

# Tarea 8 de Modelos Probabilistas Aplicados

Aplicación del teorema de Bayes en datos relacionados con el Covid-19

5271

27 de octubre de 2020

## 1. Introducción

En este trabajo se realiza un acercamiento a uno de los temas más relevantes de la actualidad, la pandemia del Covid-19. En el mismo se realiza un análisis sobre artículos que emplean el teorema de Bayes en datos de las pruebas diagnóstico realizadas en la detección de la enfermedad, teniendo en cuenta la sensibilidad, especificidad y valores predictivos (positivos, negativos) de las pruebas. Además se presenta el cálculo de los valores predictivos (positivos, negativos) con datos abiertos sobre la aplicación de pruebas de Covid-19 en México.

## 2. Teorema de Bayes

Sea  $B_1, B_2, \dots, B_n$  una partición del espacio muestral  $(\Omega)$  tal que  $P(B_i) \neq 0$  para  $i = 1, 2, \dots, n$  y sea  $A$  un evento tal que  $P(A) \neq 0$ . Entonces para cada  $j = 1, 2, \dots, n$  se tiene:

$$P(B_j | A) = \frac{P(A | B_j)P(B_j)}{\sum_{i=1}^n P(A | B_i)P(B_i)}. \quad (1)$$

## 3. Sensibilidad y especificidad de una prueba diagnóstico

En esta sección se explicarán dos conceptos básicos a tener en cuenta en el análisis de una prueba diagnóstico, como son la sensibilidad y la especificidad.

La sensibilidad representa la probabilidad de que una persona enferma obtenga un resultado positivo. Esta se define como:

$$\blacksquare \text{ Sensibilidad} := \frac{VP}{VP+FN}$$

$VP$  (verdaderos positivos) y  $FN$  (falsos negativos).

La especificidad representa la probabilidad de que una persona sana obtenga un resultado negativo. Esta se define como:

Cuadro 1: Matriz de confusión con posibles resultados de una prueba diagnóstico

		Valor real		Total
		$p$	$n$	
Predicción	$p'$	Verdaderos Positivos	Falsos Positivos	$P'$
	$n'$	Falsos Negativos	Verdaderos Negativos	$N'$
Total		$P$	$N$	

- *Especificidad*  $:= \frac{VN}{VN+FP}$   
 $VP$  (verdaderos negativos) y  $FN$  (falsos positivos).

A partir de estos conceptos se puede crear una matriz de confusión donde se reflejen los cuatro posibles resultados de un experimento a partir de  $P$  instancias positivas y  $N$  instancias negativas. Esta matriz se puede observar en el cuadro 1 de la página 2.

De la matriz en el cuadro 1 se derivan los conceptos, valores predictivos (positivo y negativo) que son hallados por el teorema de Bayes. Estos valores miden la eficacia real de una prueba diagnóstica, los mismos dan la probabilidad de padecer o no una enfermedad una vez conocido el resultado de la prueba. A continuación, se muestra cómo se definen dichos valores.

- Valor predictivo positivo ( $PV+$ ): es la probabilidad de tener la enfermedad si el resultado de la prueba es positivo.

$$PV+ := \frac{VP}{FP+VP}$$

- Valor predictivo negativo ( $PV-$ ): es la probabilidad de no tener la enfermedad si el resultado de la prueba es negativo.

$$PV- := \frac{VN}{VN+FN}$$

## 4. Análisis de artículos donde se aplica el teorema de Bayes

En esta sección se presenta un resumen de lo propuesto de varios artículos que utilizan el teorema de Bayes con datos de pruebas diagnóstico que para detectan el Covid-19. El objetivo de estos documentos es poder hallar con que probabilidad una persona que da positivo en un examen está realmente enferma y con los pacientes que resultan negativos que probabilidad existe que estas personas realmente no posean la enfermedad.

### 4.1. Suplemento de precisión de la prueba Covid-19: las matemáticas del teorema de Bayes

En el artículo “Suplemento de precisión de la prueba Covid-19: las matemáticas del teorema de Bayes” disponible en la liga: (<https://www.statnews.com/2020/08/20/covid-19-test-accuracy-supplement-the-math-of-bayes-theorem/>), los autores aplican los valores predictivos (positivo y negativo), variando la prevalencia ( $P$ ) de la enfermedad.

En el primer caso toman como prevalencia o probabilidad de estar enfermo bajo, con el valor de  $P = 1\%$  y valores de sensibilidad ( $SE$ ) =  $80\%$  y especificidad ( $SP$ ) =  $100\%$ . Dando como resultado  $PV+ = 100\%$  y un  $PV- = 99,8\%$ . En el segundo caso se mantienen los valores de  $SE$  y  $SP$  y se varia la prevalencia con  $P = 30\%$  esto significa que la probabilidad de dar positivo es mayor. Con estos valores se obtiene  $PV+ = 100\%$  y un  $PV- = 92\%$ . Aquí se puede observar cómo afecta la prevalencia en los valores predictivos.

## 4.2. Teorema de Bayes y pruebas de Covid-19

En el artículo “Teorema de Bayes y pruebas de Covid-19” disponible en la liga: (<https://www.significance-magazine.com/science/660-bayes-theorem-and-covid-19-testing>), el autor calcula al igual que en artículo anterior el valor  $PV+$  mediante la aplicación del teorema de Bayes para un  $SE = 99\%$  y un  $SP = 99\%$ . Con valores de  $P = (0,1\%, 1\%, 10\%)$  en cada caso. Con lo cual se muestra que incluso cuando se utiliza una prueba muy sensible  $SE = 99\%$  cuanto menor sea la prevalencia, más probabilidades tenemos de obtener falsos positivos. También se puede demostrar que cuanto mayor sea prevalencia, más probabilidades tenemos de obtener falsos negativos. Esto significa que la calidad de las pruebas de Covid-19 depende de la magnitud del brote.

## 4.3. COVID-19, teorema de Bayes y toma de decisiones probabilísticas

Este artículo disponible en la liga: (<https://towardsdatascience.com/covid-19-bayes-theorem-and-taking-data-driven-decisions-part-1-b61e2c2b3bea>), plantea la realización de pruebas de personas al azar en lugares con prevalencia de la enfermedad muy pequeña ( $0.0001$ ) no resulta útil en la detección del Covid-19. Además, plantea que en estos casos sería mejor hacerlo por grupos de muestras.

## 4.4. Interpretación de los resultados de la prueba COVID-19: un enfoque bayesiano y Teorema de Bayes, COVID19 y pruebas de detección

En los artículos “Interpretación de los resultados de la prueba COVID-19: un enfoque bayesiano” y “Teorema de Bayes, COVID19 y pruebas de detección” disponibles en las ligas: (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7269418/>) y (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7315940/>) respectivamente, llegan a una conclusión similar planteando que las acciones y recomendaciones posteriores al diagnóstico se basan en gran medida en las probabilidades posteriores a la prueba, no en los resultados categóricos de “positivo.” o “negativo”. Por lo que un resultado negativo de una prueba sin conocer el  $VP-$  de un individuo puede limitar la capacidad de los médicos para realizar las siguientes acciones y disposiciones apropiadas. Para todas las pruebas de detección, ya sea para COVID19 u otros diagnósticos, la comprensión de los valores predictivos y las razones de probabilidad con la ayuda del teorema de Bayes garantizará una interpretación sólida y las recomendaciones y acciones resultantes por parte del personal de salud y las partes interesadas. Un resultado de prueba negativo, en este paradigma, nunca es absolutamente negativo. Más bien, ajusta la probabilidad previa a la prueba de tener una enfermedad más baja.

## 5. Aplicación del teorema de Bayes a datos de pruebas de Covid-19 en México

Como se pudo apreciar en las secciones anteriores en lo que a las pruebas diagnósticos se refiere, no solo interesa si el resultado es “positivo” o “negativo”, además hay que tener en cuenta los valores predictivos (positivo y negativo). Estos últimos nos van a dar la probabilidad de un diagnóstico certero y los pasos a seguir del personal de salud a la hora de descartar o no la enfermedad en los pacientes.

### 5.1. PCR

En México durante el enfrentamiento a la pandemia del Covid-19 la prueba diagnóstica más utilizada es la prueba de detección de ácidos nucleicos: reacción en cadena de la polimerasa (PCR). El PCR tiene las siguientes características [2]:

- Especificidad, próxima al 100 %.
- Sensibilidad variable dependiendo del momento del proceso infeccioso, es decir, de la carga viral, y del lugar de toma de la muestra. Entre cero y siete días tras el comienzo de la enfermedad, las sensibilidades para pacientes leves como severos fueron:
  - Espudo: 89 %
  - Nasal: 73 %
  - Oro-faringe: 60 %

### 5.2. Datos utilizados

Para el cálculo de  $VP+$  y  $VP-$  de pruebas diagnósticas PCR en Mexico, se utilizan datos abiertos proporcionados por el gobierno de México y disponibles en [1], en la utilización del 16 de octubre del 2020. De dichos datos se extraen los siguientes valores:

- Total de pruebas realizadas ( $TP$ ) = 1,903,285.
- Total de pruebas positivas ( $T+$ ) = 837,445.
- Total de pruebas negativas ( $T-$ ) = 1,065,840.
- Tasa positiva ( $P$ ) = 0.295.

Con los valores de  $SE = (60\%, 73\%, 89\%)$ ,  $SP = 99,9\%$  del PCR y los valores  $TP$ ,  $T+$ ,  $T-$  se crean matrices de confusión, para la obtención de los valores a emplear en teorema de Bayes, en el cuadro 2 de la página 5 se muestra un ejemplo de matriz de confección para  $SE = 60\%$ . En el cuadro 3 de la página 5 se puede observar los resultados obtenidos del cálculo de los valores predictivos para cada una de las variantes. Como prevalencia de la enfermedad se utiliza el valor la tasa positiva que es una buena aproximación a la probabilidad previa a la prueba.

De los resultados obtenidos se puede concluir que el personal médico debe tener mayor cuidado con los valores negativos de las pruebas de PCR, dado que los  $VP-$  nos muestran que existe altas probabilidades de falsos negativos cuando la sensibilidad es baja y la prevalencia aumenta, estos son parámetros a tener en cuenta a la hora de emitir el diagnóstico final.

Cuadro 2: Matriz de confusión para  $SE = 60\%$ 

		Valor real		Total
		$p$	$n$	
Predicción	$p'$	502,467.00	1,065.00	P'
	$n'$	334,978.00	1,064,774.00	N'
Total		P	N	

Cuadro 3: Resultados de la aplicación del teorema de Bayes

$SE(\%)$	$SP(\%)$	$P$	$VP+(\%)$	$VP-(\%)$
60.00	99.90	0.3	99.61	85.35
		0.5	99.83	71.41
73.00	99.90	0.3	99.68	89.61
		0.5	99.86	78.72
89.00	99.90	0.3	99.73	95.49
		0.5	99.88	90.08

## Referencias

- [1] Esteban Ortiz-Ospina Max Roser, Hannah Ritchie and Joe Hasell. Coronavirus pandemic (covid-19). *Our World in Data*, 2020. <https://ourworldindata.org/coronavirus>.
- [2] Yang Yang, Minghui Yang, Chenguang Shen, Fuxiang Wang, Jing Yuan, Jinxiu Li, Mingxia Zhang, Zhaoqin Wang, Li Xing, Jinli Wei, Ling Peng, Gary Wong, Haixia Zheng, Mingfeng Liao, Kai Feng, Jianming Li, Qianting Yang, Juanjuan Zhao, Zheng Zhang, Lei Liu, and Yingxia Liu. Evaluating the accuracy of different respiratory specimens in the laboratory diagnosis and monitoring the viral shedding of 2019-ncov infections. *medRxiv*, 2020.