

Aquí va el título. Favor de respetar el formato para los autores y sus direcciones

Ciria, Briones-García      Rubén , Blancas-Rivera  
Víctor, Vázquez-Guevara  
Benemérita Universidad Autónoma de Puebla,  
Facultad de Ciencias Físico Matemáticas,  
Av. San Claudio y 18 Sur, Col. San Manuel,  
C.P. 72570, Puebla, Puebla,  
216470295@alumnos.fcfm.buap.mx, 216470561@alumnos.fcfm.buap.mx,  
vvazquez@fcfm.buap.mx

**Resumen.** Aquí va el resumen.

**Abstract.** Aquí va el resumen en inglés.

*Palabras clave:* Tres o cuatro de preferencia.

## 1. Introducción

Esta es la plantilla para escribir los capítulos de libro. Voy a poner algunos ejemplos de cosas que podrían ser de utilidad.

En la literatura se pueden encontrar diversos trabajos que hablan acerca de la estadística bayesiana, (véase [1] y [5] ). También ha sido aplicada a problemas sociales y políticos como se pueden consultar en [10] y [6]. El objetivo de este trabajo es estudiar el enfoque bayesiano y mostrar algunos métodos para calcular proporciones de votantes y éxito de candidatos a ganar la próxima elección presidencial del 2018.

En la estadística existen dos tipos de enfoques, el clásico (o frecuentista) y el bayesiano. La bayesiana es un tipo de inferencia estadística en la que las evidencias u observaciones se emplean para actualizar la probabilidad de que una hipótesis pueda ser cierta.

El nombre bayesiana proviene del uso frecuente que se hace del teorema de Bayes, este teorema se deriva de un trabajo realizado por el matemático Thomas Bayes, donde introduce el concepto de probabilidad inversa y además permitió entender cómo las personas cambian su juicio sobre la ocurrencia de algún evento debido a información adicional. Pero fue Pierre-Simon Laplace quien aplica el Teorema de Bayes de manera sistemática al análisis de datos, esto a finales del siglo XVIII [4].

**Teorema 1.1** (*Teorema de Bayes*)

Sea  $\{A_1, A_2, \dots, A_i, \dots, A_n\}$  un conjunto de sucesos mutuamente excluyentes y exhaustivos, y tales que la probabilidad de cada uno de ellos es distinta de cero. Sea  $B$  un suceso cualquiera del que se conocen las probabilidades condicionales  $P(B|A_i)$ . Entonces,

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B)},$$

donde

1.  $P(A_i)$  son las probabilidades a priori,
2.  $P(B|A_i)$  es la probabilidad de  $B$  en la hipótesis  $A_i$ ,
3.  $P(A_i|B)$  son las probabilidades a posteriori.

La inferencia Bayesiana usa un estimador numérico del grado de creencia en una hipótesis aún antes de observar la evidencia y calcula un estimador numérico del grado de creencia después de haber observado la evidencia.

Dada una nueva evidencia, el teorema de Bayes ajusta las probabilidades de la siguiente manera:

$$P(H_0|E) = \frac{P(E|H_0)P(H_0)}{P(E)},$$

donde

1.  $H_0$  representa una hipótesis, llamada hipótesis nula, que ha sido inferida antes de que la nueva evidencia,  $E$ , resultara disponible.
2.  $P(H_0)$  se llama probabilidad a priori de  $H_0$ .
3.  $P(E|H_0)$  se llama la probabilidad condicional de que se cumpla la evidencia  $E$  si la hipótesis  $H_0$  es verdadera. Se llama también la función de verosimilitud cuando se expresa como una función de  $E$  dado  $H_0$ .
4.  $P(E)$  se llama la probabilidad marginal de  $E$ .
5.  $P(H_0|E)$  se llama probabilidad a posteriori de  $H_0$  dado  $E$ .

## 2. Inferencia Bayesiana

En general, el objetivo de la estadística Bayesiana es representar la incertidumbre previa sobre los parámetros del modelo con una distribución de probabilidad y actualizar esta incertidumbre anterior con nuevos datos para producir una distribución de probabilidad posteriori para el parámetro que contiene incertidumbre; es decir, la estadística bayesiana se pregunta explícitamente cómo cambia nuestro estado de información acerca del valor del parámetro mediante los datos observados.

La diferencia con la inferencia clásica es que esta toma a los parámetros fijos y en la Bayesiana suponemos que los parámetros son variables aleatorias con una distribución de probabilidad.

Cuando los datos se encuentran en un espacio muestral discreto, el teorema de Bayes visto en términos de funciones de masa se encuentra caracterizado de la siguiente manera:

$$f(\theta|\text{datos}) = \frac{f(\text{datos}|\theta)f(\theta)}{f(\text{datos})},$$

donde  $\theta$  es el parámetro a estimar,  $f(\theta|\text{datos})$  es la función a posteriori para el parámetro,  $f(\text{datos}|\theta)$  se le llama función de verosimilitud,  $f(\theta)$  es la función a priori del parámetro y  $f(\text{datos})$  es la función marginal de los datos. Si los datos se encuentran en un espacio muestral continuo, la función de densidad marginal se puede encontrar de la siguiente forma:

$$f(\text{datos}) = \int f(\text{datos}|\theta)f(\theta)d\theta.$$

Se supone que se tiene una base de datos sobre el parámetro que se desea estimar, luego la densidad marginal de los datos es un número ya conocido, de esta forma puede ser tomado como una constante fija. Así se deduce que la distribución a posteriori es proporcional al producto de la función de verosimilitud por la función a priori como se observa en la siguiente ecuación:

$$f(\theta|\text{datos}) \propto f(\text{datos}|\theta)f(\theta). \tag{1}$$

donde el símbolo  $\propto$  significa “proporcional a”.

En resumen los pasos para realizar una estimación bayesiana son los siguientes:

- Establecer un modelo probabilístico completo: una distribución de probabilidad conjunta para todas las cantidades del problema, observables y no observables.
  - Función de verosimilitud:  $f(\text{datos}|\theta)$ .
  - Distribución a priori:  $f(\theta)$ .
- Condicionar los datos: obtener la distribución a posteriori, es decir, la distribución condicionada a los parámetros del modelo, dados los datos.
  - Teorema de Bayes:  $f(\theta|\text{datos}) \propto f(\text{datos}|\theta)f(\theta)$ .

Para construir la distribución a priori existen distintas formas de hacerlo, a continuación algunas de ellas.

- Distribución a priori informativa. Ver [8].
  1. Estudios empíricos previos.
  2. Conocimiento del investigador:
    - Por intervalos.
    - Estimación de momentos y supuesto de simetría.
    - Reparametrización de distribuciones. Ej:  $beta(c \cdot \tau, (1 - m) \cdot \tau)$ .
- Distribución a priori no-informativa. Ver [7].

- Impropias:  $U(-\infty, \infty)$  o  $U(0, \infty)$ .
- Distribución poco informativas:  $\theta$  tenga una distribución  $N(\mu, 10000)$ .

En este trabajo se utiliza una distribución a priori informativa con estudios empíricos previos.

### 3. Distribución beta a priori

En esta sección se expone el caso de una distribución a priori beta con función de verosimilitud Bernoulli. La función de densidad de una distribución Beta esta dada por:

$$f(p) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1} (1-p)^{\beta-1},$$

donde  $0 \leq p \leq 1$  y  $\alpha, \beta$  positivas, con esperanza  $\frac{\alpha}{\alpha+\beta}$ .

Sea una muestra aleatoria  $X = (X_1, \dots, X_n)$  que tienen una distribución *Bernoulli*( $p$ ). Así la función de verosimilitud es la siguiente:

$$f(X|p) = p^x (1-p)^{1-x}, \quad x \in [0, 1],$$

entonces la distribución a posteriori según la ecuación (1),

$$\begin{aligned} f(p|X) &\propto p^{\sum_{k=1}^n x_k} (1-p)^{n-\sum_{k=1}^n x_k} \cdot p^{\alpha-1} (1-p)^{\beta-1} \\ &= p^{\sum_{k=1}^n x_k + \alpha - 1} (1-p)^{n - \sum_{k=1}^n x_k + \beta - 1}. \end{aligned}$$

Por lo tanto  $f(p|X) \sim \text{beta} \left( \sum_{k=1}^n x_k + \alpha, n - \sum_{k=1}^n x_k + \beta \right)$ . En conclusión la distribución a posteriori es de la misma familia paramétrica que la a priori. Cuando lo anterior ocurre decimos que las distribuciones Bernoulli y beta son **conjugadas**.

Veamos que sucede con nuestras funciones a priori y posteriori para distintos valores de  $\alpha$  y  $\beta$ .



Figura 1: Gráfica de  $f(p), f(X|p), f(p|X)$ , donde  $f(p)$  tiene parámetros  $\alpha = 5$  y  $\beta = 2$ .

### 3.1. Ejemplo 1

Pensemos en la población que consiste de todas las personas en México que pueden votar en la próxima elección presidencial del 2018, sea  $p$  que representa la proporción de esa población que va ejercer su voto. La creencia de una persona acerca de la incertidumbre en esta proporción se representa por una distribución de probabilidad sobre el parámetro. Esta distribución refleja la opinión subjetiva previa de la persona sobre los valores plausibles de  $p$ .

Según información que publicó el Instituto Nacional Electoral (INE), solo el 65.44% de las personas que podían votar en el año 2012 asistieron a las urnas. Basándose en esta información se cree que la proporción  $p$  debe tener un valor mayor que 0.5, o siendo más específicos decimos que el valor de  $p$  pertenece al intervalo de 0.5 a 1.

Para obtener la verosimilitud, se considera que si una persona vota el resultado será  $X = 1$  y  $X = 0$  de lo contrario. De esta manera podemos considerar a  $X$  del tipo Bernoulli. Si consideramos una sola observación tenemos que la función de verosimilitud es la función de densidad de una distribución Bernoulli con parámetro  $p$ .

Nuestra distribución a priori para  $p$  esta dada de acuerdo a la experiencia que se ha visto en la última elección presidencial del 2012 ya que 50,323,153 votaron y 2,913,649 no lo hicieron de acuerdo a datos publicados por el INE (poner referencia). Si  $\alpha = 50,323,153$  y  $\beta = 2,913,649$  de acuerdo a () la función a posteriori es:

$$f(p|\text{datos}) \propto p^{50,323,154-1}(1-p)^{2,913,648-1}.$$

Por lo tanto la función a posteriori para la proporción  $p$  es del tipo Beta con parámetros  $\beta_0 = 50,323,154$  y  $\alpha_0 = 2,913,648$ .

### 3.2. Ejemplo 2

De acuerdo a una encuesta nacional en vivienda de El Financiero (Ver [9]) que se realizó del 19 al 25 de enero del 2017 en todas las entidades federativas, a 1008 personas, se obtuvieron los resultados siguientes.

Candidato	Porcentaje				
	Mar 16	Jun	Sep	Nov	Feb 17
A. M. L. O.	28	31	29	29	33
Margarita Z.	24	26	28	29	27
M. A. O. C.	24	26	27	26	20
M. Ángel M.	14	10	10	9	10
El Bronco	9	7	7	7	10

Según el periódico La Jornada (Ver [2]), Jueves 23 de Marzo de 2017 p.6, Andrés Manuel López Obrador se coloca a la cabeza en todos los escenarios. Se realizó una encuesta, con fecha de 3 de marzo del 2017, a 5275 personas de todo el país. Los resultados son los siguientes:

Marzo 2017	
Candidato	Porcentaje
Andrés Manuel López Obrador	32.694
Margarita Zavala	27.974
Miguel Ángel Osorio Chong	19.214
Miguel Ángel Mancera	12.264
Jaime Rodríguez, El Bronco	7.854

Estamos interesados en estimar la proporciones  $\theta_1, \theta_2, \theta_3, \theta_4$  y  $\theta_5$  de que gane López Obrador, Margarita Zavala, Osorio Chong, Mancera y el bronco respectivamente, de acuerdo a las encuestas presentadas anteriormente. Considerando que las proporciones no son fijas y se comportan de manera aleatoria, esto debido a distitos factores como pueden ser: qué partido ocupa la presidencia actualmente, campañas políticas, puestos que han ocupado los candidatos durante su carrera política, entre otras.

Se utiliza el modelo propuesto al inicio de la sección, donde  $\Theta = (\theta_1, \theta_2, \dots, \theta_k)$  tiene distribución a priori Dirichlet y la muestra aleatoria  $X = (x_1, \dots, x_n)$  tiene una distribución multinomial.

En nuestra distribución a priori tenemos  $n = 1008, k = 5, \alpha = (\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5)$ , donde  $n$  es el total de adultos encuestados,  $k$  el número de candidatos a la presidencia y:

- $\alpha_1$  es el número de votos por Andrés Manuel,
- $\alpha_2$  es el número de votos por Margarita Zavala,
- $\alpha_3$  es el número de votos por M. Ángel Osorio,
- $\alpha_4$  es el número de votos por M. Ángel Mancera,
- $\alpha_5$  es el número de votos por Jaime Rodríguez.

## 4. Conclusiones

En este trabajo mostramos el enfoque bayesiano donde a diferencia de la estadística clásica el parámetro a estimar se toma variable, más aún, se dice que tiene alguna distribución de probabilidad. Mediante el uso de este enfoque vemos la necesidad de utilizar distribuciones las cuales se le llaman posteriori, a priori y versomilitud. La primera resulta importante ya que se puede realizar predecciones para muestras aleatorias posteriores. Aunque en el enfoque bayesiano se puede pensar que si tenemos dos opiniones iniciales diferentes posiblemente se obtendrán conclusiones que difieren, sin embargo, en un sentido estricto esto puede ser cierto, mediante la acumulación de datos estas conclusiones deben coincidir mediante la distribución predictiva.

Se trabajaron los casos de distribución beta-Bernoulli conjugadas y Dirichlet-multinomial conjugadas, ejemplificando con las votaciones que serán realizadas en

el 2018. En el primer ejemplo se estimó el número de votantes lo cual no resulto fácil ya que por el momento no se encuentran encuestas en el país sobre si la gente votará o no, pero se muestra un método con el cual se puede hacerlo una vez que se tienen las encuestas pertinentes. En el segundo ejemplo se analiza quién tiene mayor probabilidad de ganar en el 2018 de acuerdo a encuestas publicadas por periódicos de México, esto mediante la distribución a posteriori que se encuentra con una verosimilitud multinomial y una distribución a priori Dirichlet.

En base a esto, utilizando la distribución predictiva se puede decir algo sobre el posible ganador de las elecciones 2018 como también la proporción de personas que ejercieran su voto. Otro punto importante, es que se usó información adicional o previa que sirvió para nuestros parámetros, se asignaron probabilidades subjetivas y se consideró el valor del parámetro aleatorio.

## Referencias

- [1] BERNARDO M. *Bayesian Statistics*. Departamento de Estadística de la Facultad de Matemáticas, Valencia Spain, 2002.
- [2] CANO A. *Humillante derrota del PRI en 2018: encuesta de Presidencia*. La Jornada. Jueves 23 de marzo de 2017, p.6. Recuperado desde: <http://www.jornada.unam.mx/2017/03/23/politica/006n1pol>
- [3] EULACIO N.R. *La familia de distribuciones de Pólya truncada*. Tesis de Maestría. Instituto de Enseñanza e Investigaciones de Ciencias Agrícolas. México: Chapingo.
- [4] GUTIÉRREZ E.. *El desarrollo de la estadística bayesiana*. Revista Digital Universitaria, Vol. 14 No. 11, 2013. Recuperado desde: <http://www.revista.unam.mx/vol.14/num11/art42/#>
- [5] JIM ALBERT. *Bayesian Computation with R*. Springer Second Edition, USA, 2001.
- [6] LAWRENCE J., THERESA W. y LOIS C. *Bayesian Estimatin of Disease Prevalence and the Parameters of Diagnostic Test in the Absence of a Gold Standard*. American Journal of Epidemiology, Vol. 141 No.3, 1995.
- [7] MENDOZA R. MANUEL y REGUEIRO M. PEDRO *Estadística Bayesiana*. Departamento de Estadística, Instituto Tecnológico Autónomo de México, 2011.
- [8] MICHAEL H. y LEONARD H. *Bayesian Estimation of the Size of a Population*. Department of Statistics Biostatistics Unit, University Zurizh and University of Munich, Sonderforschungsbereich 386, Paper 499, 2006.
- [9] MORENO A. *Toma AMLO ventaja*. El Financiero. 2017. Recuperado desde: <http://graficos.elfinanciero.com.mx/2017/encuestas/enc-01febrero17/index.html>
- [10] SCOTT M. LYNCH. *Introduction to Applied Bayesian Statistics and Estimation for Social Scientists*. Springer, New York, 1995.