Politechnika Wrocławska

Prognozowanie wyników wyborów przy pomocy modelu epidemiologicznego SIS

Jakub Koral

Wrocław, 28 listopada 2021

HR EXCELLENCE IN RESEARCH

Kilka słów o modelach epidemiologicznych

SIS (ang. Susceptible-Infectious-Susceptible) to najprostszy model epidemiologiczny. Populację dzieli się na dwie grupy: osoby podatne na zarażenie (susceptible) i osoby zarażone (infectious). Używa się go do modelowania chorób, po których przejściu nie utrzymuje się długotrwała odporność i można ponownie zachorować (np. grypa, przeziębienia).

Model SIS, czyli układ równań różniczkowych

Załóżmy, że S i I są odpowiednio proporcjami osób podatnych na zarażenie (zdrowych) i osób zarażonych (chorych), w dobrze wymieszanej populacji, w czasie t. Rozwój epidemii można wtedy opisać przy pomocy układu równań różniczkowych

$$\begin{cases} \frac{\mathrm{d}\,S}{\mathrm{d}\,t} = \gamma I - \beta S I, \\ \frac{\mathrm{d}\,I}{\mathrm{d}\,t} = -\gamma I + \beta S I, \\ S + I = 1, \end{cases} \tag{1}$$

gdzie γ to współczynnik wyzdrowień, a β to współczynnik zachorowalności.

Rozwiązanie modelu SIS

Rozwiązanie układu równań (1) dla I zadane jest wzorem

$$I(t) = \frac{\beta - \gamma}{e^{(\gamma-\beta)t}\left(\frac{\beta-\gamma}{I_0} - \beta\right) + \beta}, \tag{2}$$

gdzie $I_0 = I(0)$. Rozważmy co dzieje się w "długim" horyzoncie czasowym

$$\lim_{t\to\infty} I(t) = 0 \quad \text{dla} \quad \frac{\beta}{\gamma} \leqslant 1, \tag{3}$$

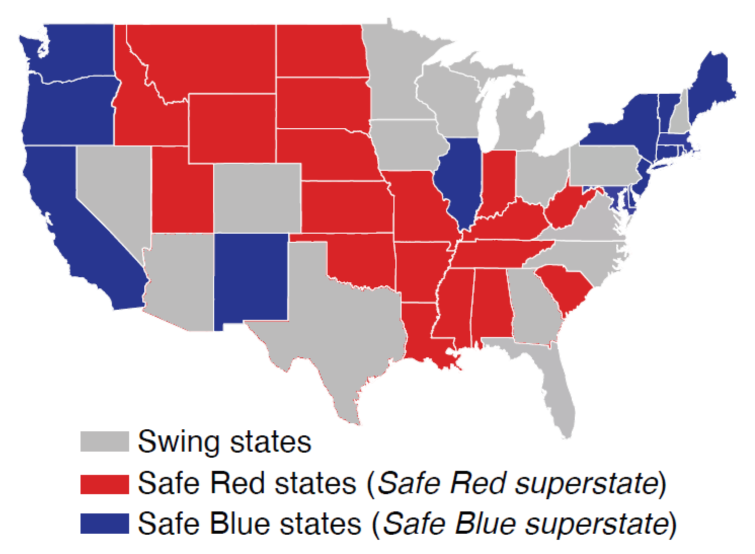
$$\lim_{t\to\infty} I(t) = 1 - \frac{\gamma}{\beta} \quad \text{dla} \quad \frac{\beta}{\gamma} > 1. \tag{4}$$

W pierwszym przypadku epidemia ostatecznie wygasa, w drugim zaś zawsze co najmniej $1 - \frac{\gamma}{\beta}$ populacji będzie chora.

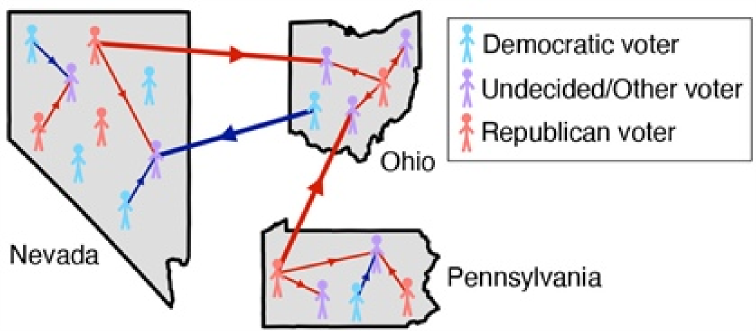
Jak wybiera się prezydenta w Stanach Zjednoczonych?

Amerykańskie wybory prezydenckie mają charakter pośredni. Wyborcy tak naprawdę wybierają 538 elektorów, którzy później głosują na kandydatów partii. W 48 stanach i Dystrykcie Kolumbii przewaga jednym głosem nad kandydatem, który zajął drugie miejsce, skutkuje wyborem wszystkich elektorów reprezentujących daną partię w określonym stanie. Prezydentem zostaje osoba, która zdobędzie co najmniej 270 głosów elektorskich.

Stany bezpieczne i wahające się



Porównanie modelu SIS i wyborów



Rozbudowanie modelu SIS

Dostosujmy model SIS do amerykańskiej sceny politycznej. Zatem:

- S^i = proporcja wyborców niezdecydowanych w stanie i,
- I_D^i = proporcja wyborców Demokratów w stanie i,
- I_R^i = proporcja wyborców Republikanów w stanie i,

gdzie oczywiście $S^i + I_D^i + I_R^i = 1$. Niech N będzie liczbą wszystkich wyborców w USA i N^j liczbą osób uprawnionych do głosowania w stanie j. Dodatkowo zakładamy, że wybory odbywają się w M stanach.

Deterministyczny model wyborów

Analogicznie do (1) można zapisać układ równań różniczkowych

$$\begin{cases} \frac{\mathrm{d}\, I_D^i}{\mathrm{d}\, t} = -\gamma_D^i I_D^i + \sum_{j=1}^M \beta_D^{ij} \frac{N^j}{N} S^i I_D^i, \\ \frac{\mathrm{d}\, I_R^i}{\mathrm{d}\, t} = -\gamma_R^i I_R^i + \sum_{j=1}^M \beta_R^{ij} \frac{N^j}{N} S^i I_R^i, \\ \frac{\mathrm{d}\, S^i}{\mathrm{d}\, t} = \gamma_D^i I_D^i + \gamma_R^i I_R^i - \sum_{j=1}^M \beta_D^{ij} \frac{N^j}{N} S^i I_D^i - \sum_{j=1}^M \beta_R^{ij} \frac{N^j}{N} S^i I_R^i, \\ S^i + I_D^i + I_R^i = 1, \end{cases} \tag{5}$$

z odpowiednimi współczynnikami zmiany poglądów.

W stronę modelu niedeterministycznego

Wyżej wspomniany model jest deterministyczny (zatem wynik będzie zawsze ten sam dla tych samych współczynników). Uwzględnia jedynie dane zebrane w ramach sondaży. By lepiej odzwierciedlić rzeczywistość, warto uwzględnić również dane statystyczne o pewnych kluczowych grupach społecznych.

Niedeterministyczny model wyborów

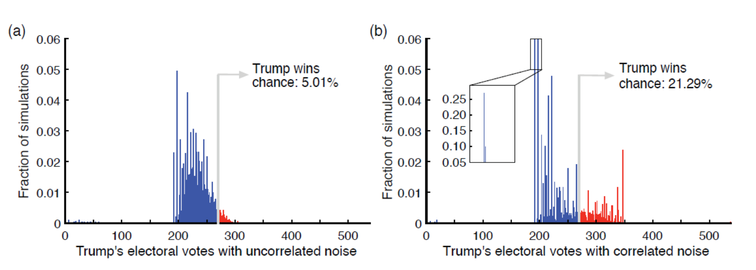
Dodając do modelu "szum", otrzymujemy układ równań różniczkowych

$$\begin{cases} \mathrm{d}\, I_D^i = \left(-\gamma_D^i I_D^i + \sum_{j=1}^{M} \beta_D^{ij} \frac{N^j}{N} S^i I_D^i\right) \mathrm{d}\, t + \sigma\, \mathrm{d}\, W_D^i, \\ \mathrm{d}\, I_R^i = \left(-\gamma_R^i I_R^i + \sum_{j=1}^{M} \beta_R^{ij} \frac{N^j}{N} S^i I_R^i\right) \mathrm{d}\, t + \sigma\, \mathrm{d}\, W_R^i, \\ \mathrm{d}\, S^i = \left(\gamma_D^i I_D^i + \gamma_R^i I_R^i - \sum_{j=1}^{M} \beta_D^{ij} \frac{N^j}{N} S^i I_D^i - \sum_{j=1}^{M} \beta_R^{ij} \frac{N^j}{N} S^i I_R^i\right) \mathrm{d}\, t + \sigma\, \mathrm{d}\, W_S^i, \\ S^i + I_D^i + I_R^i = 1, \end{cases} \tag{6}$$

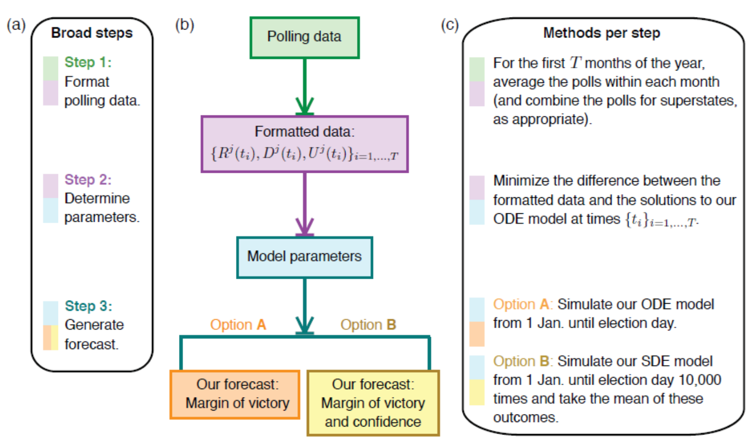
gdzie procesy Wienera W_D^i, W_R^i, W_S^i są odpowiednio skorelowane pomiędzy stanami.

Skorelowane i nieskorelowane procesy Wienera

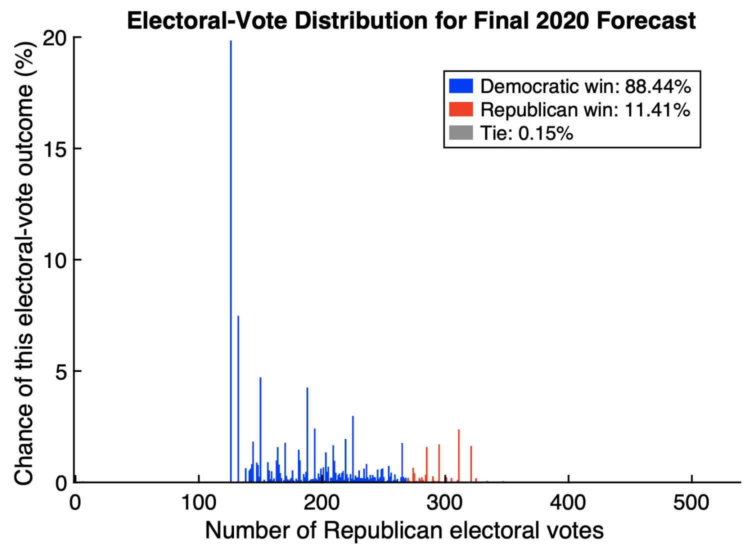
Dobrym przykładem jak ważne jest skorelowanie procesu Wienera były wybory prezydenckie w 2016 roku:



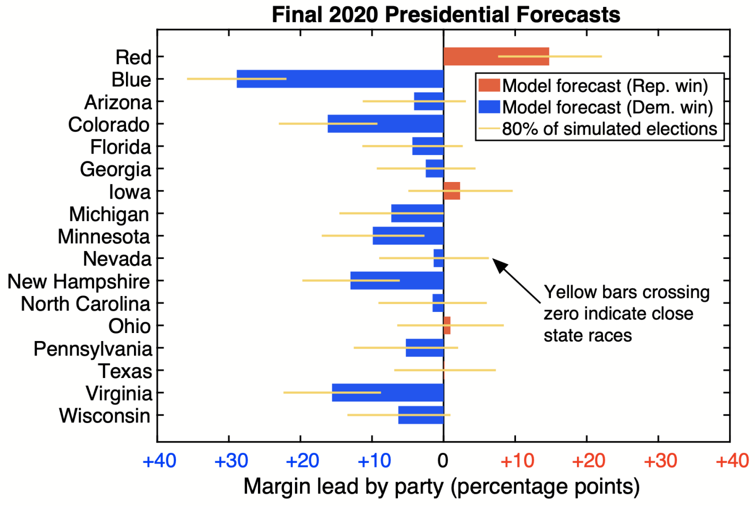
Podsumowanie procesu tworzenia prognozy



Wyniki symulacji dla wyborów prezydenckich z 2020 roku (1)



Wyniki symulacji dla wyborów prezydenckich z 2020 roku (2)

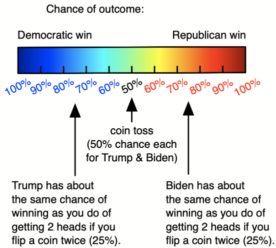


Wyniki symulacji dla wyborów prezydenckich z 2020 roku (3)

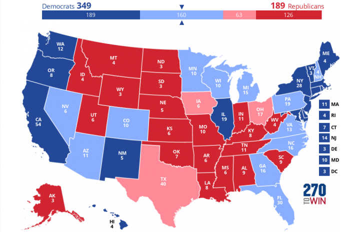
State ratings (leading candidate's chance of winnning)

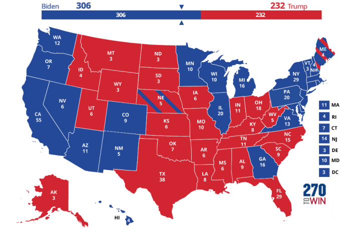
	Aug. 3	Aug. 14	Sep. 3	Sep. 14	Sep. 27	Oct. 5	Oct. 19	Oct. 23	Oct. 25	Oct. 27	Oct. 28	Oct. 29	Oct. 31	Nov. 2
Red	87.0	88.6	91.6	92.6	95.0	98.5	99.5	99.6	99.6	99.6	99.4	99.4	99.4	99.6
Blue	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Arizona			74.7	75.8	81.0	81.6	75.1	78.5	77.9	76.9	78.3	78.0	78.3	76.9
Colorado	99.9	99.8	99.4	99.0	99.7	99.8	99.8	99.8	99.9	99.9	99.9	100.0	99.9	99.8
Florida	90.4	87.9	85.5	80.7	82.1	85.0	79.1	79.2	79.5	77.5	78.6	77.4	78.3	78.5
Georgia						60.5	62.9	67.4	65.9	66.1	67.0	66.5	65.8	67.8
Iowa	58.0	52.2	64.5	62.2	63.4	56.2	66.5	64.3	64.5	64.8	64.6	65.5	64.3	65.1
Michigan	89.4	88.2	88.2	87.7	90.7	93.5	89.4	89.8	89.9	89.8	89.7	88.7	89.8	90.0
Minnesota	88.8	85.6	88.8	93.4	95.4	95.5	95.1	96.0	95.4	95.6	96.4	95.9	96.0	96.1
Nevada	77.7	74.2	79.1	83.6	69.2	74.7	62.5	61.7	62.7	62.1	61.5	61.5	63.1	59.3
New Hampshire	94.4	96.7	87.3	89.8	91.2	95.7	99.2	99.4	99.4	99.3	99.3	99.5	99.4	99.4
North Carolina	59.4	53.4	57.0	56.5	56.8	55.9	58.2	60.3	63.4	63.2	60.5	60.0	63.1	60.2
Ohio	66.0	50.0	70.9	65.4	53.4	51.6	55.2	58.2	57.7	59.7	57.5	57.9	58.4	56.3
Pennsylvania	79.9	79.5	78.7	79.9	80.6	83.4	82.6	81.8	84.5	81.3	84.0	83.4	82.1	81.9
Texas					57.4	54.3	51.1	50.2	50.7	50.3	50.4	50.4	52.1	50.5
Virginia	99.1	95.2	98.9	99.4	99.7	99.1	99.8	99.8	99.8	99.8	99.8	99.9	99.8	99.8
Wisconsin	84.9	87.9	84.1	85.6	82.3	89.0	84.9	85.4	87.0	87.1	87.0	86.3	86.9	86.7

Date of forecast



Porównanie z oficjalnymi wynikami wyborów prezydenckich z 2020 roku





Źródła

- **Artykuł:** Volkening, A., Linder, D.F., Porter, M.A., Rempala, G.A. (2020). *Forecasting elections using compartmental models of infection.* SIAM Rev., 62(4), 837–865.
- **Post na SIAM Blog:** https://sinews.siam.org/Details-Page/forecasting-elections-with-a-model-of-infectious-diseases
- **Oficjalna strona internetowa projektu:** https://modelingelectiondynamics.gitlab.io/2020-forecasts/