Sistemas sociales complejos

Juan Pablo Pinasco

Universidad de Buenos Aires - CONICET

http://mate.dm.uba.ar/~jpinasco

jpinasco@gmail.com

1era clase - 2016

Sistemas complejos

La verdad sobre China: ¿El embotellamiento más largo del mundo duró 12 días?

Daniel Canal | 13-01-2015 - 05:40:58

Mito: En China el tráfico vehicular es tal que la peor congestión registrada en la historia ocurrió en Beijing y duró 12 días.

Realidad: Cuando el conductor de un Peugeot 404 pasa de segunda marcha a t
del cuento "Autopista al sur", de Juli ... un embotell...
... dura lo suficiente pa

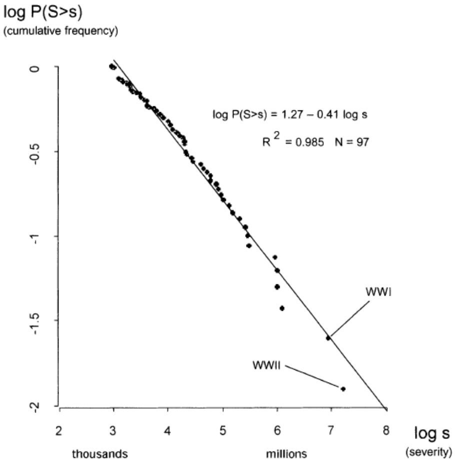
El 14 de agosto de 2010 empezó el embotellamiento más demorado en la historia, en la autopista 1 que conecta a Beijing, la capital china, con el Tíbet. Este atasco se demoró 12 días en descongestionarse y se extendió a lo largo de 100 kilómetros. Lo particular es que el embotellamie ocurrió sin ninguna razón aparente, solo porque había muchos autos en la vía, contrario al de Par ...lebres como el de Chicago en 2011, que fueron consecuencia del mal clima.

Veredicto: Sí, el embotellamiento más demorado del mundo ocurrió en China en una carretera al de Beijing. Los chinos debieron vivir casi dos semanas en sus autos, y como el protagonista de Cortázar, cuando se e... los carros y pasaron de segunda a tercera ...
dejar su vid


ADRIAN
y Los Dados Negros
RED
AV CORRIENTES A 3 CUADRAS DEL ABASTO S
ELPICAFLORBAILABLE
2014
22 hrs.

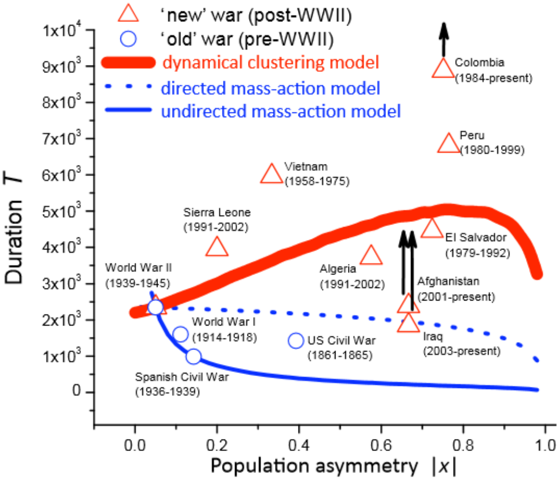
Cederman, 2003

FIGURE 1. Cumulative Frequency Distribution of Severity of Interstate Wars, 1820–1997



Source: COW data.

Zhao, 2013



'new' war (post-WWII)
'old' war (pre-WWII)
dynamical clustering model
directed mass-action model
undirected mass-action model
Duration T
1x10^4
9x10^3
8x10^3
7x10^3
6x10^3
5x10^3
4x10^3
3x10^3
2x10^3
1x10^3
0
0.0
0.2
0.4
0.6
0.8
1.0
Population asymmetry |x|
Colombia (1984-present)
Peru (1980-1999)
Vietnam (1958-1975)
Sierra Leone (1991-2002)
El Salvador (1979-1992)
Algeria (1991-2002)
Afghanistan (2001-present)
Iraq (2003-present)
World War II (1939-1945)
World War I (1914-1918)
US Civil War (1861-1865)
Spanish Civil War (1936-1939)

Algunas características

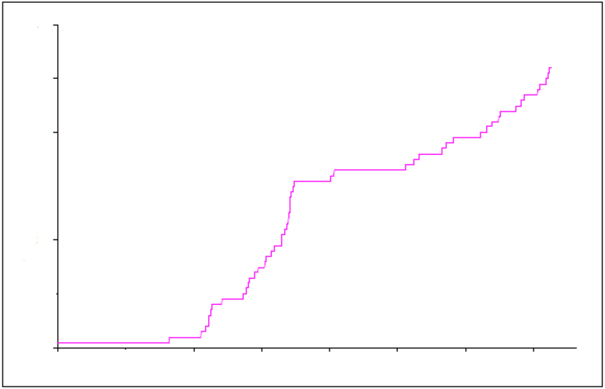
- micro motivos/interacciones, macro comportamientos/efectos (Schelling)
- feedback (beauty contest)
- auto-organización
- patrones emergentes
- transiciones al variar los parámetros
- edge of chaos

Parte I

Ejemplo

???

M.V. Simkin and V.P. Roychowdhury
Journal of Theoretical Biology 355 (2014) 111-116.



Stochastic modeling of a serial killer

M.V. Simkin and V.P. Roychowdhury - J Theor Biol 355 (2014) 111-116.

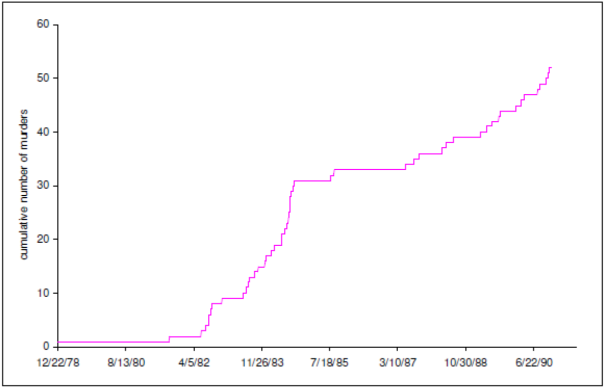


Figure 1. Chikatilo's staircase shows how the total number of his murders grew with time. The time span begins with his first murder on 12/22/1978 and ends with his arrest on 10/20/1990. The shortest interval between murders was three days and the longest – 986 days. The murder dates were determined based on the date on disappearance of the person in question.

Idea del modelo

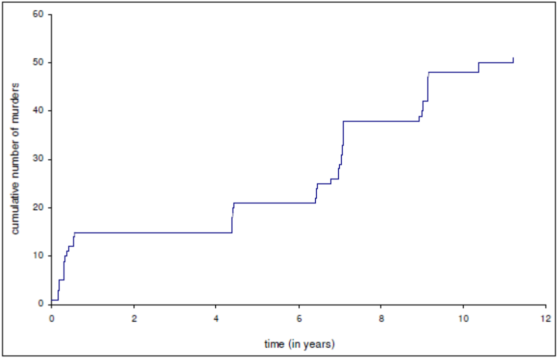
- Una neurona se activa sola con probabilidad p.
- Cuando una se activa manda un impulso a sus vecinas, que se activan con cierta probabilidad, y luego se desactiva.
- Sea Y_n al número de neuronas activas a tiempo n, y $Z_{n,j}$ a las activadas por alguna de las otras,

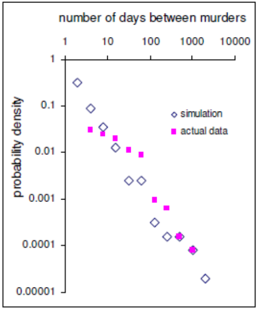
$$X_n = Y_n + \sum Z_{n,j}$$

  es el número de neuronas activas, si pasa cierto umbral o h, sale a matar gente, esto tiene un efecto sedante, y se desactivan.
- $Z \sim$Poisson, $X_n \sim$ Normal, se reduce a un paseo aleatorio y estimar los tiempos de retorno a la zona donde sale a matar.

Simulaciones

M.V. Simkin and V.P. Roychowdhury - J Theor Biol 355 (2014) 111-116.

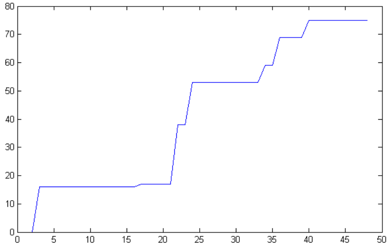


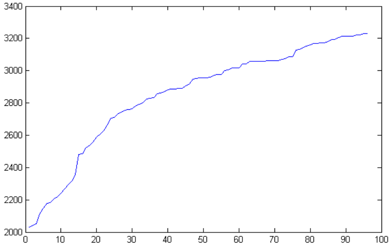


Algunas conclusiones

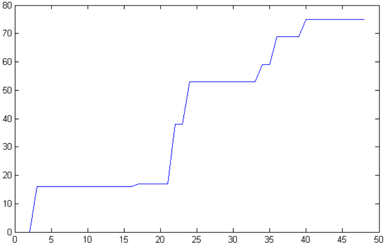
- Patrones similares de otros asesinos.
- Gran similitud con el comportamiento de los intervalos de tiempo entre ataques epilépticos (Osorio, Sornette, etc. Eur. J Neurosci. 2009, Phys Rev E 2010).
- "No killer commited enough murders to accumulate accurate intervals statistics".

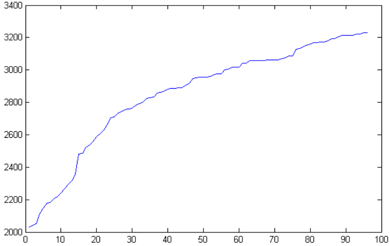
¿Y estos?


80
70
60
50
40
30
20
10
0
0
5
10
15
20
25
30
35
40
45
50


3400
3200
3000
2800
2600
2400
2200
2000
0
10
20
30
40
50
60
70
80
90
100

Fabiano Caruana (der)





Posible tema de final

"No killer commited enough murders to accumulate accurate intervals statistics".

Pero **sí** existen suficientes ajedrecistas amateurs para acumular datos y chequear el modelo.

Punto clave

Los buenos modelos abstraen mecanismos que pueden estar presentes en diferentes situaciones (y no siempre nos damos cuenta).

Se parecen mucho a un teorema general: no sirve para nada en los casos particulares, incluídos los que lo motivaron.

Parte II

Modelos

Modelos

Tres tipos de modelos:

1. Interacción con el medio. Ej: mariposas, influencia de la prensa en las opiniones [Maxwell].
2. Interacciones de a pares: formación de precios, distribución de riqueza, modelos de opinión [Boltzmann].
3. Interacciones grupales: Schelling, tráfico, muchedumbres, flocking.

Modelos

- Panorama general: propósito (qué queremos estudiar) - agentes, variables de estado, escalas de espacio y tiempo - esquema del programa.
- Diseño: puntos a considerar de las simulaciones.
- Detalles: datos iniciales, parámetros, datos reales, submodelos.

Diseño (I)

1.- Ppios básicos: qué se asume, de qué marco teórico.

2.- Emergencia: propiedades que se obtienen de las simulaciones, cuáles son obvias por las reglas y cuáles no.

3.- Adaptación: ¿cambian las variables de estado de los agentes?

4.- Objetivos: qué quieren los agentes? fitness (sexo), utilidad, etc.

5.- Learning: qué aprenden los agentes?

Diseño (II)

6.- Sensing: qué propiedades del medio son conocidas por los agentes?

7.- Interacción: cómo interactúan los agentes entre sí?

8.- Azar: hay reglas aleatorias? en qué influyen?

9.- Colectivos: ¿se forman macro-agentes?

10.- Observación: qué tipo de salida esperamos del programa? tablas? figuras? números sueltos?

Interacciones con el medio

Virtual corridors (Pe'er et al. 2005)

- Propósito.

Bajo qué condiciones aparecen corredores virtuales?

Cómo los afecta la probabilidad de no moverse al punto más alto?

- Detalles.

Usar una topografía simple, con un par de picos, y otra real, con la elevación de cada sitio.

Ponemos 500 mariposas en una pequeña región de una grilla de 150×150, cada lugar son unos $25 \times 25 m^2$.

Variar el parámetro q, y medir la longitud de la trayectoria de cada mariposa comparada con la distancia.

Programa

Virtual corridors (Pe'er et al. 2005, Railsback-Grimm 2012)

- Ubicar las mariposas en unas pocas casillas y fijar $q \in (0,1)$.
- En cada unidad de tiempo, tirar una moneda $\mathcal{U}(0,1)$. Si es menor a q, ir a la casilla vecina de mayor altura; si es mayor, moverla al azar.
- Frenar la simulación cuando está en un máximo local, o tras cierto número de pasos.
- Para cada tiempo, contar el número de casillas ocupadas (ancho del corredor virtual).
  - Corregir si las mariposas se dividen en dos o más grupos que van a distintos máximos locales.
- Para cada mariposa, calcular la distancia a la posición final y la cantidad de casillas que visitó.

Diseño

1.- Ppios básicos: trabajos de Pe'er, datos desde los '60.

2.- Emergencia: esperamos la aparición de corredores virtuales. En una topografía real podría haber más de un corredor, rutas a distintos máximos locales.

3.- Adaptación: no.

4.- Objetivos: llegar a lugares más elevados (tener sexo).

5.- Learning: nada.

Diseño

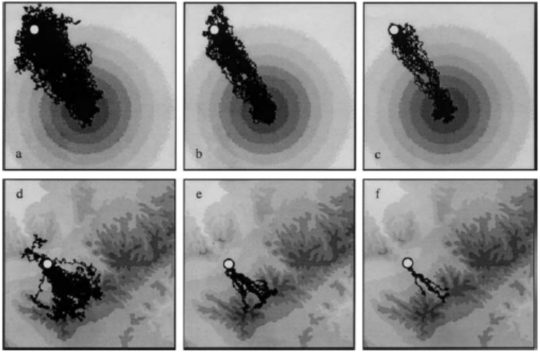
6.- Sensing: conocen la altura de los puntos vecinos

7.- Interacción: no hay.

8.- Azar: q, exploración aleatoria.

9.- Colectivos: en presencia de máximos locales, podrían formarse más de una comunidad.

10.- Observación: trayectoria de las mariposas, destinos finales, tiempos de viaje.


a
b
c
d
e
f

Modelo analítico

¿Se puede describir analíticamente este fenómeno?

Obvio! si no, elegía otro! Ver:

- Lewis et al. (2013) Dispersal, individual movement and spatial ecology, Lect. Notes in Math. 2071.

- Painter (2014) Multiscale models for movement in oriented environments and their application to hilltopping in butterflies. Theoretical Ecology 7, 53-75

Modelo analítico

Supongamos que las mariposas viven en una región $\Omega \subset \mathbb{R}^2$, y tienen una velocidad v, con $0 \leq v_{\text{mín}} \leq |v| \leq v_{\text{máx}} < \infty$.

Sea $p(t, x, v)$ la (densidad de) probabilidad de encontrar una mariposas a tiempo t en el lugar x, con velocidad v, y sea $p_0(x, v)$ la distribución inicial de mariposas a tiempo $t = 0$.

Modelo analítico

Cada individuo tiene asociado un proceso de Poisson y cuando su reloj marca que tiene que cambiar, un individuo elige una nueva velocidad.

La nueva velocidad v se elige de acuerdo a cierta distribución de probabilidad $K(t, x, v, w)$, donde w es la velocidad que traía.

Modelo analítico

Temporalmente, la densidad p varía como

$$p(t+h,x,v) = (las\ que\ estaban) - (las\ que\ se\ van) + (las\ que\ llegan)$$

$$p(t+h,x,v) = p(t,x,v) - (las\ que\ se\ van) + (las\ que\ llegan)$$

Podemos reescribir esto de manera más clara, separando la parte espacial de la velocidad:

Modelo analítico

Temporalmente, la densidad p varía como

$$p(t+h,x,v) = (\textit{las que estaban}) - (\textit{las que se van}) + (\textit{las que llegan})$$

$$p(t+h,x,v) = p(t,x,v) - (\textit{las que se van}) + (\textit{las que llegan})$$

Podemos reescribir esto de manera más clara, separando la parte espacial de la velocidad:

$$\partial_t p(t,x,v) = -\nabla \cdot vp(t,x,v) + \int_V K(x,v,w)\Big(p(t,x,w) - p(t,x,v)\Big)dw$$

Modelo analítico

$$\partial_t p(t,x,v) = -\nabla \cdot vp(t,x,v) + \int_V K(x,v,w)\Big(p(t,x,w) - p(t,x,v)\Big)dw$$

Esta es la famosa ecuación de Boltzmann [1844-1906].

Existen distintas aproximaciones posibles, y podemos tener transporte puro, pura difusión, o algo intermedio:

$$\begin{aligned}\partial_t p =& \nabla \cdot Ap \\ \partial_t p =& B\Delta p \\ \partial_t p =& B\Delta p + \nabla \cdot Ap.\end{aligned}$$

Modelo analítico

Recordemos que p era una medida (de probabilidad)... ¿qué quiere decir derivarla? ¿En qué sentido tenemos ahí una ecuación diferencial?

En las simulaciones del modelo de las mariposas tenemos un caso particular:

- hay una función W que nos da la altura del terreno,
- sabemos exactamente quién es la medida p (suma de deltas en la posición y velocidad de cada una),
- y cómo cambian de velocidad (una uniforme entre los vecinos con probabilidad q, la dirección de máximo crecimiento ∇W con probabilidad $1 - q$).

Modelo analítico

Problemas claramente matemáticos:

- ¿Cómo se relacionan estas cosas con los coeficientes A y B?
- ¿Podemos tomar el límite de infinitas mariposas, para tener una densidad p continua?
- ¿Cómo comparar el problema discreto con finitas mariposas y soluciones suaves de las ecuaciones que quedan?

La parte teórica del curso es responder este tipo de preguntas, y justificar -en la medida de lo posible- los distintos pasos al límite.

No nos interesará tanto la existencia, unicidad, y regularidad de las soluciones, *pero sí estudiar el límite* $t \to +\infty$.

Evolución

Otro problema:

La evolución temporal de la población es un problema interesante, más allá de la convergencia a tiempos largos.

Supongamos que cada agente tiene una característica, y que en el estado final ésta pudo variar.

Si recomenzamos todo el proceso, pero con la característica modificada, nos podemos preguntar si su distribución converge a algo, o si se presentan fenómenos nuevos.

Evolución

Por ejemplo:

Cada mariposa podría tener su propia tasa q_i, y en el estado final, los pares de mariposas (i,j) en una misma ubicación podrían reproducirse generando dos nuevas con tasas q_i', $q_j' \in [q_i, q_j]$. Las que quedan aisladas son removidas de la simulación.

Podemos analizar los *tiempos de extinción*, o mantener la población constante reemplazando las que mueren sin reproducirse por otras con q elegido al azar con la misma distribución inicial, o con la distribución de las mariposas que están actualmente.

Parte III

Todo muy lindo... sirve de algo esto?

Joshua M. Epstein

Ver Joshua M. Epstein *Why Model?* Journal of Artificial Societies and Social Simulation vol. 11, no. 4 (2008).

- Por qué modelar?
- Qué podés predecir?
- Otras 16 razones.

Por qué modelar?

Su respuesta al que pregunta es devolverle la pregunta.

Todos tienen un modelo. En general, es implícito, con lo cual su consistencia interna, sus consecuencias lógicas, y la relación con los datos no se conocen.

La idea no es construir modelos, sino modelos *explícitos*: escribir las hipótesis, ver qué generan, modificarlas y ver qué pasa, reducir las variables y comparar su influencia.

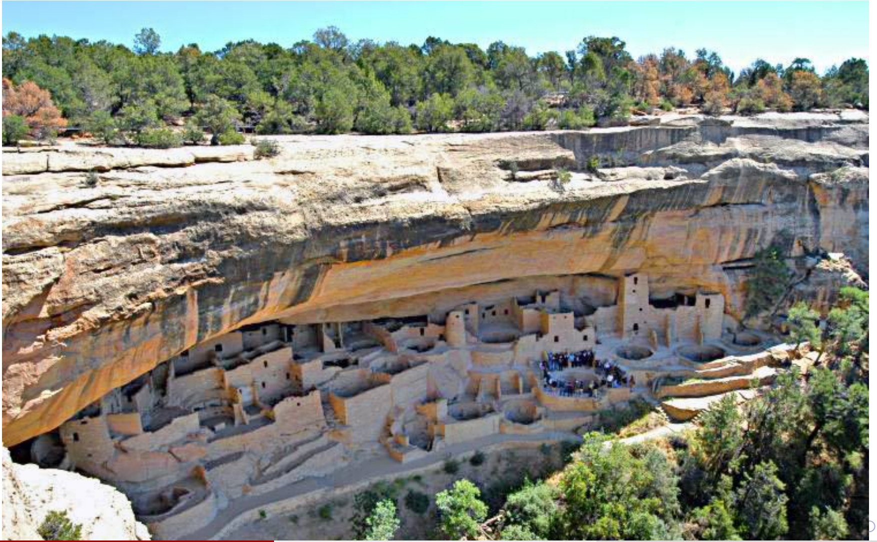
Qué podés predecir?

Es lo de menos: es un objetivo más, y no es el principal.

Otros 16 motivos:

- Explicar.
- Guiar la recolección de datos.
- ...
- Ignoracia militante (vs. militancia ignorante): modelar refuerza el principal hábito de una mente científica, cuestionar lo que nos dicen que vale, testear el conocimiento científico; nos entrena en cuestionar cómo creemos que funciona el mundo.

Ejemplo I - Anasazi, J. Epstein



Ejemplo I

Table 6. Optimized parameter settings based on the average over 15 runs of the model

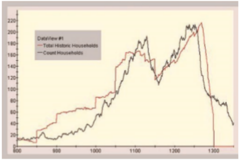
Parameter/norm	L^1, L	L^2
Minimum death age	30	25
Maximum death age	36	38
Minimum age, end of fertility	30	30
Maximum age, end of fertility	32	38
Minimum fission probability	0.125	0.125
Maximum fission probability	0.125	0.125
Average harvest	0.6	0.6
Harvest variance	0.4	0.4

persed across suitable farming habitats located primarily in areas of high potential crop production in the northern part of the valley. The fact that in the real world of Long House Valley, the supportable population chose not to stay behind but to participate in the exodus from the valley indicates the magnitude of sociocultural "push" or "pull" factors that induced them to move (20).[q] Thus, comparing the model results with the actual history helps differentiate external (environmental) from internal (social) determinants of cultural dynamics. It also provides a clue—in the form of the population that could have stayed but elected to go—to the relative magnitude of those determinants.

Discussion

As noted, in these initial inquiries our models include only the most basic environmental and demographic specification, permitting calibration with a minimum number of parameters. Introducing more agent and physical heterogeneity, quite accurate results have been obtained. Richer treatments of household characteristics are possible. For example, in calculating mean household values for size, fissioning, and "death," we have envisioned disaggregating the households into individuals of varying ages in the life course.[r] Similarly, the average caloric values used can be adjusted for age of individuals within the

[q]Dean, J. S., 31st Annual Meeting of the Soc. Am. Archaeol., May 5–7, 1966, Reno, NV.



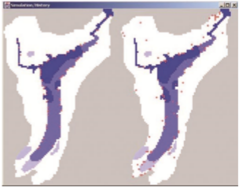


Fig. 3. Simulated and historical settlement patterns, in red, for Long House Valley in A.D. 1125; North is to the top of the page.

household. Nonuniform distributions can be explored. It is, however, interesting that even without implementing these refinements, the output from the current model closely reproduces the record of the archaeological survey.

Issues remain regarding the interpretation of our findings that some inhabitants of Long House Valley could have remained after the archaeologically determined date of abandonment. The fact that environmental conditions may not have been sufficient to drive out the entire population suggests that additional push or pull factors impelled the complete abandonment of the valley after 1300. Another possibility that can be modeled in future simulations might be a combination of environmental, demographic, and epidemiological factors. That is, synergistic interactions between nutritional stress and precolonial epidemic disease might have decimated the population beyond what our model indicates. In addition, the depressed population may simply have been insufficient to maintain cultural institutions, precipitating a collective decision to leave the valley (26). These are ripe topics for future research.

Conclusions

Our model closely reproduces important spatial and demographic features of the Anasazi in Long House Valley from about A.D. 800 to 1300. To "explain" an observed spatiotemporal history is to specify agents that generate—or grow—this history. By this criterion, our strictly environmental account of the evolution of this society during this period goes a long way toward *explaining* this history.

[r]Using reasonable estimations based on model life tables (24) and fertility schedules (25) for horticultural subsistence populations would create a reasonable set of propensities, or probabilities, that can be used in future simulations.

We thank Samuel Bowles and Colloquium participants for valuable suggestions. We gratefully acknowledge financial support from the National Science Foundation (IIS-9820872), the John D. and Catherine T. MacArthur Foundation, the Alex C. Walker Foundation, the National

Ejemplo II - Robert Smith?

Mathematical Modelling of Zombies

Ejemplo II

remaining states and DC. The mean per-person zombie reference rate is 21.4 with a standard deviation of 13.6 across states. The range was lowest (5.4) in Pennsylvania and highest in Alaska.

Age-adjusted death rate per 100,000. A higher death rate per unit population means more potential per-person zombies. Duh. Even so, a higher relative death rate may also indicate the presence of zombies who are more angry or vindictive than usual due to increased and competitive zombie populations. In a future article, we will explore the economics of unemployment rates on zombies.

Percentage of the state population voting for George Bush in 2004. It was not clear whether the presence of zombies is related to conservative political thinking or a reaction to it. This measure is related to general attitudes against taxes and the power of a centralized federal government. Witness the current rise in the US Tea Party, many of whom are zombies.[4] Zombies do not have the right to vote in the United States[5] although many long-dead individuals appear to have this privilege and regularly make use of it, especially in Chicago.

Per-capita shopping malls. When times get tough, zombies go shopping. Or so we thought initially. This is more likely a measure of income and suburban

Ejemplo II

Figure 9.5: A case of flight. A survivor runs away from a zombie (ceasing to be its neighbour) and reconnects to a random new neighbour, here another survivor.

either a survivor ($S_{m,n} \to S_{m+1,n-1}$) or a zombie ($S_{m,n} \to S_{m,n}$). As in Section 9.3.3, this causes 'side effect transitions' both to the zombie he flees from and to that zombie's new neighbour. The new neighbour will be in a state whose probability is directly proportional to the population of that state.

The derivation of the system of ODEs is left as an exercise. Here we provide only the final result:

$$\begin{aligned}\frac{d}{dt}S_{m,n} = &\left(h\langle z\rangle_{z,s}+1\right)\langle z\rangle_{s,s}\alpha\left[(m+1)S_{m+1,n-1}-mS_{m,n}\right]\\ &+\left(h\langle s\rangle_{s,z}+1\right)\left(\langle s\rangle_{s,z}-1\right)\beta\left[(n+1)S_{m,n+1}-nS_{m,n}\right]\\ &+(1+hm)\,\beta\left[(n+1)S_{m,n+1}-nS_{m,n}\right]-\left(h\langle z\rangle_{z,s}+1\right)n\alpha S_{m,n}\\ &+\gamma\frac{S}{S+Z}\left[(n+1)S_{m-1,n+1}-nS_{m,n}\right]+\gamma\langle z\rangle_{s}\frac{S_{m-1,n}-S_{m,n}}{S+Z}\end{aligned} \tag{9.4}$$

Parte IV

Aprobación de la materia

Cronograma

- 9/8 al 3/11: clases normalmente
- 7/11-11/11: SoFiA

  http://wp.df.uba.ar/sociofis/
- Ultimas dos semanas: consultas, repartir temas de final, posiblemente haya alguna charla invitada.

Cómo se aprueban los TPs?

Analizar tres (cuatro) modelos, programarlos, escribir un breve informe:

- Principios de septiembre, mariposas.
- Mediados de octubre, juegos evolutivos (Schelling, Dilema del prisionero, menor oferta única, evolución de normas).
- Fines de noviembre, modelos de opinión.

Cómo se aprueban los TPs?

Cada informe debe tener como máximo 6 páginas y tendrá:

1. una descripción del problema,
2. la explicación de un programa que lo simule (adjuntarlo por mail),
3. 3-5 gráficos con los principales resultados y su descripción.

Los informes se pueden hacer en grupos de 2 ó 3, y no necesariamente deben ser siempre los mismos autores.

Cómo se aprueba el final?

- Exponer un paper: elegir un trabajo, entenderlo, reproducir las simulaciones [o simularlo, si era teórico], y contarlo.
- Escribir un paper: elegir un modelo, analizar la literatura, pensar una variante, programarlo, obtener las ecuaciones.

Durante el curso voy a contar algunas cosas que se pueden intentar, si queda algo publicable, mejor!

Tarea para el hogar

Programar el modelo de las mariposas.

Comparar con alguna de las variantes siguientes:

1.- Considerar que con probabilidad p mantienen la dirección que traían del paso anterior (recuerden asignar una posición y una dirección inicial). Suponer que salen todas de la misma casilla, y analizar ahora el área barrida por el grupo de mariposas entre la casilla de salida y el final del recorrido (cuántas van ocupando en cada tiempo). ¿Cómo es el comportamiento del área en función de p? ¿Cómo dependen las longitudes de los caminos en función de p?

Tarea para el hogar

2.- Combinar lo anterior con el modelo clásico.
Considerar que con probabilidad p mantienen la dirección que traían del paso anterior y con probabilidad q se mueven al azar. ¿Qué pasa cuando barren los valores de p y q, manteniendo siempre $p + q \leq 1$?

Tarea para el hogar

3.- En el modelo clásico, cuando llegan a las posiciones finales, armar parejas y eliminar las que no consigan reproducirse. Reiniciar la simulación con las que quedaron, en las posiciones asignadas originalmente, hasta que se extingan todas. Graficar el tiempo de extinción respecto a q.

Tarea para el hogar

4.- Asignar q diferentes a cada mariposa. Cuando dos se estacionan en un máximo local, se las remueve y se reemplazan por dos nuevas, cada una con un q elegido uniforme entre el del padre y la madre. Las mariposas que no se aparean son eliminadas, pero para conservar la cantidad introducimos al azar nuevas con un q elegido al azar entre el mínimo y el máximo de los de la población actual, o en el intervalo $(0, 1)$. ¿Converge a algo el q de la población? ¿El q promedio?

(Ayuda)

- Ojo con los detalles: buscar en cada casilla cuál es el máximo consume tiempo, conviene hacerlo una sola vez y guardar la info.
- Se puede guardar el espacio como un vector V de $M \times M$ lugares, y en cada uno indicar a cuál se va si busca el más alto (ej: el lugar i tiene el número j).
- Las mariposas son una matriz de $N \times 2 + T$, en la primera fila tienen un número del 1 al $M \times M$, en la segunda el q (si hace falta, si todos tienen el mismo, no). En los T restantes guardan los lugares a los que se fue moviendo en cada paso del algoritmo.
- Al sortear cómo se mueve una mariposa parada en i, va a $j = V(i)$, o con una uniforme van a $i - N - 1$, $i - N$, $i - N + 1$, $i - 1$, $i + 1$, $i + N - 1$, $i + N$, $i + N + 1$, con bordes periódicos, por ejemplo.
- Si lo quieren graficar, pueden cambiar V por una matriz (y agreguen otra coordenada a la matriz de las mariposas).

Suficiente por hoy, nos vemos la próxima!