-)=\frac{P(-|E)}{P(-|E^C)}=\frac{1-P(+|E)}{P(-|E^C)}=\frac{1-S}{S_p}$$

Por ejemplo, es interesante saber qué ocurre con un paciente que tiene una probabilidad *a priori* “ $P$ ” de tener una enfermedad, al que se le aplica la prueba y se obtiene un resultado positivo. Para este paciente, sus *odds a priori* corresponderán a  $O_0 = P/(1 - P)$ , y usando el Teorema de Bayes, podemos obtener sus *odds a posteriori* ( $O_p$ ):

$$O_p = \frac{P(E|+)}{P(E^C|+)}$$

donde

$$P(E|+) = \frac{P(+|E)P}{P(+|E)P + P(+|E^C)(1 - P)} = \frac{SP}{SP + (1 - S_p)(1 - P)}$$

$$P(E^C|+) = 1 - P(E|+) = 1 - \frac{SP}{SP + (1 - S_p)(1 - P)} = \frac{(1 - S_p)(1 - P)}{SP + (1 - S_p)(1 - P)}$$

Luego:

$$O_p = \frac{P(E|+)}{P(E^C|+)} = \frac{S}{1 - S_p} \frac{P}{1 - P} = LR(+)\ O_0$$

Este resultado nos muestra que las *odds a posteriori* de presentar la enfermedad, dado un resultado positivo, corresponden al producto de las *odds a priori* por la razón de verosimilitudes positiva. Del mismo modo, se puede obtener que las *odds a posteriori* de presentar la enfermedad, dado un resultado negativo, corresponden al producto de las *odds a priori* por la razón de verosimilitudes negativa.

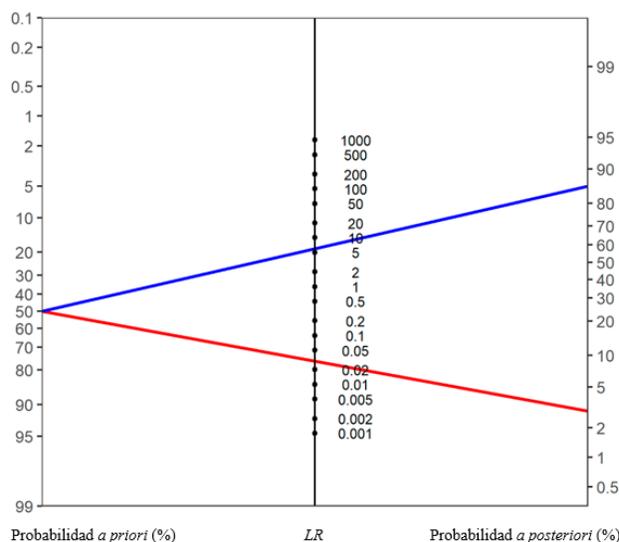
#### 4. Nomograma de Fagan

Como los conceptos de *odds* ( $O$ ) y probabilidad ( $p$ ) son intercambiables a través de la relación  $O =$

$p/(1 - p)$ , o  $p = O/(O + 1)$ , siempre es posible calcular la probabilidad *a posteriori* cuando se ha obtenido una prueba positiva, en base a la probabilidad *a priori* (presunción) y la  $LR$  de la prueba.

Utilizando esta última consideración, Fagan, en el año 1975 [11], propuso un nomograma para calcular las probabilidades *a posteriori*. En este nomograma hay tres ejes graduados. El eje izquierdo representa la probabilidad previa a la prueba, o presunción, el eje central representa el valor de  $LR$  (positivo y negativo) y el eje derecho muestra la probabilidad *a posteriori* (Figura 1). Luego, se dibuja una línea recta desde los dos puntos marcados en los ejes izquierdo y medio hacia el eje derecho. El punto en el que la línea cruza el eje derecho corresponde al valor de la probabilidad *a posteriori*.

La Figura 1 nos muestra que, por ejemplo, si la probabilidad *a priori* de un evento es de 0,5 y las razones de verosimilitud positiva y negativa son de 6 y 0,03, respectivamente, las probabilidades *a posteriori* de ocurrencia y no ocurrencia del evento serán de 86 % y 3 %, respectivamente.



**Figura 1:** Nomograma de Fagan considerando una probabilidad *a priori* de 0,5, razón de verosimilitud positiva de 6 y razón de verosimilitud negativa 0,03. La línea azul muestra la probabilidad *a posteriori* de ocurrencia del evento, y la línea roja la probabilidad *a posteriori* de no ocurrencia del evento. Realizado en R (función `nomogrammer`).

#### 5. Aplicación a la radiología del COVID-19

Recientemente se han propuesto para PCR en tiempo real (RT-PCR), radiología de tórax (RxTx) y TAC de Tórax (CT), los siguientes valores de  $S$ ,  $S_p$  y  $LR$  para resultados positivos y negativos para el diagnóstico de COVID-19 [8], mostrados en la Tabla 2.

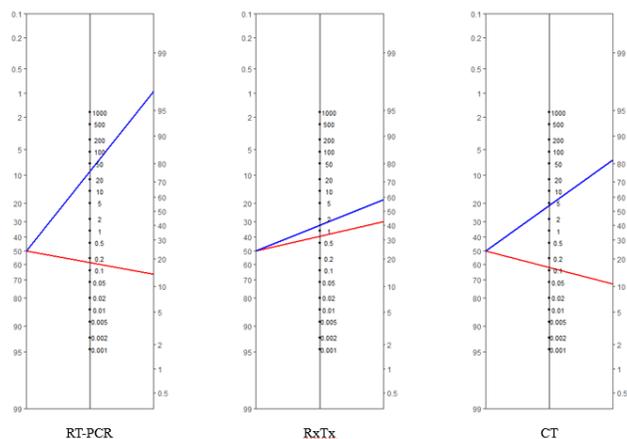
**Tabla 2:** Sensibilidad ( $S$ ), Especificidad ( $S_p$ ), razones de verosimilitud positiva ( $LR(+)$ ) y negativa ( $LR(-)$ ), según examen.

Examen	$S$	$S_p$	$LR(+)$	$LR(-)$
RT-PCR <sup>(1)</sup>	0,846	0,974	32,538	0,158
RxTx <sup>(2)</sup>	0,560	0,600	1,400	0,733
CT <sup>(3)</sup>	0,905	0,797	4,451	0,119

(1) PCR en tiempo real. (2) Radiografía de tórax.

(3) TAC (escáner) de tórax.

Estos indicadores permiten estimar las probabilidades post-test de tener COVID-19 en un paciente con una presunción de un 50% de tener COVID-19 antes del test (Figura 2).



**Figura 2:** Nomograma de Fagan para estimar las probabilidades post-test para resultados positivos y negativos para COVID-19 en un paciente con una presunción de un 50% de tener COVID-19 antes del test. De izquierda a derecha: RT-PCR, RxTx en negro y CT. Líneas azules: presencia de la enfermedad, líneas rojas: ausencia de ésta.

Este diagrama de Fagan nos muestra que el mejor predictor de enfermedad ante un resultado positivo es el RT-PCR, pero que ante un resultado negativo es mejor el CT. Nos muestra además que la RxTx aporta poco, pero que el CT es un buen examen, produciendo altas probabilidades post-test. Nos muestra además que los mejores test son los que tienen mayores ángulos entre las líneas asociadas a  $LR(+)$  y las líneas finas asociadas a  $LR(-)$ .

Así, la aplicación del Teorema de Bayes al análisis de pruebas diagnósticas nos permite actualizar fácilmente la probabilidad de determinados diagnósticos y saber cómo cambia ante los resultados de una prueba. El Nomograma de Fagan nos permite graficar y visualizar de manera rápida la utilidad de un test en un diagnóstico particular como el COVID-19. El ejemplo presentado nos muestra la gran aplicabilidad del Teorema de Bayes y la inferencia bayesiana en problemas de gran relevancia como la pandemia del COVID-19.

## Referencias

- [1] Moreno, J.C.T., *et al*, Un análisis de los significados de enseñanza del Teorema de Bayes, *Conference paper*, 2016.
- [2] Albert, J., Teaching introductory statistics from a Bayesian perspective, In *Proceedings of the Sixth International Conference on Teaching Statistics*. Cape Town, South Africa, 2002, Voorburg, The Netherlands: International Statistical Institute [Online: [www.stat.auckland.ac.nz/~iase/publications/1/3f1\\_albe.pdf](http://www.stat.auckland.ac.nz/~iase/publications/1/3f1_albe.pdf)].
- [3] Berry, D.A., Teaching elementary Bayesian statistics with real applications in science, *The American Statistician*, 1997; 51(3): 241-246.
- [4] Haynes, R.B., *et al*, Evidence-based medicine: How to practice & teach EBM, *Canadian Medical Association Journal*, 1997; 157(6): 788.
- [5] Canals, M., Bases cuantitativas del análisis de imágenes médicas: información e inferencia bayesiana, *Revista Médica de Chile*, 2003; 131(5): 553-561.
- [6] Canals, M., *et al*, Neumonías en el inmunocomprometido: perspectiva desde el diagnóstico por imágenes, e inferencia Bayesiana, *Revista Chilena de Infectología*, 2014; 31(2): 139-152.
- [7] Canals, M., Bases científicas del razonamiento clínico: inferencia Bayesiana, *Revista Médica de Chile*, 2019; 147(2): 231-237.
- [8] Canals, M., Canals, A., How accurate are radiography and computed tomography in the diagnosis of COVID-19? A Bayesian approach, *Acta Radiologica Open*, 2022 (en prensa).
- [9] Bernardo, J.M., Smith, A.F., Bayesian theory (Vol. 405), *John Wiley & Sons*, 2009.
- [10] Lee, P.M., Bayesian Statistics: An Introduction, *Oxford University Press*, 1989.
- [11] Fagan, T.J., Letter: Nomogram for Bayes Theorem, *N. Engl. J. Med.*, 1975; 293(5): 257.

---

## Anexo: Script de R

```
library(ggplot2)
library(scales)
library(ggpubr)
source("https://raw.githubusercontent.com/
      achekroud/nomogrammer/master/nomogrammer.r")

figura1 <- nomogrammer(Prevalence = .50, Plr = 6,
                      Nlr = 0.03, Verbose=TRUE)
+ scale_color_manual(values = c("red", "blue"))

a = nomogrammer(Prevalence = .50, Plr = 32.538,
                Nlr = 0.158)
+ scale_color_manual(values = c("red", "blue"))

b = nomogrammer(Prevalence = .50, Plr = 1.4,
                Nlr = 0.733)
+ scale_color_manual(values = c("red", "blue"))

c = nomogrammer(Prevalence = .50, Plr = 4.451,
                Nlr = 0.119)
+ scale_color_manual(values = c("red", "blue"))

figura2 <- ggarrange(a, b, c, ncol = 3, nrow = 1)
```