Estimación Bayesiana de los Resultados de las Encuestas Electorales

Lic. Dindo Valdez Blanco
dindovaldez@hotmail.com

Resumen. En el presente artículo se revisan los elementos básicos del paradigma bayesiano y se realiza una estimación bayesiana puntual y de intervalo de la proporción de votación a favor del actual presidente Evo Morales basado en los resultados de la encuesta realizada por la empresa Ipsos en el mes de agosto de 2014.

Palabras clave: Inferencia Bayesiana, encuestas electorales, función a priori, función a posteriori.

1. Introducción

El presidente del Estado Evo Morales, domina con amplia ventaja en las encuestas seis semanas antes de las elecciones generales del 12 de octubre. En la encuesta realizada por la encuestadora privada Ipsos, Morales se encuentra a la cabeza de intención electoral con un caudal del 59 %, y 42 % de voto seguro. La ficha técnica de dicho sondeo indica que se entrevistó a 3000 ciudadanos entre el 1 y 16 de agosto, la muestra tiene una confianza de 95 % y un error máximo de estimación de 1.79 %, la encuesta fue efectuada en 10 ciudades capitales, 10 localidades urbanas y 88 localidades rurales de Bolivia.

Estos resultados son estimaciones del tipo frecuentista denominado estimación clásica, sin embargo es posible obtener una estimación bayesiana y calcular la precisión que tienen ambos estimadores.

2. El paradigma bayesiano

Si definimos el parámetro π como la proporción de electores de la población que tienen la intención de votar por Evo Morales, definimos también una variable binaria X, tal que $x = 0$ si una persona no piensa votar por el actual presidente, y $x = 1$ si tiene la intención de votar por el candidato oficialista. Entonces la función de probabilidad de X condicionada en el parámetro π es Bernoulli, tal que su función de probabilidad paramétrica es:

$$f(x|\pi) = \pi^x (1-\pi)^{1-x}; \qquad x = 0{,}1., \ 0 \leq \pi \leq 1. \tag{1}$$

El estimador máximo verosímil de π basado en una muestra aleatoria de tamaño n de X está definido por: $\widehat{\pi}_F = \sum_{i=1}^{n} x_i/n$, el mismo basado en los datos de la encuesta electoral resulta igual a $\widehat{\pi}_F = 0.59$. En este caso se considera a π como un parámetro desconocido pero fijo.

Sin embargo desde el punto de vista bayesiano, el parámetro π es una variable aleatoria. Este aspecto tiene sentido porque la intención de voto de los electores es cambiante, debido a muchos factores como la información que emiten los medios de comunicación y las propuestas de los candidatos. La estimación bayesiana se basa en la función de probabilidad condicional denominada función de probabilidad a posteriori definida por:

$$f(\pi|\text{muestra}) = \frac{f(\pi) \times f(\text{muestra}|\pi)}{f(\text{muestra})} \tag{2}$$

3. Estimación bayesiana tomando una función a priori Beta

Si se supone una función a priori $Beta(\alpha, \beta)$ para π con α y β conocidos:

$$f(\pi) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \pi^{\alpha-1}(1-\pi)^{\beta-1}; \qquad 0 \leq \pi \leq 1 \tag{3}$$

La función de probabilidad a posteriori de π dada la información de la muestra resulta:

$$f(\pi|\text{muestra}) = \frac{\Gamma(\alpha+\beta+n)}{\Gamma(\alpha+\sum x_i)\Gamma(\beta+n-\sum x_i)}\pi^{\alpha+\sum x_i-1}(1-\pi)^{\beta+n-\sum x_i-1}; \qquad 0 \leq \pi \leq 1. \tag{4}$$

Claramente se reconoce que la distribución posterior corresponde a una ley $Beta(\alpha' = \alpha + \sum x_i, \beta' = \beta + n - \sum x_i)$. Basado en este hecho el estimador de Bayes se define como la media condicional de la función posterior:

$$\widehat{\pi}_B = E(\pi|\text{muestra}) = \frac{\alpha+\sum x_i}{\alpha+\beta+n}. \tag{5}$$

Para la elección de los valores iniciales α y β, se puede utilizar un sistema de ecuaciones basados en la media y varianza de la distribución a priori de π, si se considera una media igual a 0.5 y una varianza de 0.01, el sistema resulta:

$$E(\pi) = \frac{\alpha}{\alpha+\beta} = 0.5 \qquad V(\pi) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} = 0.01 \tag{6}$$

De donde se obtienen los valores iniciales $\alpha = 12$, $\beta = 12$. Tomando la información de la muestra $n = 3000$, $\sum x_i = 1770$, el estimador bayesiano resulta $\widehat{\pi}_B = 0.5893$. De la misma manera se puede calcular un intervalo de confianza al 95 % para la estimación bayesiana basado en los percentiles 2.5 y 97.5 de la distribución $Beta(\alpha' = 1782, \beta' = 1242)$. De donde resulta el intervalo (0.5717; 0.6068).

4. Estimación bayesiana tomando una función a priori Uniforme

Si no se tiene ninguna suposición sobre la distribución del parámetro π se puede asumir una función de probabilidad uniforme $U(0,1)$, es decir: $f(\pi) = 1$, $0 \leq \pi \leq 1$. De igual manera se obtiene la función de densidad posterior de π dada la información de la muestra:

$$f(\pi|\text{muestra}) = \frac{\Gamma(n+2)}{\Gamma(\sum x_i + 1)\Gamma(n - \sum x_i + 1)}\pi^{\sum x_i}(1-\pi)^{n-\sum x_i}; \qquad 0 \leq \pi \leq 1. \tag{7}$$

La expresión anterior muestra que la distribución a posteriori también sigue una ley $Beta(\alpha' = \sum x_i + 1, \beta' = n - \sum x_i + 1)$. Nuevamente reemplazando las observaciones de la muestra, se tiene que la distribución a posteriori es $\pi|\text{muestra} \sim Beta(\alpha' = 1770, \beta' = 1231)$. Por tal razón el estimador puntual bayesiano de la proporción de la población que piensa apoyar al presidente en las próximas elecciones es igual a la media condicional de la densidad posterior.

$$\widehat{\pi}_B = E(\pi|\text{muestra}) = \frac{\alpha'}{\alpha'+\beta'} = 0.5898 \tag{8}$$

Finalmente el intervalo de confianza al 95 % para el parámetro π desde el enfoque bayesiano es (0.5626; 0.5976).

5. Medida de precisión de los estimadores

El estimador frecuentista clásico es insesgado, en cambio los estimadores bayesianos por lo general tienen un sesgo, la diferencia radica en que el Error Cuadrático Medio de las estimaciones es menor en el caso Bayesiano en comparación del estimador frecuentista. El error cuadrático medio (ECM) de un estimador $\widehat{\theta}$ está definido como:

$$ECM(\widehat{\theta}) = E[(\widehat{\theta}-\theta)^2] = (E[\widehat{\theta}]-\theta)^2 + Var[\widehat{\theta}] \tag{9}$$

En el caso del estimador frecuentista $\widehat{\pi}_F$,su media y varianza son: $E[\widehat{\pi}_F] = \pi$, $V[\widehat{\pi}_F] = \pi(1-\pi)/n$. Por lo que su ECM resulta:

$$ECM(\widehat{\pi}_F) = \frac{\pi(1-\pi)}{n} \tag{10}$$

Con $n = 3000$ y $\pi = 0.5$, se tiene un error cuadrático medio igual a $ECM(\widehat{\pi}_F) = 0.000083$.

De igual manera para el caso bayesiano la media y varianza de $\widehat{\pi}_B$ es igual a:

$$E(\widehat{\pi}_B) = \frac{\alpha + n\pi}{n+\alpha+\beta} \qquad V(\widehat{\pi}_B) = \frac{n\pi(1-\pi)}{(n+\alpha+\beta)^2} \tag{11}$$

De tal forma que su error cuadrático medio es igual a:

$$ECM(\widehat{\pi}_B) = \left(\frac{\alpha-(\alpha+\beta)\pi}{n+\alpha+\beta}\right)^2 + \frac{n\pi(1-\pi)}{(n+\alpha+\beta)^2} \tag{12}$$

Con $n = 3000$, $\pi = 0.5$, $\alpha = 12$ y $\beta = 12$, se obtiene $ECM(\widehat{\pi}_B) = 0.000082$. Esto confirma que el estimador bayesiano tiene menor error en comparación con el estimador frecuentista.

La siguiente tabla muestra las estimaciones del parámetro π con sus respectivos intervalos de confianza.

Tabla 1 Estimaciones del parámetro π

Método	Función a priori	Estimación puntual	Intervalo de confianza	Error cuadrático medio
Frecuentista	—	0.5900	(0.5724; 0.6076)	0.000083
Bayesiano	$Beta$(12,12)	0.5893	(0.5717; 0.6068)	0.000082
Bayesiano	$Uniforme$(0,1)	0.5898	(0.5626; 0.5976)	0.000083

6. Conclusión

Mediante esta aplicación se concluye que los estimadores bayesianos tienen igual o mejor precisión que el estimador frecuentista, adicionalmente la estimación bayesiana es muy parecida en los dos casos, no importando la función a priori inicial que se tome.

Referencias

[1] Bernardo, J. M., Bayarri, M. J., Berger, *Bayesian Statistics*. Oxford: University Press, 2003.

[2] Lee, P. M. *Bayesian Statistics: An Introduction* (3rd. ed.) London University Press, 2004.