

# Teorema de Bayes para datos de Covid-19

---

Alberto Benavides  
26 de octubre de 2020

## 1. TEOREMA DE BAYES

Según Grinstead y Snell [4], en problemas en los que hay  $m$  hipótesis  $H_i, i = [1, m]$  que pueden ser confirmadas por alguna evidencia  $E$ , se puede utilizar el teorema de Bayes para conocer la probabilidad de una determinada hipótesis dada la evidencia  $P(H_i|E)$  según la fórmula

$$P(H_i|E) = \frac{P(H_i)P(E|H_i)}{\sum_{k=1}^m P(H_k)P(E|H_k)}. \quad (1.1)$$

## 2. PRUEBAS DE DETECCIÓN DE COVID-19

Estos conceptos de la probabilidad se pueden aplicar a ramas de la epidemiología donde se usan pruebas para detectar enfermedades. En 2020 se vive una pandemia dada por el contagio del virus SARS-CoV-2, responsable de la enfermedad Covid-19 [7]. Existen diferentes tipos de pruebas para detectar el contagio presente o pasado de esta enfermedad [1], las cuales tienen algunas características [6] que ayudan a determinar la posibilidad de que sus resultados sean correctos. En este respecto, existen cuatro posibilidades (véase la tabla 2.1 en la p. 2):

- Verdadero positivo (VP): El paciente está infectado  $p$  y fue diagnosticado positivo  $p'$ .
- Falso positivo (FP): El paciente no está infectado  $n$  y fue diagnosticado positivo  $p'$ .
- Falso negativo (FN): El paciente está infectado  $p$  y fue diagnosticado negativo  $n'$ .

Tabla 2.1: Representación visual de las combinaciones entre diagnóstico e infección de pacientes de Covid-19.

		Infección	
		$p$	$n$
Diagnóstico	$p'$	Verdadero positivo	Falso positivo
	$n'$	Falso negativo	Verdadero negativo

- Verdadero negativo (VN): El paciente no está infectado  $n$  y fue diagnosticado negativo  $n'$ .

Además, a partir de estas posibilidades se pueden calcular

- Exactitud:  $\frac{VP+VN}{P+N}$ ,
- Precisión:  $\frac{VP}{VP+FP}$ ,
- Sensibilidad:  $\frac{VP}{VP+FN}$ ,
- Especificidad:  $\frac{FN}{VN+FP}$ .

### 3. DATOS UTILIZADOS

Con todo esto, se puede aplicar este teorema a valores de contagio obtenidos de datos reales. Se extraen de la Secretaría de Salud de México [3] datos de contagios de Covid-19 en México. De estos datos se extraen solamente los casos que se consideran positivos y negativos, o sea  $p' = 864,696$  positivos y  $n' = 1,072,760$  negativos; con un total de 1,937,456 pruebas.

### 4. EXACTITUD DE LAS PRUEBAS

Entre las distintas pruebas que existen, el porcentaje de exactitud puede variar desde un 20 % a un 80 % según las características y tipos de pruebas aplicadas. Los porcentajes más bajos corresponden a pruebas consideradas rápidas que se realizan a partir de muestras de sangre. Los resultados altos para las pruebas se asocian a pruebas virales (de mucosas y tejidos del sistema respiratorio). Para pacientes asintomáticos, las pruebas virales tienen una exactitud de 30 % a 50 %, mientras que en los pacientes con síntomas, están en el rango de 60 % a 80 % [2, 5]. Generalmente, el porcentaje de exactitud para pruebas bayesianas que se suele elegir es de 70 %, por lo que se tomará dicho valor en este análisis también.

### 5. ESTIMACIONES A PARTIR DE LOS DATOS

De los valores que se tienen a partir de los datos de contagios obtenidos de la Secretaría de Salud de México y del 70 % estimado de exactitud de las pruebas, se pueden obtener la

Tabla 5.1: Diagnósticos e infecciones a partir de los datos de la Secretaría de Salud de México y el estimado de 70 % de exactitud en las pruebas realizadas.

	$p$	$n$
$p'$	605,287	259,409
$n'$	321,828	750,932

cantidad de verdaderos positivos  $VP = 0.7p'$ , falso positivo  $FP = 0.3n'$ , falsos negativos  $FN = 0.3p'$ , y verdaderos negativos  $VN = 0.7n'$ . Esto queda representado en la tabla 5.1. Con estos resultados, se puede estimar el número de contagiados en México es  $p = VP + FN = 927,115$ , en tanto el número de personas no contagiadas que se hicieron las pruebas sería  $n = p + VN + FP = 1,010,341$ . Estos estimados incrementan un 7.2 % los casos positivos y disminuyen un 5.9 % los casos negativos, en relación a los reportados.

## 6. TEOREMA DE BAYES APLICADO A PRUEBAS DE COVID-19

En el caso de las pruebas de Covid-19, se pueden definir algunas variables. Por ejemplo,  $H_1$  : “estoy contagiado de Covid-19”;  $E$  : “la prueba salió positiva”. De este modo, se podría desear averiguar si un paciente tiene Covid-19 dado que recibió una prueba con resultado positivo, esto es  $P(H_1|E)$ . Por otro lado,  $H_2$  : “no estoy contagiado de Covid-19”. De aquí, se tiene  $P(H_1) = \frac{927,115}{1,937,456} = 0.4785$  como la proporción de pacientes con Covid-19 entre todos los que se hicieron la prueba; y  $P(H_2) = 1 - P(H_1) = 0.5215$  en tanto la proporción de pacientes no contagiados de Covid-19 de entre los que se hicieron la prueba.

Así, se pueden obtener algunos otros datos de interés como la proporción de pruebas positivas para los pacientes que sí están infectados  $P(E|H_1) = 0.7$  y para los que no lo están  $P(E|H_2) = 0.3$ .

Con estos valores, se puede calcular  $P(H_1|E)$  a partir de la ecuación 1.1 como sigue:

$$P(H_1|E) = \frac{P(H_1)P(E|H_1)}{P(H_1)P(E|H_1) + P(H_2)P(E|H_2)} = 0.6816,$$

es decir que se tiene una probabilidad de estar contagiado de un 68.16 % si es que una prueba sale positiva, a partir de los supuestos aquí descritos.

## 7. NOTAS ADICIONALES

El teorema de Bayes es una manera de conocer probabilidades de eventos dependientes entre sí que puede aplicarse para dar estimaciones más certeras a partir de resultados observados. En este caso, es posible repetir los cálculos con valores que reflejen la prueba usada para la detección del Covid-19 y sus porcentajes de exactitud asociados, además del número de pruebas realizadas, los casos positivos y negativos, tanto falsos como verdaderos.

## REFERENCIAS

- [1] FOOD, U. S. y DRUG ADMINISTRATION (2020), «Conceptos básicos de las pruebas para el coronavirus», <https://www.fda.gov/consumers/articulos-en-espanol/conceptos-basicos-de-las-pruebas-para-el-coronavirus>.
- [2] GARCÍA, S. (2020), «¿Qué tan certero puede ser el resultado de una prueba para diagnosticar Covid-19?», <https://verificado.com.mx/que-tan-certero-puede-ser-el-resultado-de-una-prueba-para-diagnosticar-la-covid-19/>.
- [3] GOBIERNO DE MÉXICO (2020), «Covid-19 México», <https://datos.covid-19.conacyt.mx/#DownZCSV>.
- [4] GRINSTEAD, C. M. y J. L. SNELL (1997), *Introduction to probability*, American Mathematical Society.
- [5] LISBOA BASTOS, M., G. TAVAZIVA, S. K. ABIDI, J. R. CAMPBELL, L.-P. HARAOU, J. C. JOHNSTON, Z. LAN, S. LAW, E. MACLEAN, A. TRAJMAN, D. MENZIES, A. BENEDETTI y F. AHMAD KHAN (2020), «Diagnostic accuracy of serological tests for Covid-19: systematic review and meta-analysis», *BMJ*, **370**.
- [6] RANJAN, A. (2020), «Covid-19, Bayes' theorem and taking probabilistic decisions», <https://towardsdatascience.com/covid-19-bayes-theorem-and-taking-data-driven-decisions-part-1-b61e2c2b3bea>.
- [7] SCHNIPPER, J. L. y P. E. SAX (2020), «Covid-19 test accuracy supplement: The math of Bayes' theorem», <https://www.statnews.com/2020/08/20/covid-19-test-accuracy-supplement-the-math-of-bayes-theorem/>.