

Seguimiento de la opinión pública usando MCMC: Elecciones presidenciales Chile 2021

Alonso Rojas Tiare Letelier

Resumen

El presente proyecto propone dos modelos de seguimiento de la opinión pública simplificados, aplicado en particular a las elecciones presidenciales en Chile el año 2021, tanto para segunda y primera vuelta, aunque en esta última solo se consideraron los candidatos más populares de las encuestas, es decir: Jose Antonio Kast, Sebastian Sichel, Gabriel Boric y Yasna Provoste. Para ello se hará uso de modelos bayesianos dinámicos, junto a métodos de Markov Chain Monte Carlo, logrando obtener un seguimiento diario de las intenciones de voto de los 100 días previos a las elecciones.

1 Formulación

Se utilizó un modelo lineal Dirichlet-Multinomial para representar el comportamiento de las intenciones de voto, ya que es considerado el más adecuado para trabajar en sistemas democráticos con varios partidos.

Primeramente se define para cada tiempo t un vector y_t que representa la cantidad de encuestados que votan por cada candidato, es decir, si son c candidatos, se tiene

$$y_t = (y_{t1}, y_{t2}, \dots, y_{tc})$$

y proponemos $y_t \sim \text{Multinomial}(\alpha_{t1}, \alpha_{t2}, \dots, \alpha_{tc}, N_t)$, donde N_t es la cantidad de encuestados en el tiempo t , y cada α_{ti} representa la probabilidad de que en dicho tiempo una persona tenga la intención de votar por el candidato i .

Queremos realizar inferencias sobre estos últimos parámetros, y para ello definimos un modelo bayesiano y consideramos para cada vector α_t una distribución a priori de tipo Dirichlet de parámetros b_{t1}, \dots, b_{tc} , lo que garantiza que los α_{ti} sumen 1. Como resultado de esto la distribución a posteriori también es una Dirichlet, pero de parámetros $y_{t1} + b_{t1}, \dots, y_{tc} + b_{tc}$. Se quiere utilizar las proporciones obtenidas en el tiempo $t - 1$ como prior para el tiempo t , y para ello se escogió los valores de los b_{ti} como $b_{ti} = 500 \times \alpha_{t-1,i}$, con lo que estamos suponiendo que en cada día se entrevistan al menos 500 personas. Para el primer día se utiliza un prior uniforme, en el sentido que todos los candidatos tienen la misma proporción de votos.

Para obtener una muestra de los α_t utilizamos el algoritmo de Metropolis-Hastings, de manera que en cada iteración se actualizan los valores de α_t para todos los tiempos. Para obtener un valor de prueba de α_t se implementaron dos modelos: el primero, obtiene una muestra de una distribución $\text{Gamma}(b_{ti} + y_{ti}, 1)$ que luego se normaliza para obtener proporciones que sumen 1. El segundo simplemente obtiene un valor aleatorio en un intervalo $(\max\{\alpha_{ti} - d, 0\}, \alpha_{ti} + d)$ con probabilidad uniforme. Con estas proporciones de prueba se aplica en cada día el criterio de Metropolis.

2 Resultados

Con el fin de comparar la precisión de los modelos se realizaron diferentes variaciones en parámetros de interés, por ejemplo, cambios en el parámetro d utilizado en la distribución uniforme del

segundo modelo, observando los porcentajes de aceptación de los valores de prueba para cada día.

A partir de la muestra generada graficamos su media para cada día, en conjunto con intervalos de confianza del 90 %. A continuación, presentamos uno de los gráficos obtenido a partir del primer modelo.

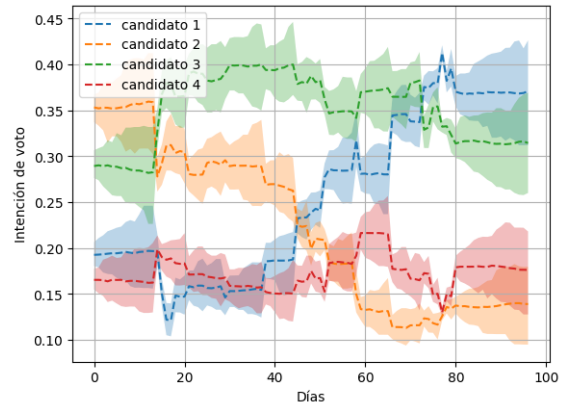


Figura 1: Seguimiento intención de voto Modelo n°1. Los candidatos son: 1. Kast, 2. Sichel, 3. Boric, 4. Provoste.

Finalmente se extraen los porcentajes de intención de voto obtenidos para el día de las elecciones presidenciales:

Kast: 37.094199 %

Sichel: 13.872800 %

Boric: 31.434299 %

Provoste: 17.598701 %

3 Referencias

[1] Bunker, Kenneth; Bauchowitz, Stefan(2016).Electoral Forecasting and Public Opinion Tracking in Latin America: An Application to Chile Política.

[2] Stoltenberg, Emil (2013). Bayesian Forecasting of Election Results in Multiparty Systems. Master's thesis University of Oslo.

[3] Tres Quintos. <http://www.tresquintos.com/>