Model Thinking

Frank Sanacory
`sanacoryf@oldwestbury.edu`

SUNY College at Old Westbury

Contents

---

1 Introdction

2 Segregation and Peer Effects

2.1 Schelling

2.2 Granovetter

2.3 Standing Ovation Model

2.4 Identification Problem

3 Aggregation

3.1 Binomial Distribution and Normal Distribution

3.2 Game of Life

3.3 One Dimensional Cellular Automata

3.4 Preferences

4 Decision Models

4.1 Multicriterion

4.2 UnderUncertainty

4.3 Value of Information

5 Categories, Linear Models and the wisdom of Crowds

5.1 Categorical Models

Categorizing data helps to explain the variation in the data.

Example. You have the calorie data for a bunch of fruits and desserts. If you compute the total variation (sum of squared difference between the data point and the mean), you get a certain large figure. When you split the objects into fruits and desserts, and calculate the variation of each object to the mean of the class of objects, then total variation drops down a lot. If Total Variation = 53,200, Fruit Variation = 200, Dessert Variation = 5000, then categorizing helped to explain

$$(53, 2005, 200)/53, 200 = 48, 000/53, 200 = 90.2\%$$

of the variation (ie R-Squared is 0.902)

5.2 Linear Models

- Linear Models Interpreting Output
- R-Squared. The % amount of total variation (calculated in the same way above using the mean of all data points) explained by the line (variation after the line is calculated by sum of squared distances between the data points and the line).
- Standard error of regression. How much variation was there in the data to begin with
- Standard error of coefficient
  - 68% confidence that coefficient value is ± 1 standard error from the estimated coefficient.
  - 95% confidence that coefficient value is ± 2 standard error from the estimated coefficient.
- P-value of a coefficient
  - Probability that the sign of the coefficient is wrong.
  - So p-value = 0 and coefficient is positive means that we are absolutely sure that the coefficient is positive. If p-value = 1.4
- Linear Models (Two ways to incorporate nonlinearity)
  - Fit straight lines to different segments of the data
  - Use non-linear terms in the linear regression
- Linear Models Pitfalls To Be Aware Of
  - Correlation is not causation
  - Extrapolation is affected by f
  - Feedback: Other variables changing as a result of the change in your independent variable
  - Big coefficient thinking (i.e. only focusing on the variables with big coefficients) can make people blind to new innovative approaches to addressing issues.

5.3 Wisdom of Crowds

- True value = θ
- average prediction = μ
- Crowd's error = $(\theta - \mu)^2$
- Average error = $\frac{1}{n} \sum_{i=1^n} (\theta - x_i)^2$
- Diversity (average variation) = $\frac{1}{n} \sum_{i=1^n} (\mu - x_i)^2$

Theorem 5.1 (Diversity Prediction Theorem): Crowd error = Average error - Diversity

$$(\theta - \mu)^2 = \frac{1}{n} \sum_{i=1^n} (\theta - x_i)^2 - \frac{1}{n} \sum_{i=1^n} (\mu - x_i)^2$$

Note

- Wisdom of crowds come from reasonably smart people who are diverse.
- Madness of crowds come from like-minded people who are all wrong.

6 Game Theory

7 Tipping Points

7.1 Percolation Model

You have a 2-D grid. Let P be the probability that you fill in a square. You can move from one filled in square to another. For a grid to percolate, you must be able to move from one side to the other side of the grid. When P ¡ 59.2746%, a graph doesnt percolate. Once P goes above that tipping point it percolates.

7.2 Diffusion, SIS and SIR

7.2.1 Diffusion Model

At time t, W_t have the disease, N is the toal number of people and τ is the transmission rate. When two people meet, the probability of one transmitting to the other $= (W_t/N)((NW_t)/N)\tau$ c is the contact rate the probability of there being a meeting between two people, so Nc is the expected number of meetings out of N people.

$Nc(W_t/N)((NW_t)/N)\tau =$ number of new people contracting the disease

Diffusion $W_(t+1) = W_t + N * c * (W_t/N) * ((NW_t)/N) * \tau$ Process starts out slow, it accelerates, then it decelerates. There is no tipping point.

7.2.2 SIS Model (Susceptible, Infected, and then Susceptible)

a is the rate of infected people getting better

$$\begin{aligned} W_(t+1) &= \text{ number of infected in previous period + number of newly infected number cured} \\ &= W_t + Nc(W_t/N)((NW_t)/N)\tau a W_t \\ &= W_t + W_t(c\tau(NW_t)/Na) \end{aligned}$$

Basic reproduction number If W_t is small, then $(NW_t)/N \approx 1$ The disease spreads if $c * \tau > a$, i.e. $c\tau/a > 1$. $c\tau/a =$ basic reproduction number Tipping point for spreading is when $c\tau/a > 1$. Example usage When you vaccinate V% of people, the reproduction number drops by that percentage. Old reproduction number * (1 V) = New reproduction number To get new production number ¡= 1 1 1 / old reproduction number ¡= V So we can calculate the % V that of the population required to be vaccinated to prevent a disease from spreading, using the reproduction number of a disease.

7.2.3 SIR Model (Susceptible, Infected, and then Recovery)

Unlike the SIS model, some diseases after you recover from them you dont get it again. In such situations, use the SIR Model. Direct vs. Contextual Tips Direct tip: small action or event that has a large effect on end state. Contextual tip: change in the environment by a tiny bit has a huge effect on the end state (e.g. percolation model)

7.3 Growth

7.3.1 Exponential Growth for Populations & Economies

7.3.2 Solow Model

Basic Growth Model

1. Assumption 1: Output is increasing and concave in labor and machines $O_t = \sqrt{L_t}\sqrt{M_t}$
2. Assumption 2: Output is consumed or invested $O_t = E_t + I_t$ $I_t = s * O_t$, where s is the savings rate
3. Assumption 3: Machines can be built but they depreciate $M_(t+1) = M_t + I_t d * M_t$, where d is the depreciation rate

Long run equilibrium occurs when Investment = Depreciation

- $I_t = d * M_t$
- $s * O_t = d * M_t$
- $s * Sqrt(L_t) * Sqrt(M_t) = d * M_t$

Solve for M_t

Solow Growth Model Adds an additional Technology paramter Output $O_t = A_t * K_t^b eta * L_t^(1 - beta)$

- $A_t =$ Technology at time t
- $K_t =$ Capital at time t
- $L_t =$ Labor at time t

Innovation multiplier

- If $A_t = 2$, equilibrium output is 4x the basic growth model.
- If $A_t = 3$, equilibrium output is 9x the basic growth model.
- As labor and capital become more productive, there are more incentives to invest in more capital.

7.4 Tip or not Tip?

Measuring Tippyness

7.4.1 Diversity Index

Tells you approximately how many different types of things there are.

$$\text{Diversity index } = \frac{1}{\sum_{i=1}^{n} p_i^2}$$

That is so for 4 possible outcomes each equally likeley $(p_i 1/4)$ the diversity index is 4.

7.4.2 Entropy

$$\text{Entropy} = \sum(p_i * log_2(p_i))$$

Entropy tells us the number of pieces of information we have to know in order to identify the outcome (i.e. type). E.g. when you have 4 possible outcomes each with 1/4 probability, entropy = 2. You just need to ask 2 questions,

1. is it the first two outcomes? yes / no,

2. after you know which two, just ask is it one of the two, and you will know what the final outcome is using 2 pieces of information.

Definition of a tipping point

When something goes over a tipping point, the diversity index or entropy goes up or down, e.g. initially you can go left or right, after the tipping point you can only go right.