Estimating transition times for a model of language change in an age-structured population

---

W. Garrett Mitchener ⋄ *MitchenerG@cofc.edu*

http://mitchenerg.people.cofc.edu

COLLEGE of CHARLESTON

Abstract:

Human languages are stable on long time scales but have a tendency to change dramatically in a matter of a few decades. Some changes can be attributed to contact between languages, but others seem to be essentially spontaneous. In this project, I explore part of the machinery of language learning and variation that seems to contribute to spontaneous changes: Children seem to detect correlations between language variation and age and social status, and amplify them as they grow up. These effects are called *regularization* and *incrementation*. To model these, I have formulated a discrete Markov chain, and its limit as a stochastic differential equation, for a population with two age groups. The population tends to settle at a stable equilibrium dominated by one language variant, then switch spontaneously to an equilibrium dominated by another. The mean transition time can be calculated using a related convection-diffusion PDE, but that turns out to be numerically stiff. Alternatively, an asymptotic estimate of the mean time between transitions can be computed using the action functional. I demonstrate the calculation as performed with the Mathematica computer algebra system.

Language change:

❄ Middle English (1100 AD to 1500 AD) *verb-raising* syntax:

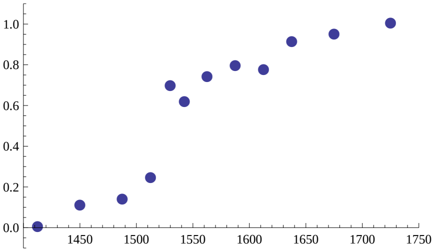
- Know you what time it is?
- I know not what time it is.

❄ Early Modern & Modern English do-*support* syntax:

- Do you know what time it is?
- I don’t know what time it is.

Problem: Human languages are stable on long time scales but can change spontaneously over decades.

Forces: Children detect correlations between language variation and age and social status, and amplify them.



Fraction of transitive affirmative questions using do*-support instead of verb-raising [2, 4]*

Learning model: Children modify language as they learn

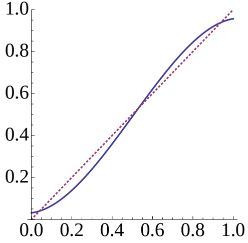
Regularization: Children sometimes drop rarely used forms

- Two idealized grammars: G_1 and G_2

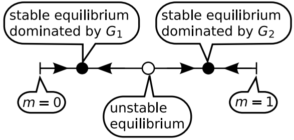
⇒ only one independent variable:

- m = mean usage rate of G_2

$$\frac{\mathrm{d}m}{\mathrm{d}t} = \overbrace{q(m)}^{\text{birth}} - \overbrace{m}^{\text{death}}$$



Mean learning function $q(m)$

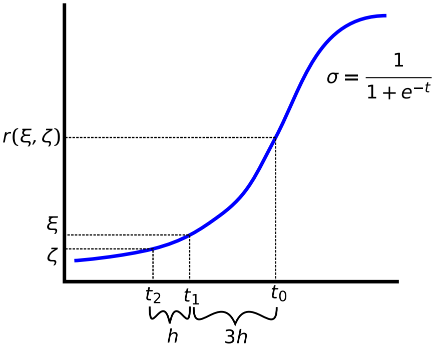


Dynamics of m

Incrementation: Children can detect and amplify trends

Two age groups:

- ξ = mean speech of young generation
- ζ = mean speech of old generation
- $r(\xi, \zeta)$ = where children predict the population will be

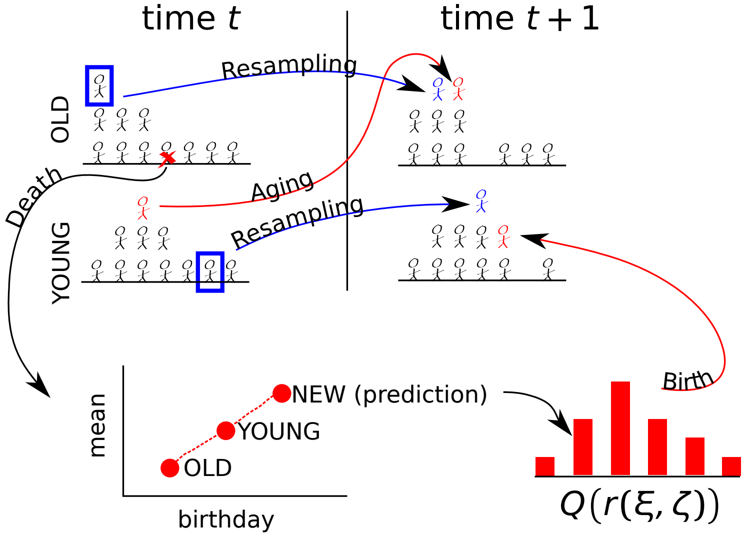


Prediction function

$$\frac{d\xi}{dt} = \overbrace{q(\underbrace{r(\xi, \zeta)}_{\text{prediction}})}^{\text{birth}} - \overbrace{\xi}^{\text{aging}} \qquad \frac{d\zeta}{dt} = \overbrace{\xi}^{\text{aging}} - \overbrace{\zeta}^{\text{death}}$$

Add some random fluctuations:

Start with a Markov chain:



Fine print about the Markov chain:

❈ N youth and N parents

❈ Most general: Each agent in a state $k \in \{0, 1, \ldots, K\}$, uses G_2 at a rate $\frac{k}{K}$

⇒ Simplify $K = 1$: Each agent uses G_1 (state 0) or G_2 (state 1)

❈ C_k = number youth in state k

❈ $X_k = \frac{1}{N} C_k$

⇒ Random sentence from uniformly selected youth is type k with probability X_k

❈ D_k = number parents in state k

❈ $Y_k = \frac{1}{N} D_k$

⇒ Random sentence from uniformly selected parent is type k with probability Y_k

❈ $Q(m) =$ binomial distribution, parameters $q(m)$ & K

$$Q_k(m) = \binom{K}{k} q(m)^k (1 - q(m))^{K-k}$$

❈ r_D = death rate: each time step each adult dies with probability $p_D = \frac{r_D}{N}$, replaced by sampling from youth; each youth ages with probability p_D, replaced by sampling from distribution vector $Q(r(X_1, Y_1))$

⇒ time-in-generation is geometric, mean $= \frac{N}{r_D}$ time steps

⇒ 1 time step $= \frac{1}{N}$ years

⇒ average life span $L = \frac{2N}{r_D}$ time steps $= \frac{2}{r_D}$ years

❈ r_R = resampling rate: each time step each agent copies the state of a random agent of the same generation with probability r_R

Limit of infinite population $N \to \infty$:

- ξ = fraction of youth using G_2
- ζ = fraction of parents using G_2

Rescale time: time in units of $\frac{1}{r_D} = \frac{L}{2}$ years

$$\text{Let } \varepsilon = \sqrt{\frac{1-(1-r_R)^2}{r_D}}$$

$$d\xi = (q(r(\xi,\zeta)) - \xi)\,dt + \varepsilon\sqrt{\xi(1-\xi)}\,dB^{\xi}$$
$$d\zeta = (\xi - \zeta)\,dt + \varepsilon\sqrt{\zeta(1-\zeta)}\,dB^{\zeta}$$

where B^{ξ} and B^{ζ} are Brownian motion

- From [1, 5]: (X_1, Y_1) converges weakly to (ξ, ζ)

Change variables:

$$\theta = \arcsin(2\xi - 1) \qquad \phi = \arcsin(2\zeta - 1)$$

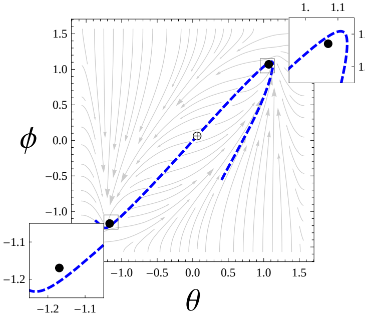
$$\mathrm{d}\theta = \overbrace{\left(\left(\frac{1}{2}\varepsilon^2 - 1\right)\tan\theta + \frac{2q\left(r(\xi,\zeta)\right) - 1}{\cos\theta}\right)}^{b_1(\theta,\phi)} \mathrm{d}t + \varepsilon\,\mathrm{d}B^{\xi}$$

$$\mathrm{d}\phi = \overbrace{\left(\left(\frac{1}{2}\varepsilon^2 - 1\right)\tan\phi + \frac{\sin\theta}{\cos\phi}\right)}^{b_2(\theta,\phi)} \mathrm{d}t + \varepsilon\,\mathrm{d}B^{\zeta}$$

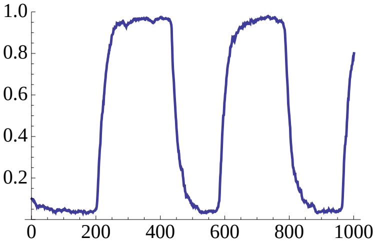
Phase space is

$$R = \left[-\frac{\pi}{2}, \frac{\pi}{2}\right] \times \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$$

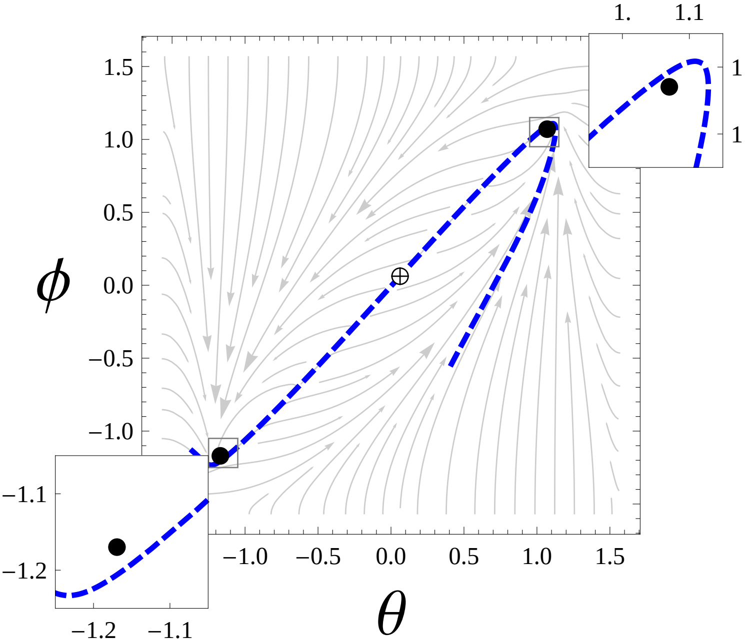
Vector field component:



(θ, ϕ) vector field: $\bullet$ = stable fixed points where both generations are dominated by G_1 or G_2; $\oplus$ = saddle, separates the basins



Mean usage rate of G_2 in young generation vs. time: The population hovers near one stable fixed point. Fluctuations drive it across the separatrix to the other stable fixed point.



1.5
1.0
0.5
0.0
−0.5
−1.0
ϕ
−1.0
−0.5
0.0
0.5
1.0
1.5
θ
1.
1.1
1
1
−1.1
−1.2
−1.2
−1.1

Estimating time between transitions:

From [3]: Given an SDE with ε noise:

$$\mathrm{d}X = b(X)\,\mathrm{d}t + \varepsilon\,\mathrm{d}B$$

❄ The *normalized action functional* is

$$S(f) = \frac{1}{2}\int_0^T \left\| f'(t) - b(f(t)) \right\|^2 \, \mathrm{d}t$$

Intuition: This measures how strongly a path $f : [0, T] \to R$ flows against the vector field—how much energy the random fluctuations must muster. If f solves $x' = b(x)$ then $S(f) = 0$.

❈ The *quasipotential* is

$$V(x_0, x) = \inf\{S(f) \mid f(0) = x_0, f(T) = x\}$$

Intuition: Minimum energy to go from x_0 to x.

❈ The *exit time* from a domain D is

$$\tau = \min\{t \mid X_t \notin D\}$$

❈ The expected exit time satisfies

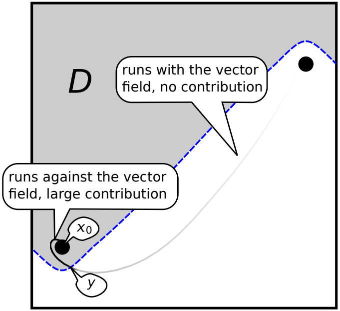
$$\lim_{\varepsilon \to 0} \varepsilon^2 \ln \mathbb{E}(\tau) = \min_{y \in \partial D} V(x_0, y)$$

Intuition: Minimum energy to best exit point y

$$\text{pick a small } \varepsilon\text{:}\ \mathbb{E}(\tau) \approx \exp\left(\frac{1}{\varepsilon^2} \min_{y \in \partial D} V(x_0, y)\right)$$

Calculation: Pick a small ε and compute V. In Mathematica, set up an `InterpolatingFunction` for the path f with unknowns for its values & derivatives at various times, then use `FindMinimum`.

❄ What to use for D? We really need the basin of attraction for D, but that's hard to compute. Instead, pin the ends of f to the stable fixed points:



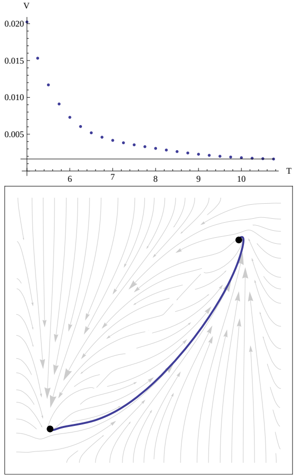
❄ Once the path leaves D, it can follow the vector field, which contributes 0 to the integral for S

❄ No need to find exit point y

❄ Let finish time T increase and look for convergence

❄ Let $\varepsilon = 0.02 \ldots$

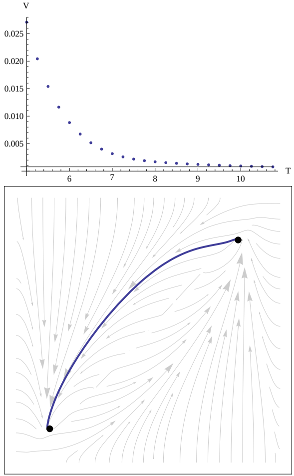
Bottom to top:



❄ $\min V \approx 0.00162956$

❄ $\mathbb{E}(\tau) \approx 58.7871$

Top to bottom:



❄ $\min V \approx 0.000777568$

❄ $\mathbb{E}(\tau) \approx 6.98613$

Why two dimensions are necessary:

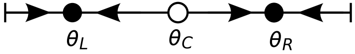
From [3]: For an SDE $\mathrm{d}X = b(X)\,\mathrm{d}t + \varepsilon dB$, if the vector field b is a gradient (always true in one dimension) then everything is simpler.

❄ Learning with no age groups, no incrementation, no prediction:

$$\mathrm{d}\xi = (q(\xi) - \xi)\,\mathrm{d}t + \varepsilon\sqrt{\xi(1-\xi)}\,\mathrm{d}B$$

❄ Change variables $\theta = \arcsin(2\xi - 1)$

$$\mathrm{d}\theta = \overbrace{\left(\left(\frac{1}{2}\varepsilon^2 - 1\right)\tan\theta + \frac{2q(\xi) - 1}{\cos\theta}\right)}^{b(\theta)}\mathrm{d}t + \varepsilon\,\mathrm{d}B$$



❈ Define

$$U(\theta_0, \theta) = -\int_{\theta_0}^{\theta} b(s)\,\mathrm{d}s$$

❈ From [3]: There's only one boundary point θ_C, and the minimum V is simple:

$$V(\theta_L, \theta_C) = 2U(\theta_L, \theta_C) \approx 0.123923$$

⇒ Mean transition time from left to right

$$\mathbb{E}(\tau) \approx \exp\left(\frac{1}{\varepsilon^2} V(\theta_L, \theta_C)\right) \approx 3.53612 \times 10^{134}$$

⇒ That language isn't going to change any time soon!

Intuition: The fluctuations have to overcome the vector field all the way from the stable fixed point to the unstable tipping point in the center. Eventually it must happen but it takes too much energy to occur on a reasonable time scale.
With two dimensions, the separatrix can be very close to the stable fixed point. With less of the vector field to overcome, the fluctuations can push the population state into another basin of attraction more easily.

What goes wrong with PDE approach:

❈ From [3]: Given a general SDE

$$dX = b(X)\,dt + \sigma(X)\,dB, \quad a(X) = \sigma(X)\sigma^*(X)$$

the corresponding differential operator is

$$Lu = \frac{1}{2}\sum_{i,j} a_{i,j}(x)\partial^2_{i,j}u(x) + \sum_i b_i(x)\partial_i u(x)$$

❈ The solution to

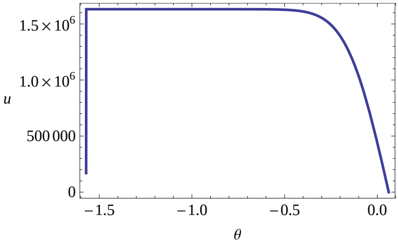
$$Lu = -1 \text{ on } D,\ u = 0 \text{ on } \partial D$$

is

$$u(x) = \mathbb{E}(\tau \mid \text{ start at } x)$$

❈ Could set it up on $D =$ a small disk around the stable fixed point, solve with a finite element method
⇒ But: numerically unstable, results are nonsense

Even in one dimension, the calculation is difficult because of the outside boundary conditions: Specifying $u(-\pi/2) = 0$ results in a numerical singularity.



Mean exit time from θ_L via diff eq, using $u(\theta_C) = 0$ and $u(-(1-10^{-6})\pi/2) = 0$ to avoid boundary singularity.

Intuition: The solution is a plateau, and the edges are very difficult to resolve numerically.

Conclusion:

❋ Formulated population dynamics with learning, including regularization and incrementation

❋ Mean time between transitions from one stable fixed point to the other:

⇒ Represents language change

⇒ Can be calculated with a PDE, but numerically unstable

⇒ Better to use an asymptotic approximation using the action functional

❋ Example calculation: Mean transition times on the order of 10 lifespans

⇒ Without age groups and incrementation: The model is first order, but the transition time is many orders of magnitude too large

⇒ Some sort of second order dynamics is necessary

References

[1] Richard Durrett. *Stochastic Calculus: A Practical Introduction*. CRC Press, New York, 1996.

[2] Alvar Ellegård. *The Auxiliary do: The Establishment and Regulation of Its Use in English*, volume II of *Gothenburg Studies in English*. Almqvist and Wiksell, 1953.

[3] M. I. Freidlin and A. D. Wentzell. *Random Perturbations of Dynamical Systems*, volume 260 of *Grundlehren der mathematischen Wissenschaften*. Springer Verlag, New York, 1984.

[4] Anthony Kroch. Reflexes of grammar in patterns of language change. *Language Variation and Change*, 1:199–244, 1989.

[5] W. Garrett Mitchener. A stochastic model of language change through social structure and prediction-driven instability. Submitted, 2009.