



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

Integrantes: Tiare Letelier , Alonso Rojas
22 Diciembre 2022



Indice

1.- Introducción

2.- Marco Teórico:

- Modelo Dinámico Lineal
- Modelo Dirichlet-Multinomial
- Métodos Markov Chain Monte Carlo

3.- Data

4.- Implementación

5.- Resultados

6.- Análisis y conclusiones

Resumen

A través de 2 modelos se realizará un seguimiento de la intención de voto aplicado al caso de las elecciones presidenciales del año 2021, a través de las herramientas mencionadas.

- Pregunta a responder:

¿Por quién votaría usted si las elecciones se realizaran este domingo?

Referencia: Bunker, K; Bauchowitz, S(2016).Electoral Forecasting and Public Opinion Tracking in Latin America: An Application to Chile Política.



Marco Teórico

Dynamic Linear Model

Los modelos lineales dinámicos (DLM) son una clase general de modelos Bayesianos temporales no estacionarios, muy adecuados para modelar la dinámica de una cantidad latente para la que sólo existen medidas imprecisas, como por ejemplo el apoyo a candidatos presidenciales.

Componentes DLM:

- 1) Una **ecuación de observación** que describe cómo la serie temporal observada y_t es generada por una serie temporal no observada α_t de estados latentes.
- 2) Una **ecuación de transición** que describe cómo evoluciona este estado latente con el tiempo.
- 3) La información inicial (**prior**)



Dirichlet-Multinomial

Adecuado para trabajar en
sistemas electorales
multipartitos

Primeramente definimos el siguiente vector:

$$y_t = (y_{t1}, y_{t2}, y_{t3}, y_{t4})$$

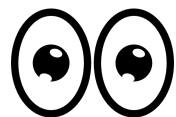
el cual agrupa la cantidad de encuestados que votan por cada candidato en el tiempo t . Con ello obtenemos la serie de tiempo observada $y_1, y_2, y_{t3}, \dots, y_N$. Luego el modelo observacional es de la forma:

$$y_t \sim \text{Multi}(\alpha_{t1}, \alpha_{t2}, \alpha_{t3}, \alpha_{t4}; N_t)$$

Donde α_{ti} ($i = 1, \dots, 4$) es la probabilidad de que un encuestado declare su intención de voto al candidato i . Y N_t es el tamaño del sondeo

(Stoltenberg, 2013)

Luego debemos especificar el prior sobre los parámetros α_{ti} . La distribución prior conjugada de una distribución Multinomial se conoce como distribución de Dirichlet



Un prior Dirichlet sobre $(\alpha_{t1}, \alpha_{t2}, \alpha_{t3}, \alpha_{t4})$ asegura que los parámetros α_{ti} sumen 1

La densidad de probabilidad de este tipo de distribuciones es de la forma

$$\pi(\alpha_{t1}, \alpha_{t2}, \alpha_{t3}, \alpha_{t4}) = \frac{\Gamma(b_{t1} + b_{t2} + b_{t3} + b_{t4})}{\Gamma(b_{t1}) \dots \Gamma(b_{t4})} \alpha^{b_{t1}-1}_{t1} \dots \alpha^{b_{t4}-1}_{t4}$$

$$b_{ti} \in \mathbb{Z}^+$$

Finalmente por Regla de Bayes, la distribución posterior para α esta dada por:

$$\pi(\alpha_{t1}, \alpha_{t2}, \alpha_{t3}, \alpha_{t4} | y_{t1}, y_{t2}, y_{t3}, y_{t4}) \sim \text{Diri}(b_{t1} + y_{t1}, \dots, b_{t4} + y_{t4})$$



Los parámetros b_{ti} son enteros positivos, mientras que $\alpha_{ti} \in [0,1]$



Recuerdo Metropolis-Hastings

Simular $(X_n) \sim CM(\mu, P)$

- $(V_n) \sim iid. U([0,1])$ y $f: [0,1] \times E \rightarrow E$ función de transición
- $(U_n) \sim iid. U([0,1])$ indep de las (V_n)
- Dado $X_n = x$ simulamos

$$Y_{n+1} = f(V_{n+1}, x) = y$$

$$X_{n+1} = \begin{cases} Y_{n+1} & \text{si } U_{n+1} \leq \frac{\pi_y R_{yx}}{\pi_x R_{xy}} \\ X_n & \text{si } U_{n+1} > \frac{\pi_y R_{yx}}{\pi_x R_{xy}} \end{cases}$$

Data: Primera vuelta

Se trabajó únicamente con los candidatos con mayor popularidad en las encuestas:

- 1) Candidato 1: José Antonio Kast
- 2) Candidato 2: Sebastián Sichel
- 3) Candidato 3: Gabriel Boric
- 4) Candidato 4: Yasna Provoste

Análisis de los últimos **100 días** de campaña electoral y de manera **diaria**

Fecha	Kast	Sichel	Boric	Provoste
13-08-2021	302	668	523	242
...
27-08-2021	3230	357	463	612
31-08-2021	1551	124	295	388
01-09-2021	1443	43	159	188
02-09-2021	0	0	0	0
03-09-2021	702	84	133	161
04-09-2021	0	0	0	0
05-09-2021	0	0	0	0
06-09-2021	1619	178	291	437
...
21-11-2021	0	0	0	0

Data: Segunda vuelta

Los candidatos en esta ocasión son únicamente:

- 1) Candidato 1: José Antonio Kast
- 2) Candidato 2: Gabriel Boric

Análisis de los últimos **94 días** de campaña electoral y de manera **diaria**

Fecha	Kast	Boric
16-09-2021	191	355
...
29-10-2021	1411	1537
30-10-2021	0	0
31-10-2021	794	699
01-11-2021	0	0
02-11-2021	4663	3957
03-11-2021	2788	2836
04-11-2021	444	404
...
19-12-2021	0	0

Implementación

A GRANDES RASGOS:

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:

$$b_{ti} = 500 * \alpha_{t-1,i}$$

1. Para el primer día se utiliza un prior uniforme
2. Para obtener una muestra de los α_t utilizamos el algoritmo de Metrópolis-Hastings, de manera que en cada iteración se actualizan los valores de α_t para todos los tiempos.
3. Para obtener un valor de prueba de α_t se implementaron dos modelos

```
post = lambda p, alpha: dirichlet.logpdf(p, alpha)
prior = [0.25, 0.25, 0.25, 0.25]
```

Modelo 1:

```
prueba = lambda alpha: np.array([gamma.rvs(a) for a in alpha])
```

Modelo 2:

```
prueba = lambda alpha, d: np.random.randint(np.maximum(alpha-d,0), alpha+d, size=4)
```

```
p_prueba = prueba(alpha)    o    prueba(alpha, d)
p_prueba = p_prueba/p_prueba.sum()
if post(p_prueba, alpha) - post(p, alpha) > np.log(np.random.uniform()):
    muestra[i, day] = p_prueba
else:
    muestra[i, day] = p
```



Intervalos de credibilidad 90%

```
q = [0.05, 0.95]  
q1, q2 = np.quantile(muestra, q, axis=0)
```

Graficamos:

```
for i in range(4):  
    plt.fill_between(range(days), q1[:,i], q2[:,i], alpha=0.3)
```

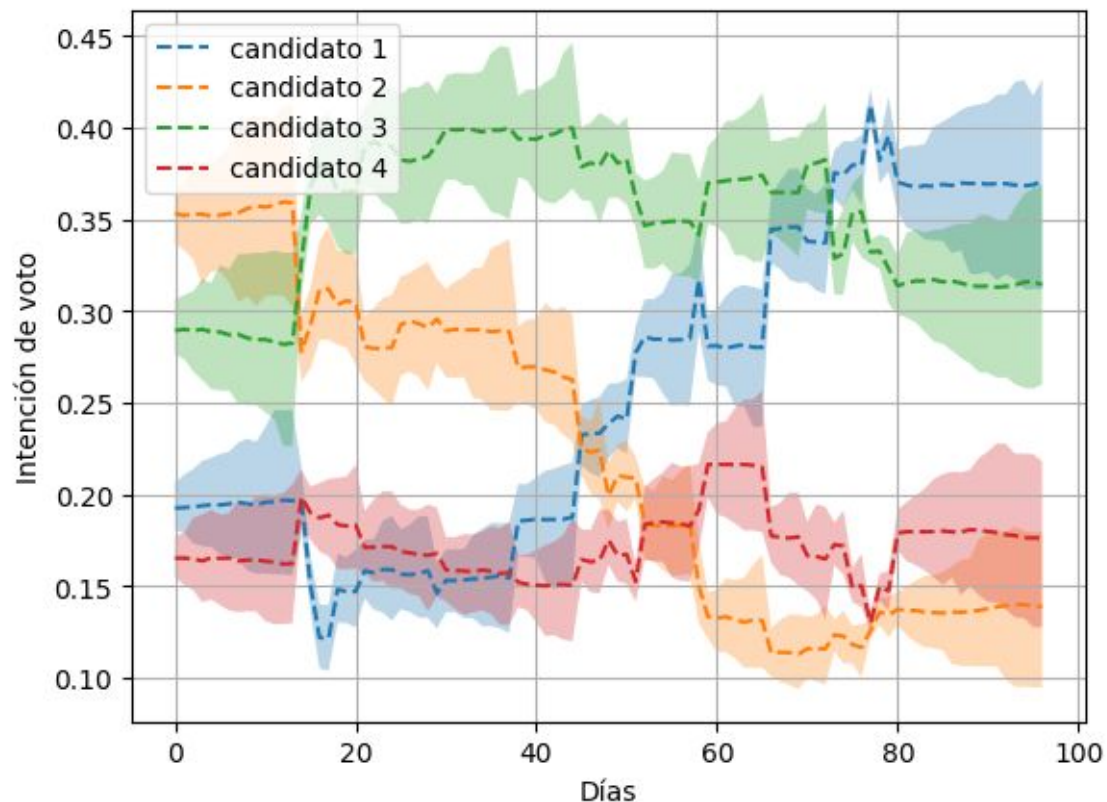


Resultados

1. Primera vuelta

Modelo 1

N = 10.000



Candidato	Porcentaje
Kast	37.094199%
Sichel	13.872800%
Boric	31.434299%
Provoste	17.598701%

Niveles de Aceptación:

Aceptados día 1: 100.0%

Aceptados día 2: 35.5%

Aceptados día 3: 35.46%

..

Aceptados día 15: 99.94%

Aceptados día 16: 100.0%

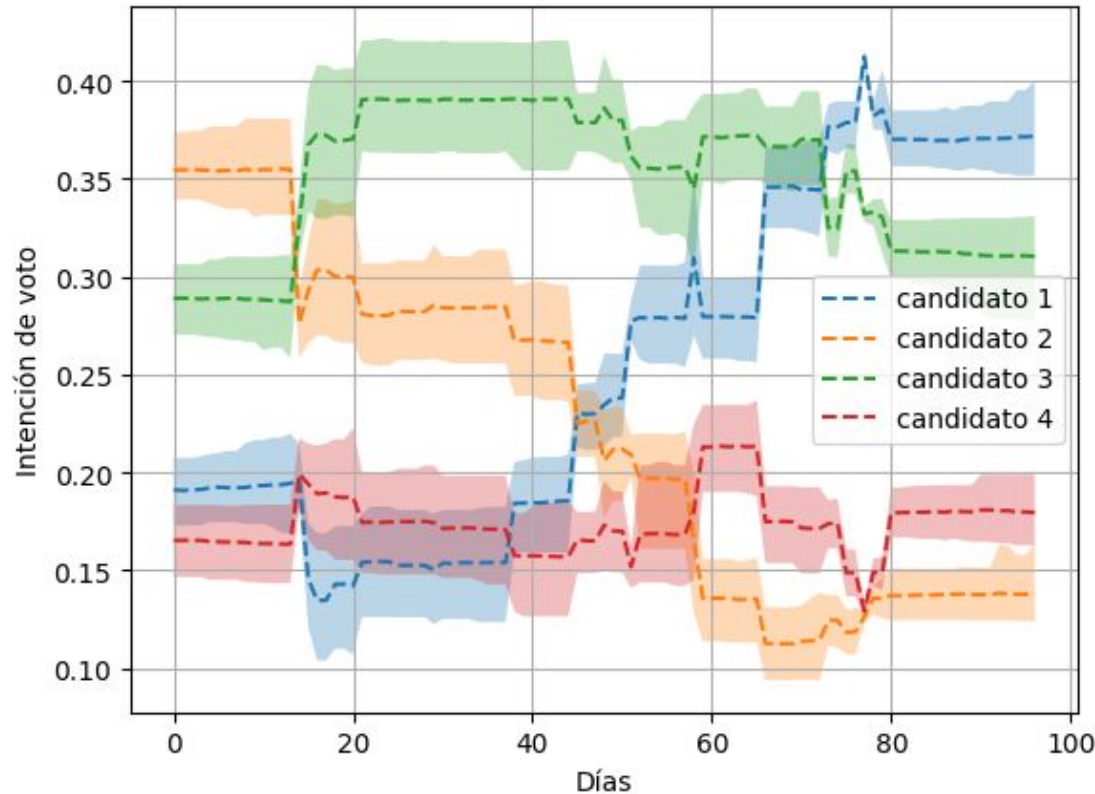
...

Aceptados día 96: 35.75%

Aceptados día 97: 35.91%

Modelo 2

$N = 10.000$, $d = 50$



Candidato	Porcentaje
Kast	37.180306%
Sichel	13.823127%
Boric	31.049486%
Provoste	17.947081%

Niveles de Aceptación:

Aceptados día 1: 100.0%

Aceptados día 2: 2.75%

Aceptados día 3: 2.62%

...

Aceptados día 18: 2.14%

Aceptados día 19: 31.83%

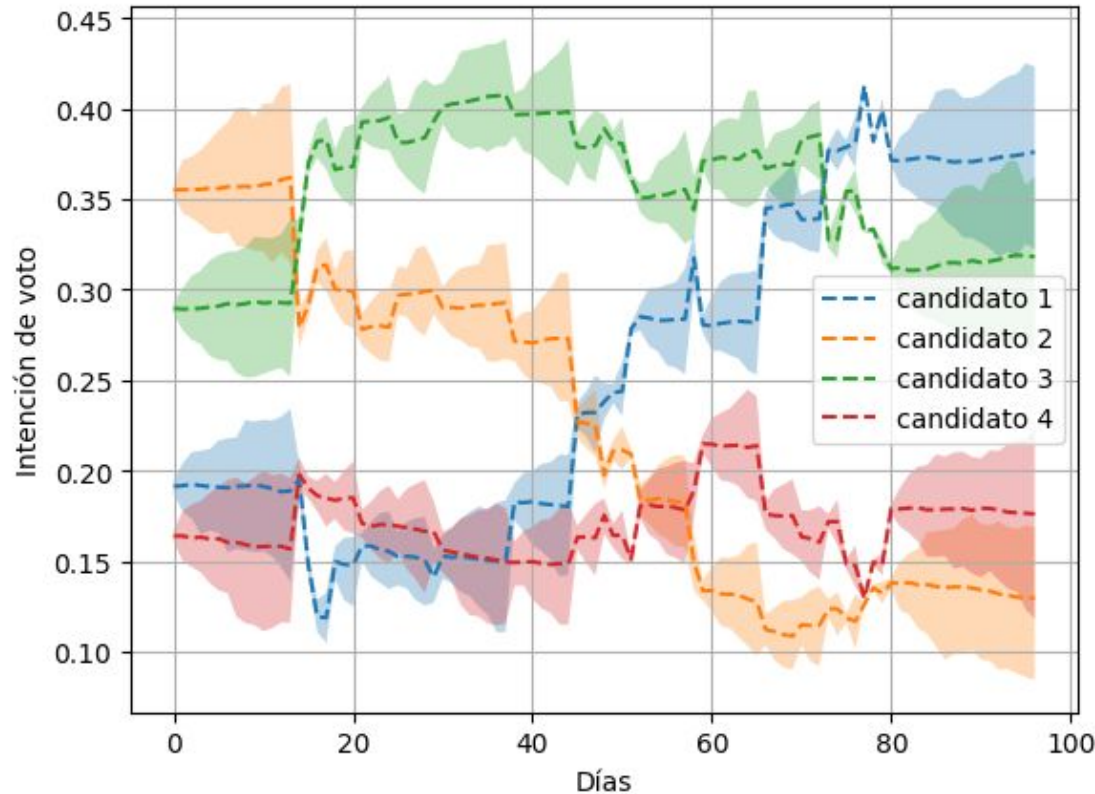
...

Aceptados día 96: 2.09%

Aceptados día 97: 2.34%

Modelo 2

$N = 10.000$, $d = 10$



Candidato	Porcentaje
Kast	37.610321%
Sichel	12.962773%
Boric	31.818190%
Provoste	17.608716%

Niveles de Aceptación:

Aceptados día 1: 100.0%

Aceptados día 2: 65.68%

Aceptados día 3: 65.39%

...

Aceptados día 18: 62.2%

Aceptados día 19: 100.0%

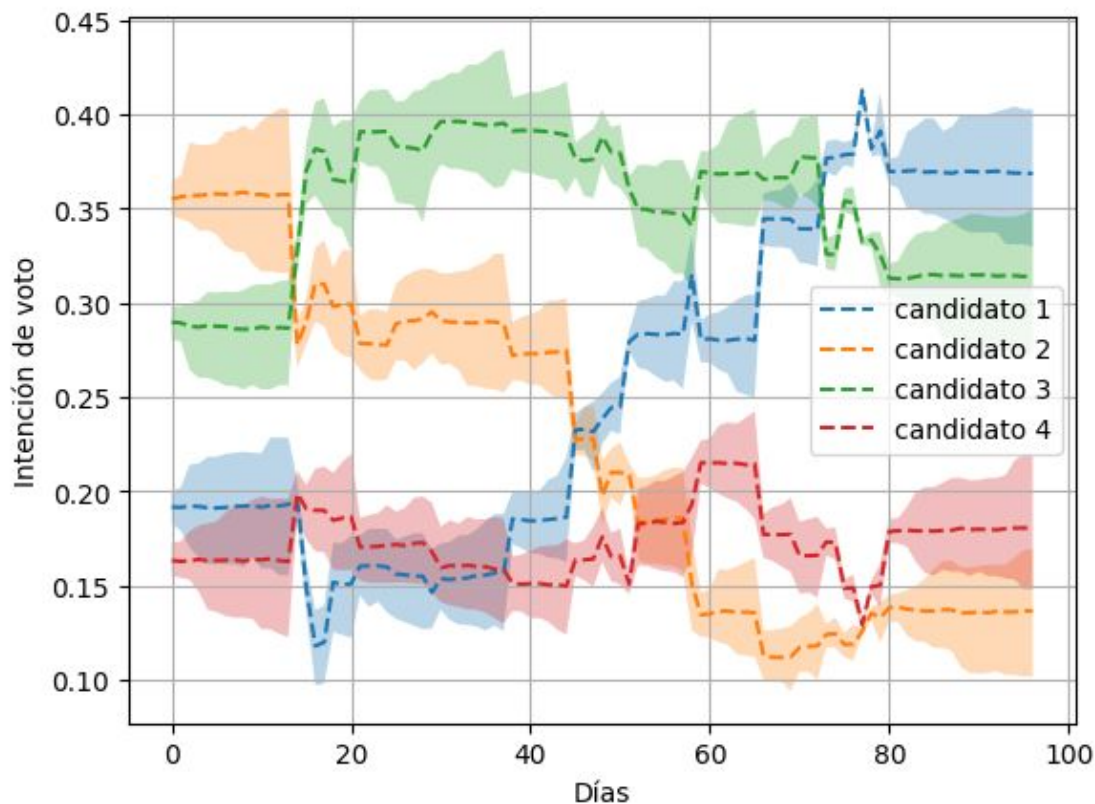
...

Aceptados día 96: 61.34%

Aceptados día 97: 61.6%

Modelo 2

$N = 10.000$, $d = 25$



Candidato	Porcentaje
Kast	36.859793%
Sichel	13.660811%
Boric	31.416833%
Provoste	18.062562%

Niveles de Aceptación:

Aceptados dia 1: 100.0%

Aceptados dia 2: 16.61%

Aceptados dia 3: 16.16%

...

Aceptados dia 18: 13.58%

Aceptados dia 19: 95.14%

...

Aceptados dia 96: 14.2%

Aceptados dia 97: 14.98%

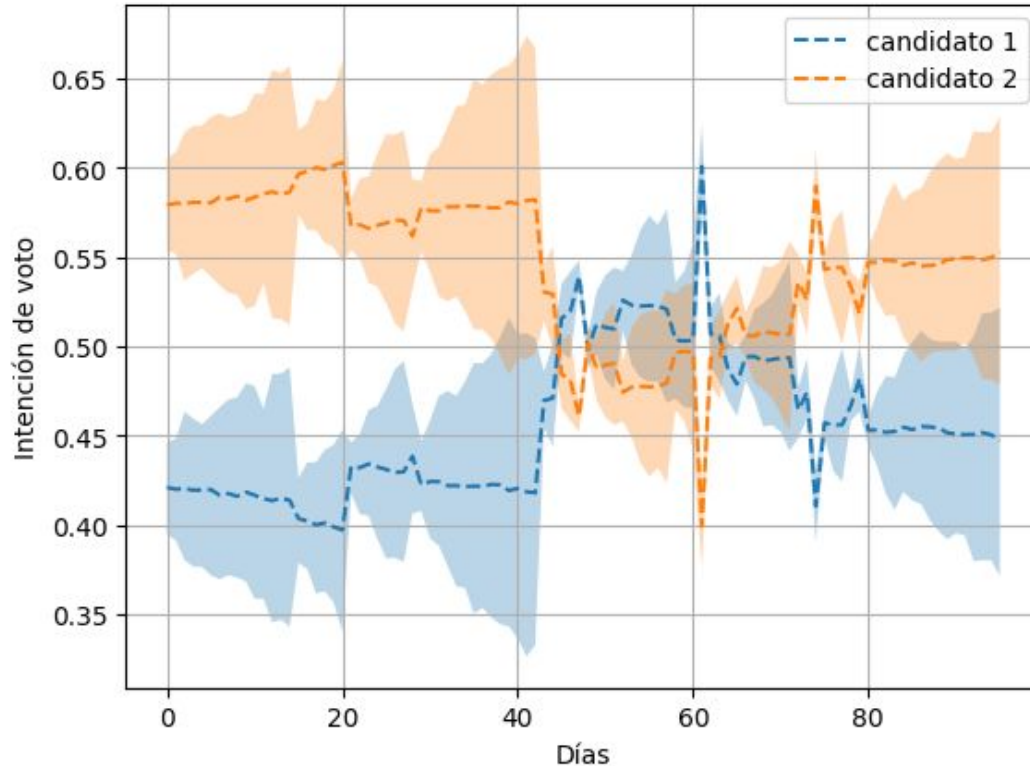


Resultados

2. Segunda vuelta

Modelo 1

N = 10.000



Candidato	Porcentaje
Kast	44.726163%
Boric	55.273837%

Niveles de Aceptación:

Aceptados dia 1: 100.0%

Aceptados dia 2: 71.35%

Aceptados dia 3: 71.01%

...

Aceptados dia 15: 70.94%

Aceptados dia 16: 94.05%

...

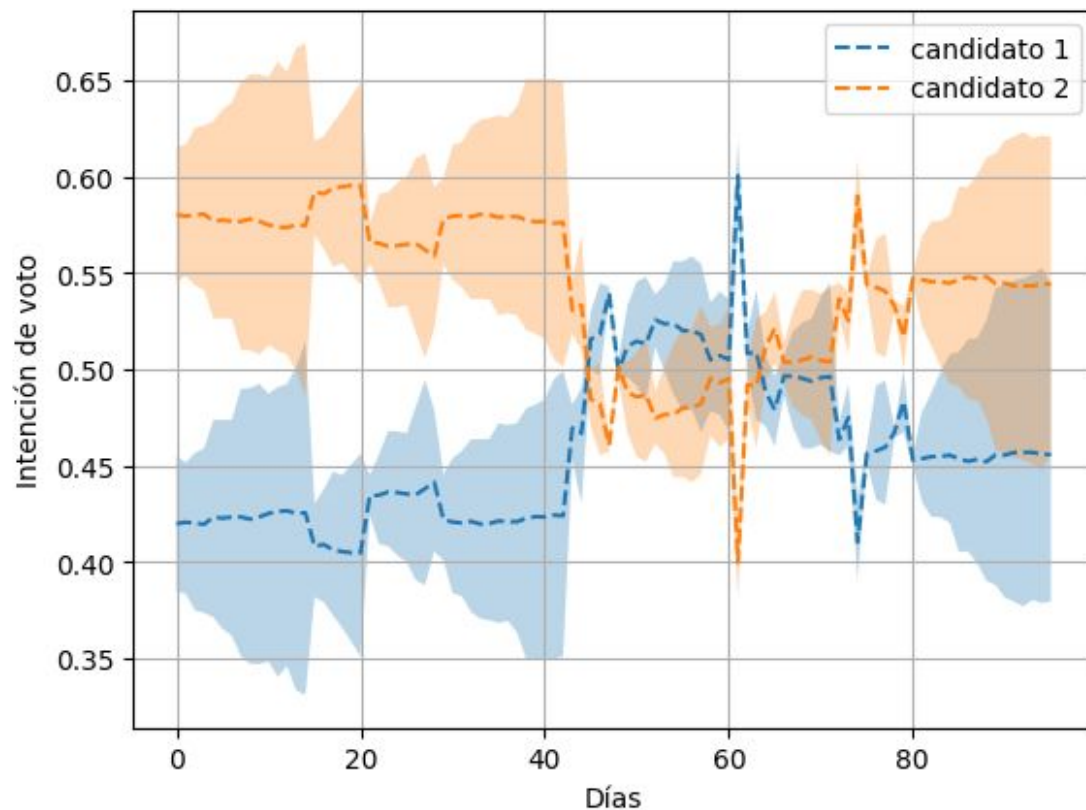
Aceptados dia 94: 70.79%

Aceptados dia 95: 70.96%

Aceptados dia 96: 70.19%

Modelo 2

$N = 10.000$, $d = 50$



Candidato	Porcentaje
Kast	45.566872%
Boric	54.433128%

Niveles de Aceptación:

Aceptados día 1: 100.0%

Aceptados día 2: 43.55%

Aceptados día 3: 43.95%

...

Aceptados día 18: 43.87%

Aceptados día 19: 43.36%

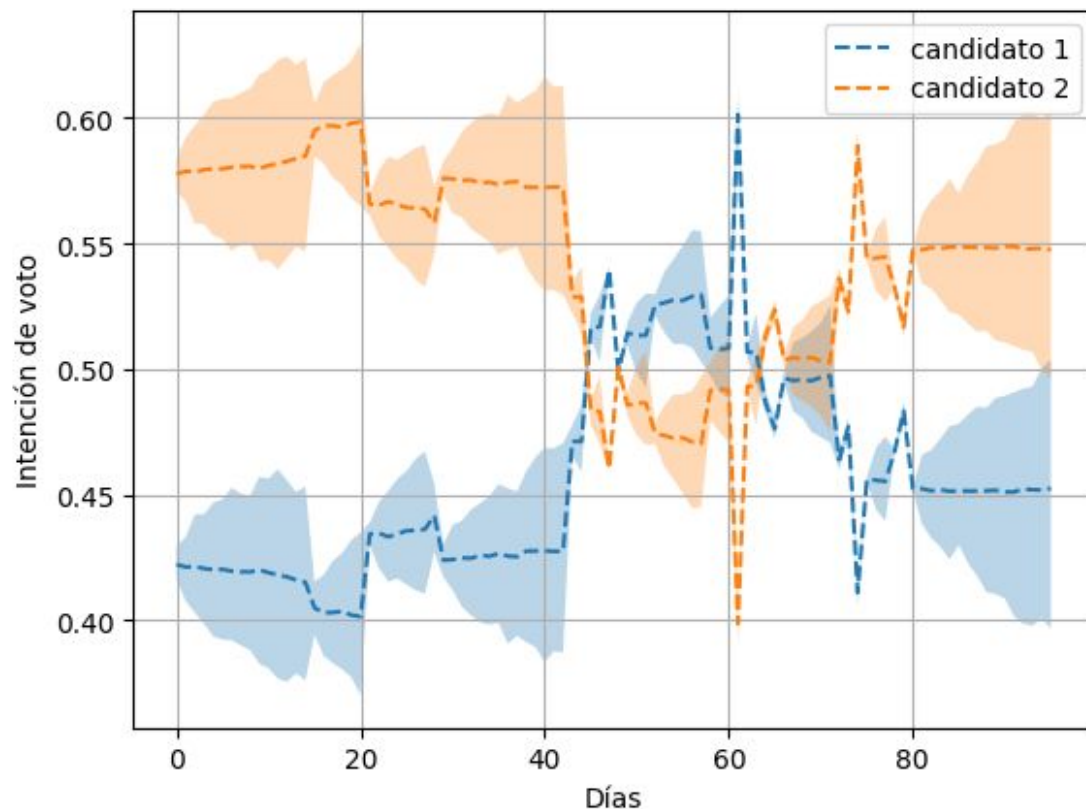
...

Aceptados día 95: 44.54%

Aceptados día 96: 45.39%

Modelo 2

$N = 10.000$, $d = 10$



Candidato	Porcentaje
Kast	45.250688%
Boric	54.749312%

Niveles de Aceptación:

Aceptados día 1: 100.0%

Aceptados día 2: 94.12%

Aceptados día 3: 93.36%

...

Aceptados día 18: 93.72%

Aceptados día 19: 93.17%

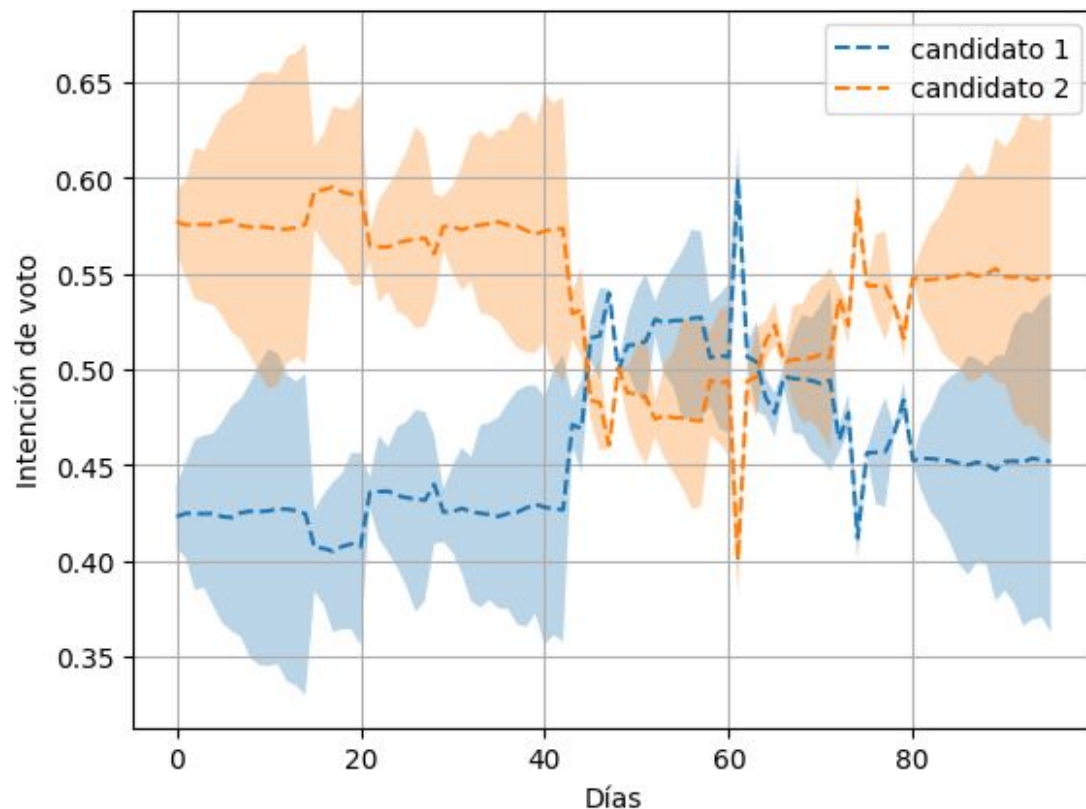
...

Aceptados día 95: 93.53%

Aceptados día 96: 93.86%

Modelo 2

$N = 10.000$, $d = 25$



Candidato	Porcentaje
Kast	45.173416%
Boric	54.826584%

Niveles de Aceptación:

Aceptados día 1: 100.0%
Aceptados día 2: 71.01%
Aceptados día 3: 70.72%
...
Aceptados día 18: 70.44%
Aceptados día 19: 71.16%
...
Aceptados día 95: 71.95%
Aceptados día 96: 71.59%



Análisis

Predicciones vs Resultados reales

	Primera Vuelta		Segunda Vuelta	
Candidatos	Votos	Porcentaje	Votos	Porcentaje
J.Kast	1.961.122	27,91%	3.649.647	44,13%
S.Sichel	895.510	12,79%		
G.Boric	1.814.809	25,83%	4.620.671	55,87%
Y.Provoste	815.558	11,61%		

(Emol,2021)



Primera Vuelta

- Kast = 27,81%
- Sichel = 12,78%
- Boric = 26,83%
- Provoste = 11,81%

	Porcentajes (%)			
Candidato	M1	M2 d = 50	M2 d= 10	M2 d=25
Kast	37.094199%	37.180306%	37.610321%	36.859793%
Sichel	13.872800%	13.823127%	12.962773%	13.660811%
Boric	31.434299%	31.049486%	31.818190%	31.416833%
Provoste	17.598701%	17.947081%	17.608716%	18.062562%



Segunda Vuelta

- Kast = 44,18%
- Boric = 55,87%

	Porcentajes (%)			
Candidato	M1	M2 d = 50	M2 d= 10	M2 d=25
Kast	44.726163%	45.566872%	45.250688%	45.173416%
Boric	55.273837%	54.433128%	54.749312%	54.826584%

Variaciones: Parámetro d

Primera vuelta

1) $d = 50$:

- Aceptación promedio: entre el 2.5%

2) $d = 10$:

- Aceptación promedio: entre el 65%

3) $d = 25$:

- Aceptación promedio: entre el 15%

Segunda vuelta

1) $d = 50$:

- Aceptación promedio: entre el 44%

2) $d = 10$:

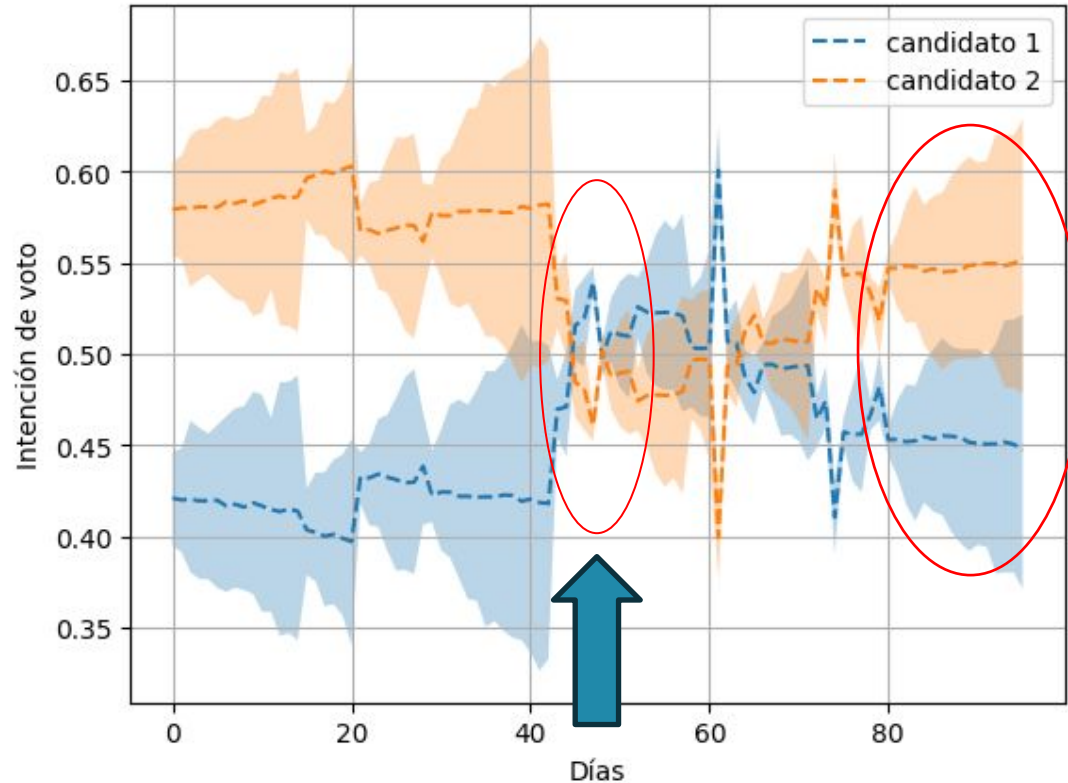
- Aceptación promedio: entre el 93%

3) $d = 25$:

- Aceptación promedio: entre el 71%

Intervalos de credibilidad

Ejemplo:





Conclusiones

¡Gracias por su atención!

Referencias

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

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

Tres Quintos. <http://www.tresquintos.com/>