

Modelado y Simulación de Sistemas Complejos con Aplicaciones en Economía

Clase 4
Leyes de Potencia

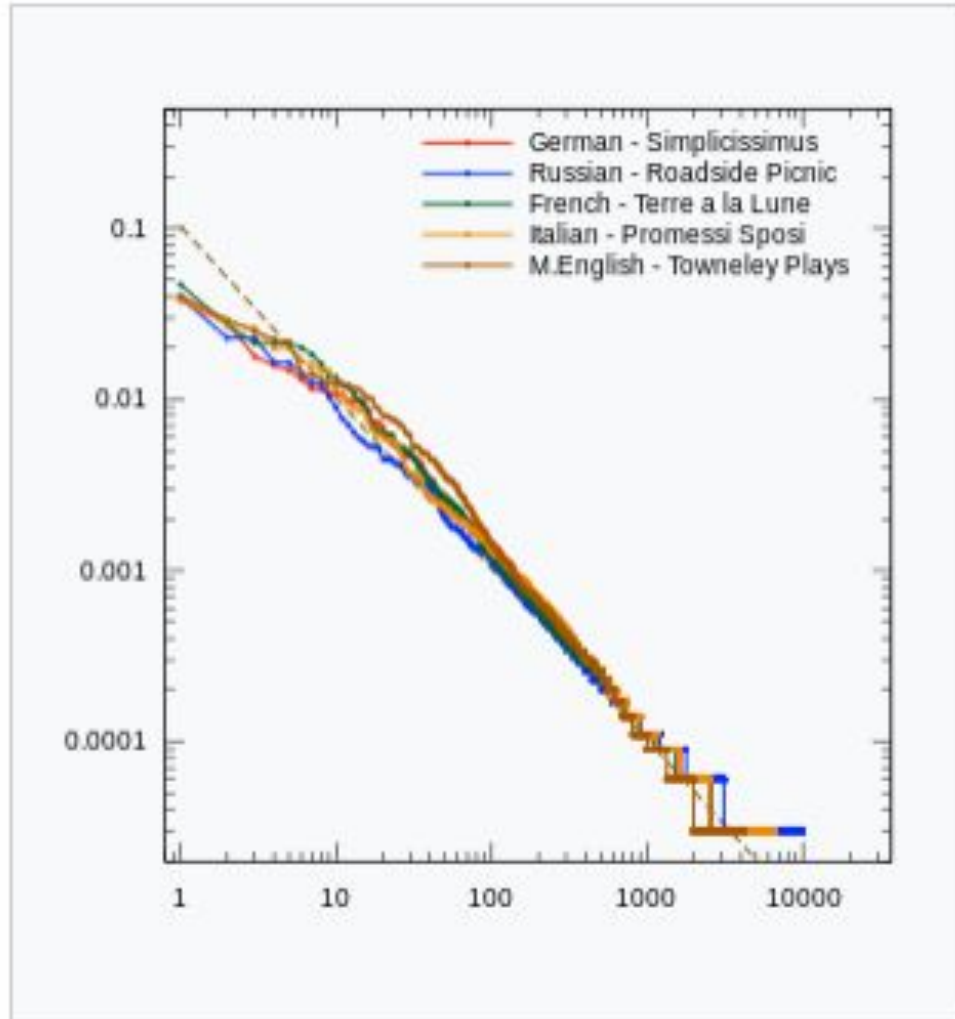
Contenido de la clase 4

- Introducción
- Leyes de Potencia
- Mercados al azar. Vuelos de Levy
- Criticalidad Autoorganizada

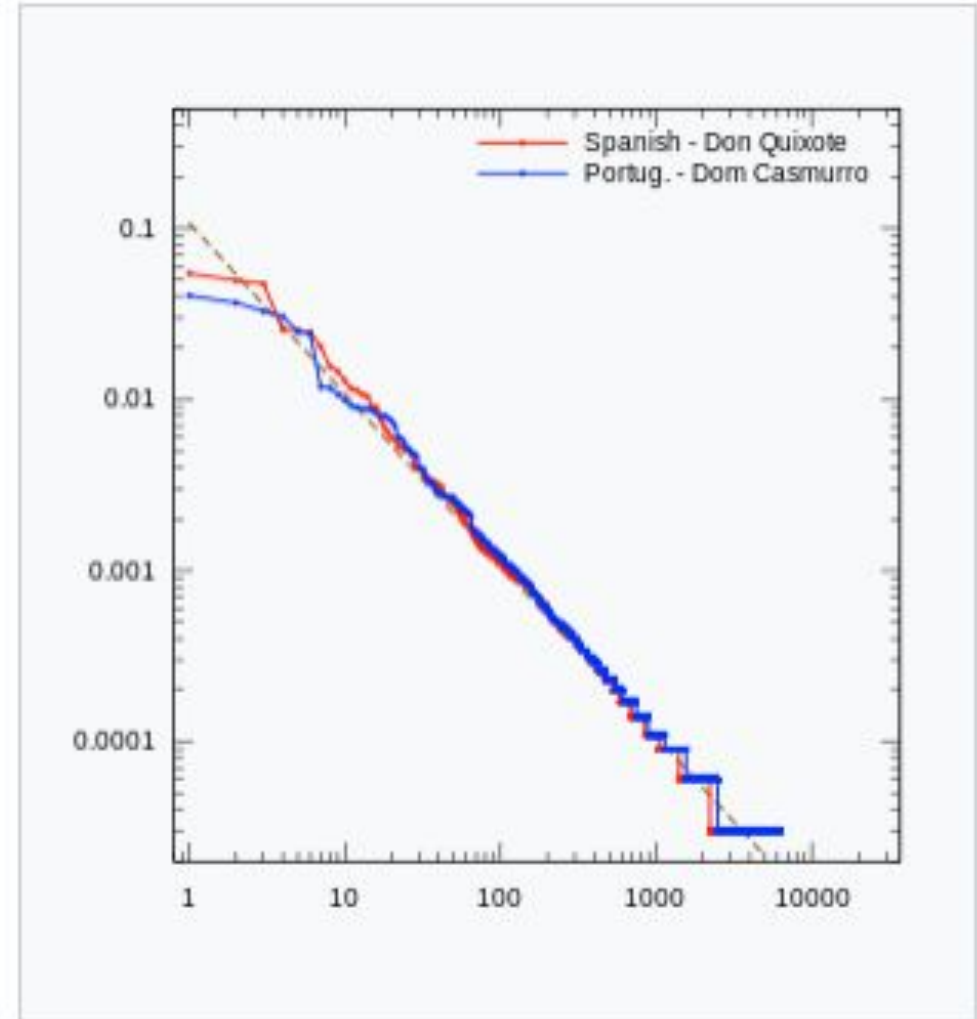
Leyes de Potencia

En la década del 40, el lingüista George Zipf descubrió una propiedad curiosa de como se comporta la frecuencia de aparición de las palabras dentro de una gran cantidad de textos del idioma inglés. Encontró que la palabra más frecuente es “the” con una frecuencia de aparición de 7%, mientras que la segunda es “of”, con una frecuencia de 3.5%, y la tercera más frecuente es . La frecuencia de aparición f es inversamente proporcional a su ranking n : $f(n) \simeq 1/n$, conocida ahora como Ley de Zipf. Lo que interesante es que la misma ley ocurría con en otros lenguajes, marcando un principio de universalidad que apunta a la forma en que nuestro cerebro en última instancia hace evolucionar al lenguaje.

Zipf's law plots for several languages



Texts in [German](#) (1669), [Russian](#) (1972), [French](#) (1865), [Italian](#) (1840), and Medieval English (1460)

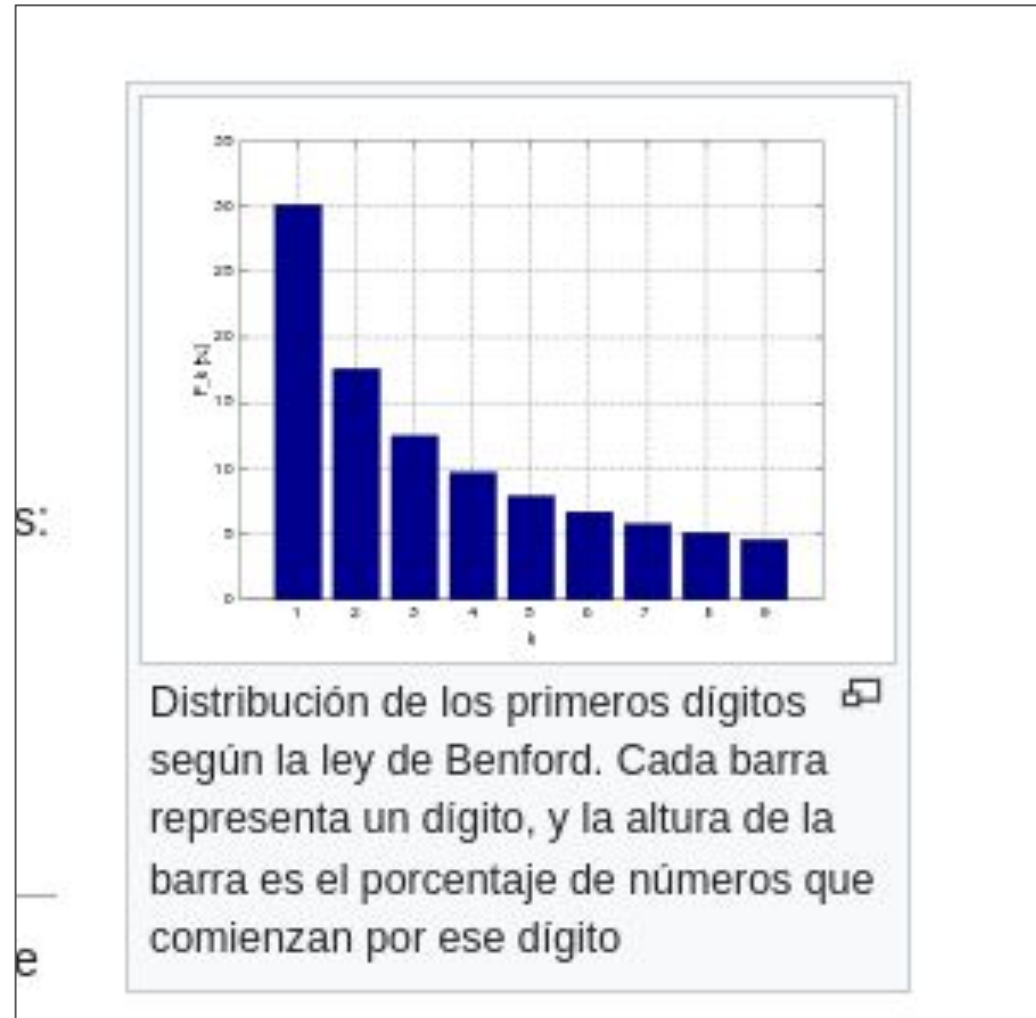


[Cervantes' Don Quixote Part I](#) ([Spanish](#), 1605) and [Assis's Dom Casmurro](#) ([Portuguese](#), 1899)

Leyes de Potencia

En 1881 el astrónomo Simon Newcomb encontró que las primeras páginas de los libros con tablas de logaritmos, estaban más usadas que las últimas. En 1938 el matemático Benford analizó una gran cantidad de números que utilizamos a diarios (registros de pagos de impuestos, alturas de calles, etc) y descubrió que el dígito 1 aparece como primer dígito en 30% de los casos, mientras que el número 2 aparece 18%, el 3 aparece 12%, el 4 aparece 9%, etc. Esta ley se conoce ahora como Ley de Benford y se la ha utilizado para encontrar fraudes contabilidad, y es aceptado como una herramienta legal en los juicios de Estados Unidos.

Ley de Benford o la ley del primer dígito

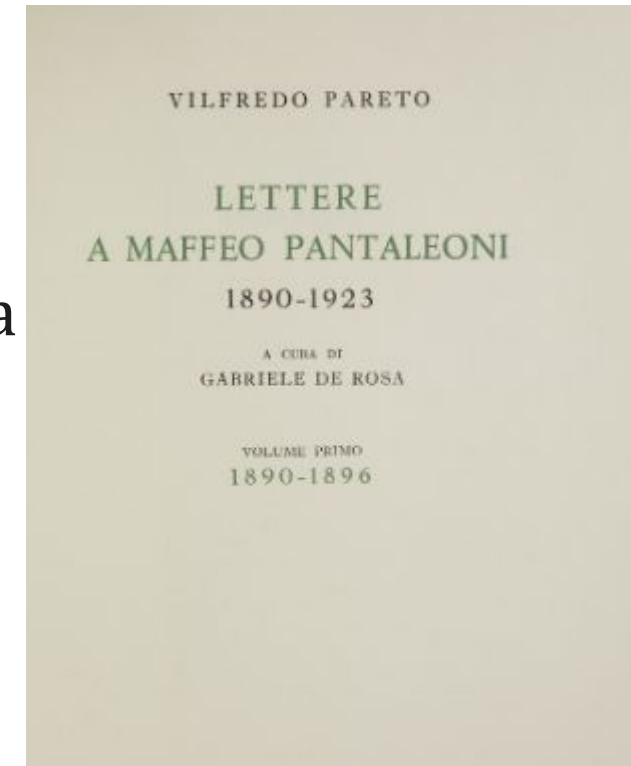


Leyes de Potencia

En 1806, Vilfredo Pareto analizó los ingresos de muchas sociedades, y descubrió que la frecuencia ingresos podía ser resumida como inversamente proporcional a los ingresos.

Lo interesante de estas distribuciones, es que son marcadamente distintas a la distribución normal o Gaussiana, y ocurren en una gran cantidad de fenómenos. Esto permitió que se indague sobre las condiciones por las cuales distribuciones distintas a la distribución Gaussiana, puedan aparecer. En este sentido Levy tuvo un papel importante en discutir sobre la 'estabilidad' de las distribuciones, introduciendo las distribuciones de Levy.

- **El Principio de Pareto o la “regla del 80-20”:** establece que aproximadamente el 80% de los efectos proviene del 20% de las causas
- Originalmente: esta distribución se utilizó para describir la asignación de la riqueza entre los individuos, ya que parecía mostrar bastante bien la forma en que una mayor parte de la riqueza de cualquier sociedad es propiedad de un porcentaje menor de las personas de esa sociedad. También lo utilizó para describir la distribución de la renta: el 20% de la población controla el 80% de la riqueza.⁵



Existen numerosos ejemplos de distribuciones que siguen leyes de potencia:

La Ley de Zipf: Dado un texto de lenguaje natural, la frecuencia de cualquier palabra es inversamente proporcional a su rango en una tabla de frecuencias de aparición o sea la frecuencia

$$f(n) \propto n^{-\alpha} \quad \text{con} \quad \alpha \approx 1$$

Distribución de Pareto: establece que la distribución de individuos según su ingreso sigue una ley de potencias con un exponente ~ -2.5

Ley de Mandelbrot de precios del algodón estableció que grandes fluctuaciones de precio ocurren con mayor frecuencia que la prevista en un paseo al azar.

Tamaño de ciudades: siguen una ley de potencias con un exponente ~ -2.3

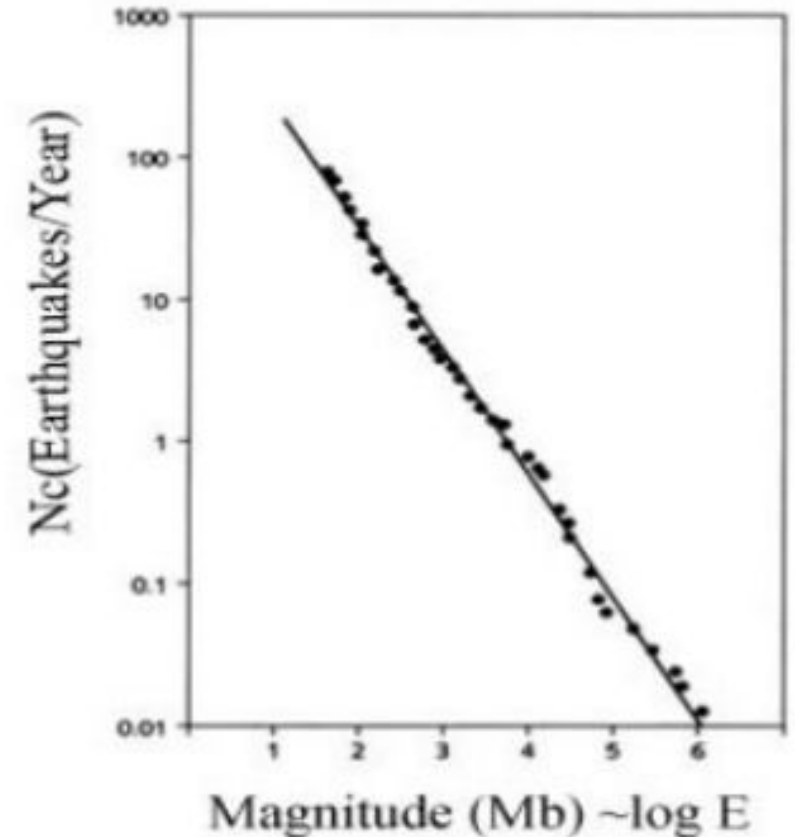
La ley de Gutenberg-Richter de la magnitud de terremotos

- La **ley de Gutenberg-Richter** establece que las magnitudes de los terremotos se distribuyen exponencialmente:

$$\text{Log}_{10} N(m) = a - bm,$$

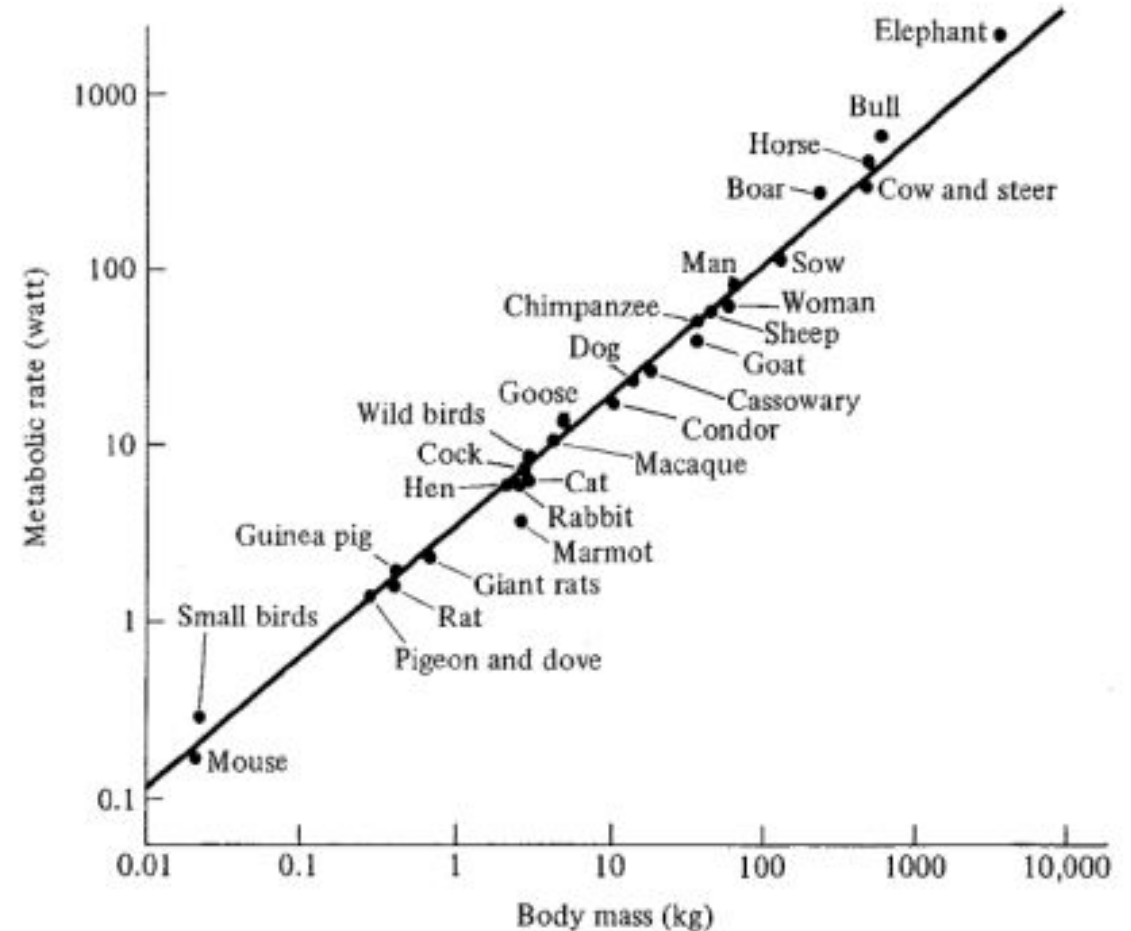
donde $N(m)$ es el número de terremotos con magnitud mayor o igual a m , b es un parámetro de escala y a es una constante.

- *Fuente:* [Gutenberg B., Richter C.F. Frequency of earthquakes in California, Bull. seism. Soc. Am., 1944](#)



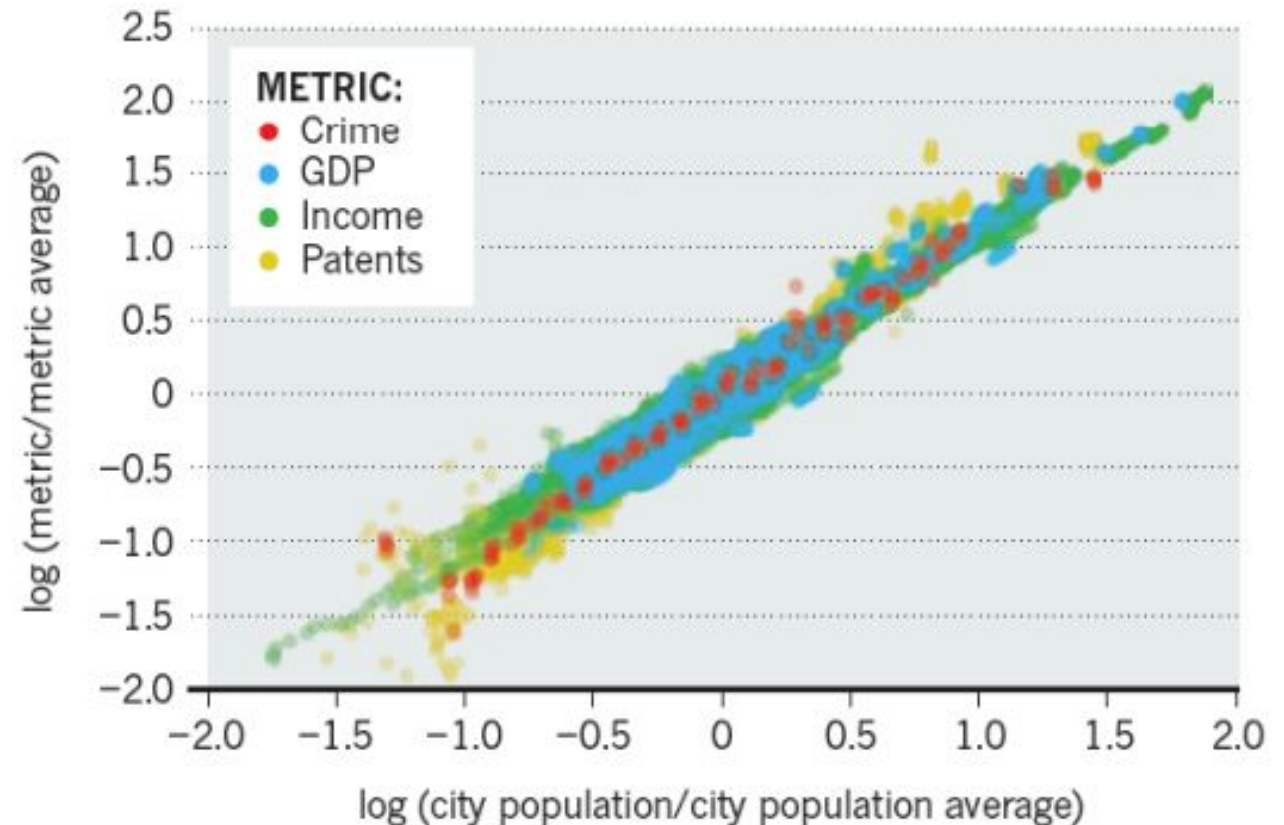
Escalamiento metabólico en animales

- Tasas metabólicas de mamíferos y aves vs sus masas corporales en coordenadas logarítmicas
- Esta línea se conoce como la "curva del ratón al elefante".
- Fuente: [K. Schmidt-Nielsen, Scaling: Why Is Animal Size So Important? Cambridge, 1984](#)



Escalamiento de crimen, ingreso, con la población de ciudades

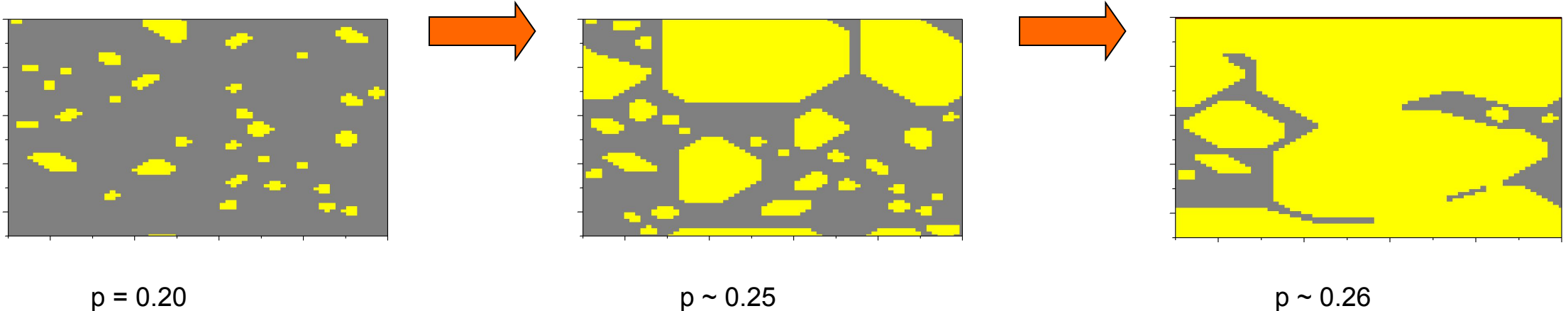
- Los datos de 360 áreas metropolitanas de EE.UU. muestran que parámetros como los salarios y la delincuencia varían en función del tamaño de la población.
- *Fuente:* [L. Bettencourt and G. West, A Unified Theory of Urban Living, Nature, 2010](#)



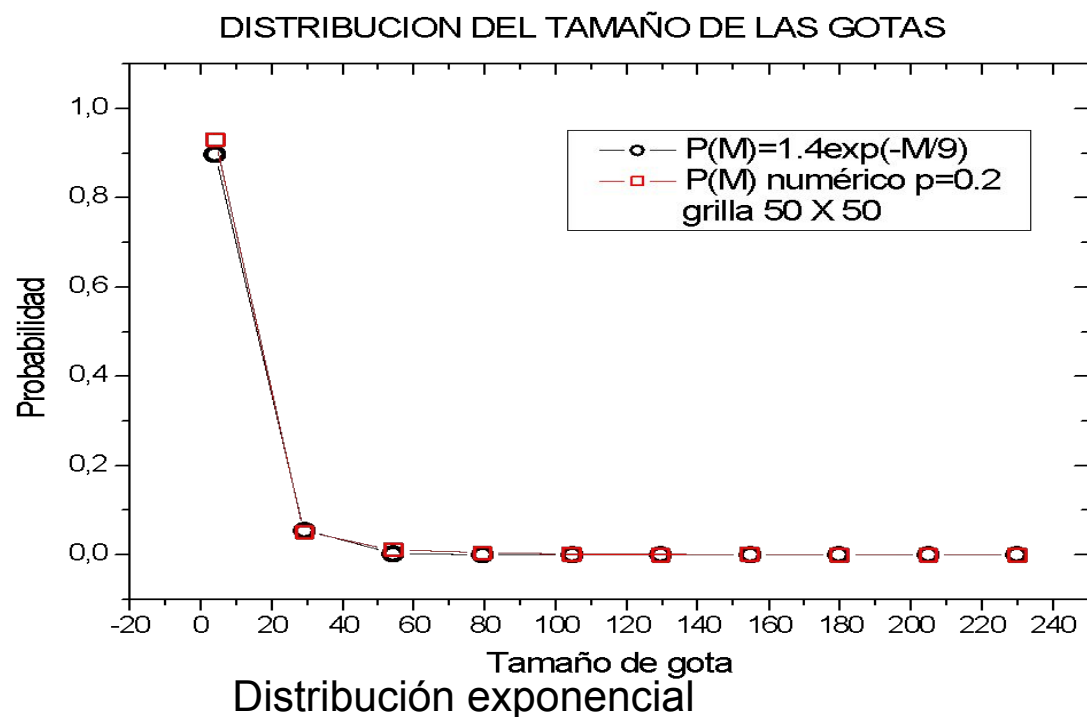
Previamente, del modelo de condensación

- Si se siembran núcleos de condensación con probabilidad p las gotas que se forman pueden cubrir todo el espacio sólo si $p \geq p_{(crítico)} = 0.25$

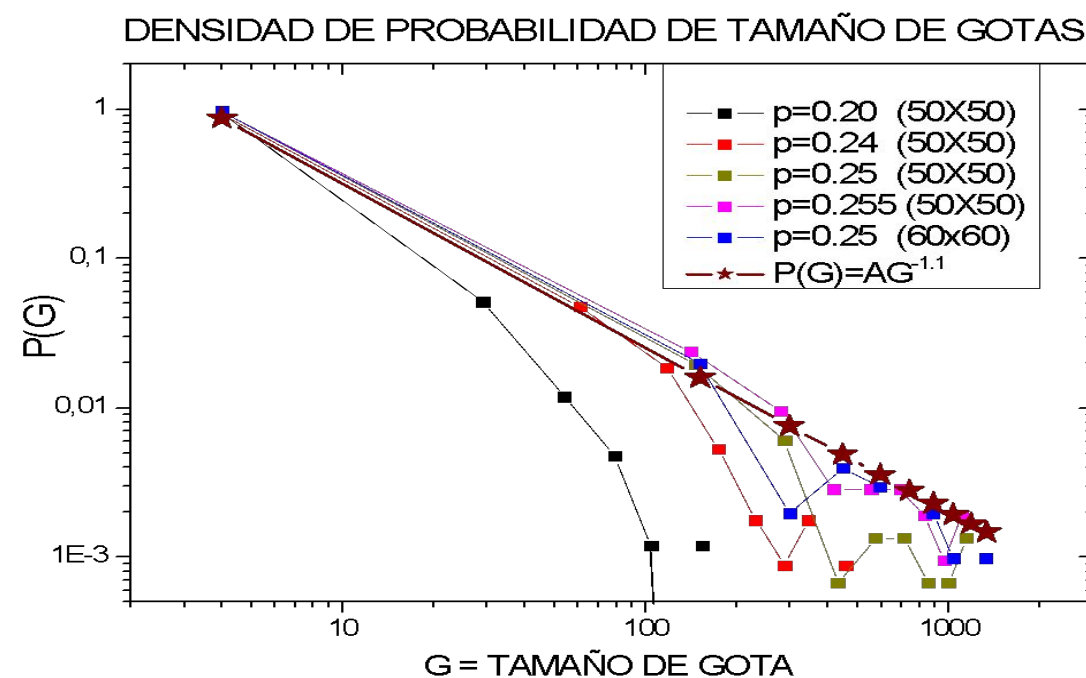
○ *Este ejemplo es una variante del fenómenos de **percolación***



- Los tamaños de las gotas se distribuyen con una distribución exponencial para $p < p_{(crítico)}$ pero en la transición de fase la distribución es una **ley de potencias**



A



B

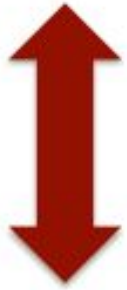
Se puede estudiar cuál es la distribución de tamaño de gotas para cada valor de p .

Se encuentra que esa distribución es exponencial pero cuando p alcanza el valor crítico pasa a ser una ley de potencias con exponente -1.1 . Una distribución así no tiene valor medio finito.

Conclusión: en la transición de fase se pueden encontrar “gotas” de todos los tamaños imaginables

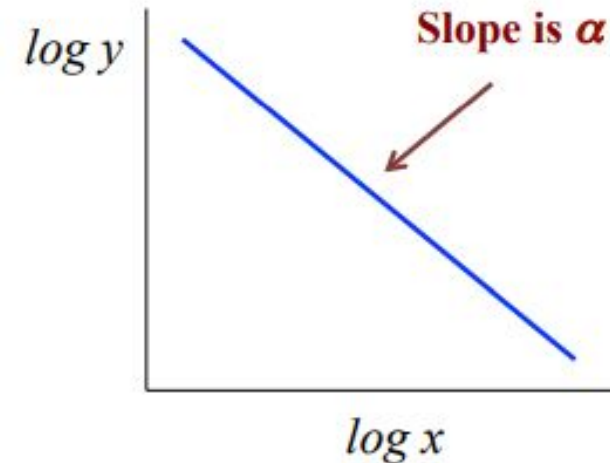
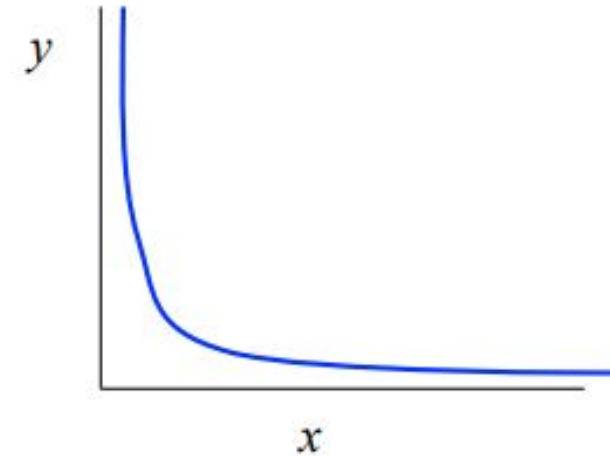
Ley de Potencia

$$y = cx^{\alpha}$$



$$\log y = \alpha \log x + \log c$$

Straight line on a log-log plot



PROPIEDAD DE LAS LEYES DE POTENCIAS

A una ley de potencias se la suele llamar “**libre de escalas**” porque cualesquiera sea la escala con que se observen los eventos, sigue la misma ley:

$$P(x) = Ax^{-\lambda} \Rightarrow P(bx) = Ab^{-\lambda}x^{-\lambda} \Rightarrow P(bx) = g(b)P(x)$$

Esto significa que si se amplía una parte de la distribución, se obtiene la misma forma que la distribución completa.

Ejemplos a tratar

- Crecimiento de firmas
- Mercados financieros
- Fenómenos críticos
- Criticalidad autoorganizada

Estas distribuciones sugieren la vecindad de cambios estructurales y están relacionados con situaciones en que las variables aleatorias siguen distribuciones **libres de escalas** o **autosemejantes**

Crecimiento de firmas

- Un punto de partida adecuado para los estudios sobre la estructura y la dinámica industrial es la distribución del tamaño de las empresas
- La distribución empírica de tamaños propuesta por Robert Gibrat (1931), conocida como la "Ley del efecto proporcional" (o "ley de Gibrat")
- Gibrat observó que la distribución de los establecimientos manufactureros franceses seguía una distribución sesgada que se asemejaba a la lognormal
- Gibrat consideró la aparición de la distribución del tamaño de las empresas como un resultado y quiso ver qué proceso de crecimiento subyacente podía ser responsable de generarla

Modelo de Gibrat (1931)

- i) la tasa de crecimiento de una empresa es independiente de su tamaño
- ii) los crecimientos en períodos sucesivos no están correlacionados
- iii) las firmas **no interactúan** entre sí

El resultado del modelo es que el tamaño de las firmas siguen un paseo al azar

$$S_{t+\Delta t} = S_t (1 + \eta_t) \quad S_t \text{ es el tamaño de la empresa; y } \eta_t \text{ es una variable aleatoria}$$

El tamaño de una empresa en el momento t puede explicarse únicamente en función de su historial idiosincrásico de choques multiplicativos

Y los tamaños sigue una distribución gaussiana

$$p(\log S_t) \approx \exp\left(-\frac{(\log S - \mu)^2}{2\sigma^2}\right)$$

También las tasas de crecimiento

$$R = \frac{S_{t+T}}{S_t}$$

Las empresas son realizaciones independientes de perturbaciones de crecimiento i.i.d. distribuidas normalmente → este proceso estocástico conduce a la aparición de una distribución lognormal del tamaño de las empresas.

Tang, A. (2015)

Does Gibrat's Law hold for Swedish energy firms?.

Empirical Economics, 49(2): 659-674

<http://dx.doi.org/10.1007/s00181-014-0883-x>

Does Gibrat's law hold for Swedish energy firms?

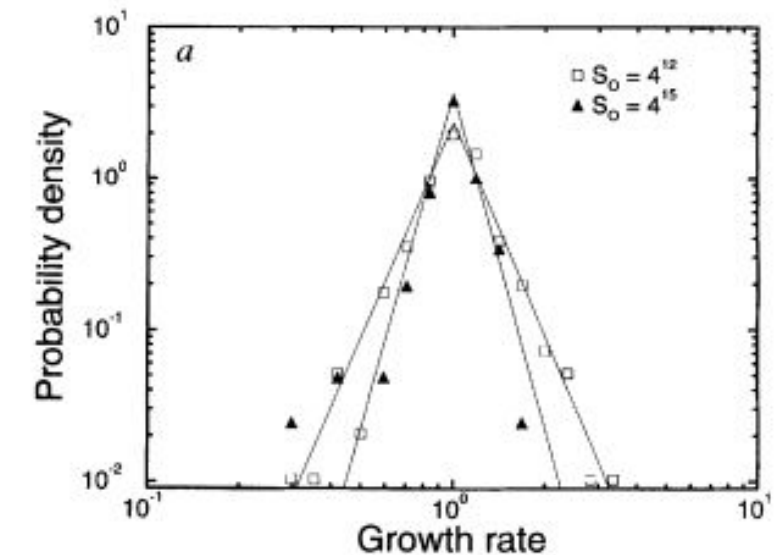
Aili Tang¹

¹ Aili Tang, Sweden, Dalarna University, SE - 791 88 Falun, +46 762 339 893, ata@du.se

Abstract Gibrat's law predicts that firm growth is purely random and should be independent of firm size. We use a random effects–random coefficient model to test whether Gibrat's law holds on average in the studied sample as well as at the individual firm level in the Swedish energy market. No study has yet investigated whether Gibrat's law holds for individual firms, previous studies having instead estimated whether the law holds on average in the samples studied. The present results support the claim that Gibrat's law is more likely to be rejected ex ante when an entire firm population is considered, but more likely to be confirmed ex post after market selection has “cleaned” the original population of firms or when the analysis treats more disaggregated data. From a theoretical perspective, the results are consistent with models based on passive and active learning, indicating a steady state in the firm expansion process and that Gibrat's law is violated in the short term but holds in the long term once firms have reached a steady state. These results indicate that approximately 70% of firms in the Swedish energy sector are in steady state, with only random fluctuations in size around that level over the 15 studied years.

Scaling behaviour in the growth of companies

Michael H. R. Stanley*, Luís A. N. Amaral*,
Sergey V. Buldyrev*, Shlomo Havlin*†,†,
Heiko Leschhorn*, Philipp Maass*,
Michael A. Salinger‡ & H. Eugene Stanley*



Amaral, Stanley et al. analizaron estadísticas de firmas en los EEUU (período 1975-1991, Compustat database) y encontraron:

- La distribución de tamaños de firmas sigue una distribución log-normal
- La distribución de tasas de crecimiento para cada tamaño inicial sigue una ley exponencial que se puede ajustar bien mediante

$$p(r|s_0) = \frac{1}{\sqrt{2}\sigma(s_0)} \exp\left[-\frac{\sqrt{2}|r - \bar{r}(s_0)|}{\sigma(s_0)}\right] \quad (1)$$

S_1 y S_0 son ventas en dos años consecutivos

Es habitual estudiar el crecimiento de las empresas en escalas logarítmicas, por lo que se defina $r \equiv \ln(S_1/S_0)$ y $s_0 \equiv \ln(S_0)$, y se calcula la distribución condicional $p(r/s_0)$ de las tasas de crecimiento con un determinado valor inicial de ventas S_0

Paseos al azar, mercados financieros y distribuciones estables

- Los mercados son sistemas abiertos donde numerosos individuos interactúan relacionando ante la llegada de información determinan colectivamente los precios y las cantidades intercambiadas.
- Una hipótesis generalizada en el estudio de mercados financieros es que los precios responden a un proceso estocástico caracterizado como un paseo al azar.
- Para que el movimiento de los precios de un activo financiero siga un paseo al azar, las transacciones realizadas deberán ser consistentes con lo que se denomina un “fair game”, en el cual la ganancia esperada de un especulador debe ser idénticamente nula.
- La “mejor” predicción del precio de mañana, es simplemente el precio de hoy.
- Una importante consecuencia de esta hipótesis, es que los cambios en precios de los activos no están correlacionados (Bachelier, 1900), lo cual implica que no es posible realizar predicciones sobre la base de algún predictor de los precios futuros basado en la secuencia de los pasados.

Paseos al azar, mercados financieros y distribuciones estables

- Se definen los rendimientos relativos como $r_{t+1} = (p_{t+1} - p_t) / p_t$, y estas se tratan como variables aleatorias independientes
- Se suele suponer que el proceso estocástico es un movimiento Browniano, por lo que, está guiado por una distribución normal
- En lo que sigue veremos ejemplos de distribuciones de mercados que no se ajustan a estas hipótesis de paseos al azar y se ajustan en cambio a leyes de potencias. Algunas distribuciones que guían al proceso generador de rendimientos son distribuciones denominadas estables que dan lugar a “colas anchas” y distribuciones que siguen leyes de potencias

Paseos al azar, mercados financieros y distribuciones estables

Una hipótesis generalizada en el estudio de mercados financieros es que los precios responden a un proceso estocástico caracterizado como un paseo al azar.

Cabe preguntarse si dichos procesos estocásticos son autosemejantes

¿Pasa lo mismo en unas horas que en todo un día?

¿Pasa lo mismo en pocos días que en un mes?

Supongamos un paseo al azar unidimensional, cuyo saltos son independientes e idénticamente distribuidos con una probabilidad $p(\mathbf{x})$. ¿Cuándo la probabilidad $P_n(\mathbf{x})$ que el paseo haya arribado a la posición \mathbf{x} luego de n pasos

$$x = x_1 + \cdots + x_n$$

es la misma que la $p(\mathbf{x})$ a menos de un factor de escala?

¿Puede existir una $p(\mathbf{x})$ que produzca un proceso estocástico con una trayectoria **autosemejante**?

Una hipótesis generalizada en el estudio de mercados financieros es que los precios poseen distribuciones normales.

Una distribución $p(\mathbf{x})$ Gaussiana con media μ y varianza σ cumple ese principio de autosemejanza ya que que $P_n(\mathbf{x})$ tiene media $n\mu$ y varianza $n\sigma$

Sin embargo esa no es la distribución más general que cumple con el requisito de autosemejanza.

En realidad las distribuciones autosemejantes tienen características diferentes y se asocian con los denominados **vuelos de Levy:**

*Una distribución se dice que corresponde a un **vuelo de Levy** o que se trata de una **distribución estable** cuando las sumas parciales de n variables aleatorias se distribuyen todas de la misma manera*

Distribuciones estables

- sean dos variables aleatorias x_1 y x_1' independientes, con la misma distribución de probabilidad $P_1(x)$, y se define $x_2 = x_1 + x_1'$ como una nueva variable aleatoria
- La distribución de probabilidad de esa nueva variable x_2 resulta $P_2(x) = \int dx_1 P_1(x_1) P_1(x - x_1)$, es decir, la probabilidad del evento x surge de la ocurrencia conjunta e independiente de los eventos x_1 y $x - x_1$
- se dice que $P_2(x)$ es la convolución $P_1(x) \star P_1(x)$
- generalizando para la distribución de la suma de N variables de una misma distribución $\rightarrow P_N(x)$ es la convolución de N veces de $P_1(x)$

Distribuciones estables

- Se espera que partiendo de cualquier distribución $P_1(x)$ la convolución sucesiva resultará en una nueva distribución cuya forma funcional no necesariamente será similar a la original
- Sin embargo, una propiedad interesante de algunas distribuciones $P_1(x)$ es que la forma funcional de $P_N(x)$ es igual a la de $P_1(x)$ a menos un factor de localización y escala $P_N(x) = \beta P_1(\gamma x)$
- cuando ocurre esta propiedad se dice que $P_1(x)$ es estable

Forma de una distribución estable

La distribución de una suma de variables aleatorias independientes es el producto de convolución de sus distribuciones:

$$P_n(x) = \int_{-\infty}^{+\infty} P_{n-m}(x-x')P_m(x')dx' \quad 0 \leq m \leq n$$

Con lo que el producto de las funciones características asociadas (transf. de Fourier) de tales distribuciones, satisfacen:

$$P_n(k) = P_{n-m}(k) \cdot P_m(k) \quad \text{con} \quad P(k) = \int_{-\infty}^{+\infty} e^{ikx} P(x)dx$$

Para que se satisfaga esta ecuación, las funciones características de una distribución estable posee la forma:

$P_n(k) = e^{(-n|k|^\beta)}$ (Distribucion estable) . Las distribuciones con funciones características se denominan distribuciones estables y corresponde a un proceso estocástico llamado como vuelo de Lévy y verifican que sumas parciales de n variables aleatorias tienen la misma función de distribución. El exponente β se denomina exponente característico.

Distribuciones estables en mercados financieros

En lo que sigue analizaremos series temporales de las volatilidades en mercados de valores y de divisas y mostraremos que son distribuciones estables

En estos casos los pasos del paseo al azar son las sucesivas cotizaciones con lo que el número de pasos n es equivalente a un intervalo de tiempo Δt

El rendimiento relativo al cabo de Δt períodos de tiempo es es

$$r(t, \Delta t) = \frac{p(t) - p(t - \Delta t)}{p(t - \Delta t)}$$

$r(t)$ es una variable aleatoria.

Veremos que su distribución de probabilidad $P(r, \Delta t)$ de ocurrencia de un valor r para un dado Δt resulta ser estable.

El valor de n está aquí relacionado con Δt . Explicitando el intervalo de tiempo Δt de cada paso del paseo al azar, la función característica de una distribución estable se puede escribir

$$P(r, n\Delta t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-n\Delta t |k|^\beta} e^{-ikr} dk = \frac{1}{\pi} \int_0^{+\infty} e^{-(n\Delta t)k^\beta} \cos(kr) dk$$

con $P_n(r) \equiv P(r, n\Delta t)$

Si se hace el cambio de variables $r_s = \frac{r}{(\Delta t)^{1/\beta}}$ se puede mostrar que

$$P_s(r_s, n\Delta t) = \frac{P(r, n\Delta t)}{(\Delta t)^{-1/\beta}}$$

La variable r_s es un *rendimiento estándar o renormalizado*

Además es posible calcular la *probabilidad de retorno al origen* que resulta

$$P_n(0, \Delta t) = \frac{1}{\pi} \int_0^{\infty} e^{-(n\Delta t)k^\beta} dk = \frac{\Gamma(1/\beta)}{\pi \beta (n\Delta t)^{1/\beta}}$$

La probabilidad de retorno al origen se comporta como una ley de potencia cuyo exponente está determinado por el exponente característico de la distribución estable

De lo dicho surge además que las distribuciones para distintos valores de n son auto-similares si se corrige sólo un factor de escala.

Por consiguiente se pueden superponer las distribuciones para diferentes Δt corrigiéndolas sólo por un factor relacionado con su escala temporal

$$P_s(r_s, n\Delta t) = \frac{P(r, n\Delta t)}{(\Delta t)^{-1/\beta}}$$

y utilizando la variable renormalizada

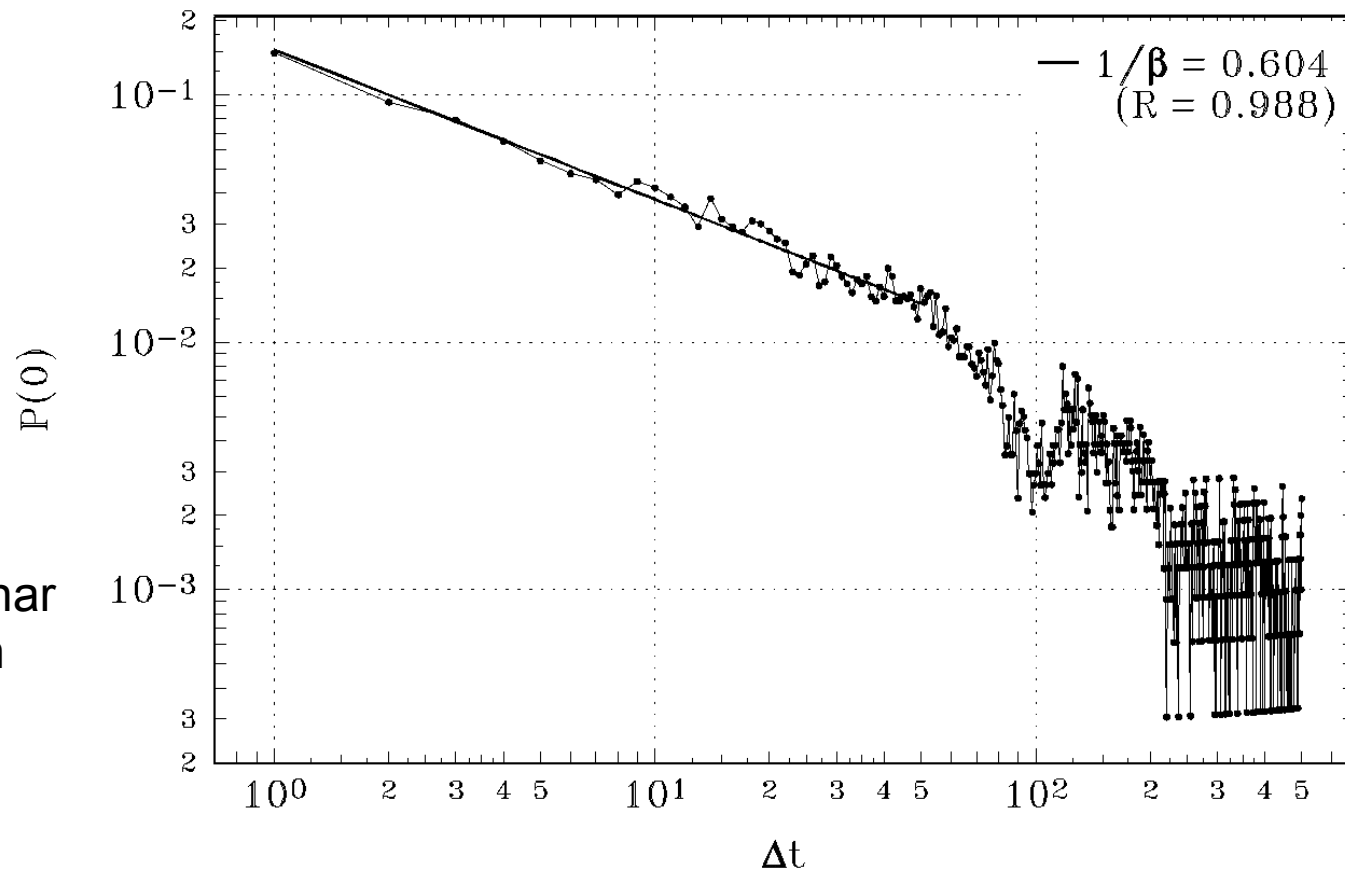
$$r_s = \frac{r}{(\Delta t)^{1/\beta}}$$

El proceso estocástico estable es por consiguiente, invariante frente a transformaciones de escala temporal

La evolución del tipo de cambio de la Libra Esterlina contra el Dólar Americano

Para estimar el exponente característico se obtuvo primero la probabilidad de retorno al origen $P(r = 0; \Delta t)$ para diversos intervalos Δt , en una escala log-log.

Los datos pueden ajustarse con una recta cuya pendiente es - 0,604. Hasta los primeros 50 valores de Δt el ajuste da un coeficiente de correlación de 0.988. La pendiente permite determinar el exponente característico que corresponde con un comportamiento no gaussiano - ya que $\beta = 0.60$



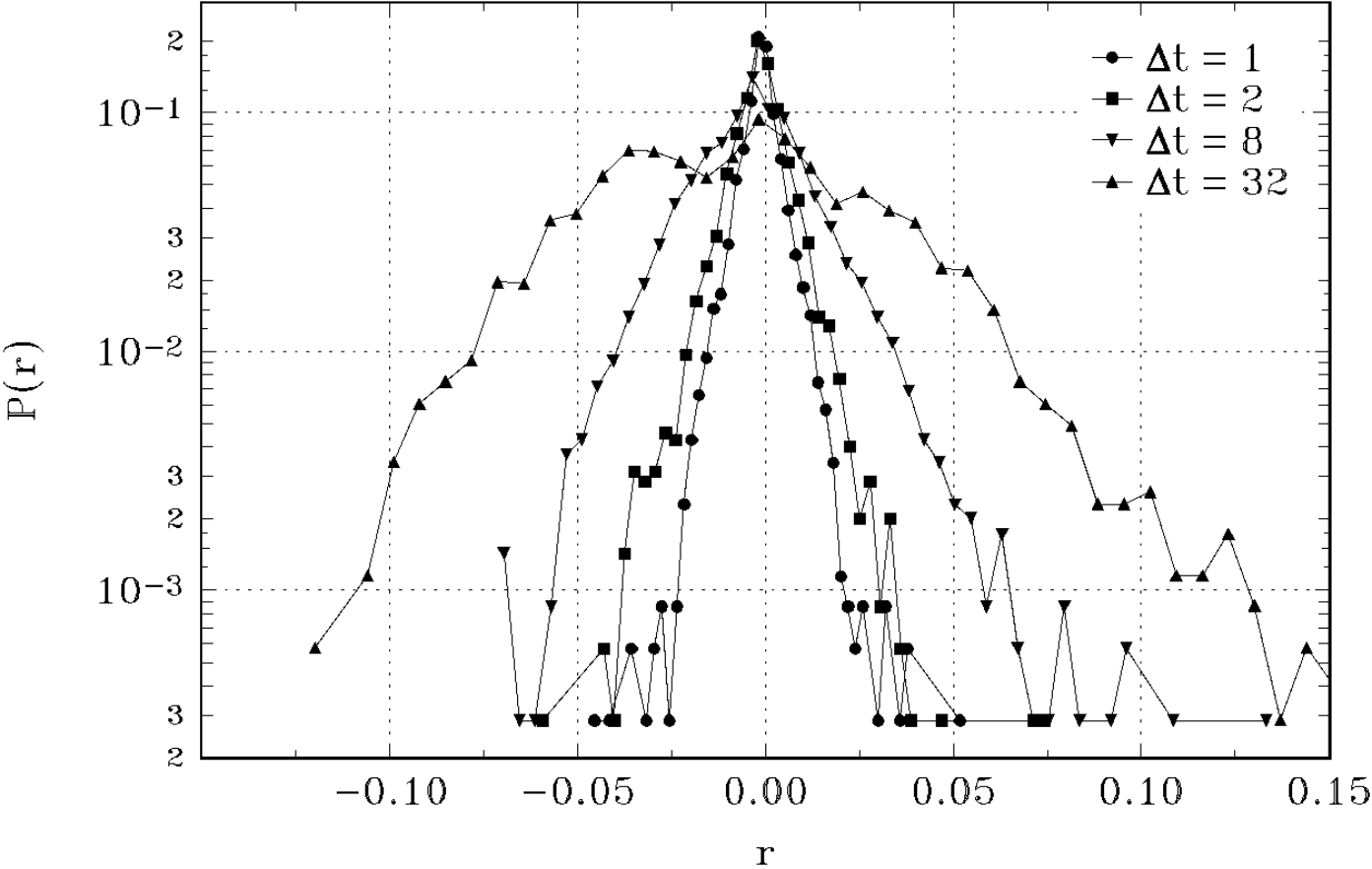
La serie de datos que se utilizó para esta distribución está formada por el tipo de cambio diario Libra Esterlina contra el Dólar Americano al cierre y corresponde al período comprendido entre el 1ro. de junio de 1973 y el 21 de mayo de 1987 y está conformada por una total de 3511 datos.

La evolución del tipo de cambio de la Libra Esterlina contra el Dólar Americano

Con la misma serie temporal se pueden calcular las probabilidades $P(r, \Delta t)$ para diversos intervalos Δt medidos en días 1, 2, 8 y 32.

Se obtiene las siguientes distribuciones.

se puede observar estas distribuciones son visiblemente simétricas y sus colas se despliegan al crecer el intervalo Δt , ya que al elegir un mayor intervalo sobre el cual se calculan los rendimientos relativos, crece la probabilidad de obtener grandes rendimientos (positivos o negativos).

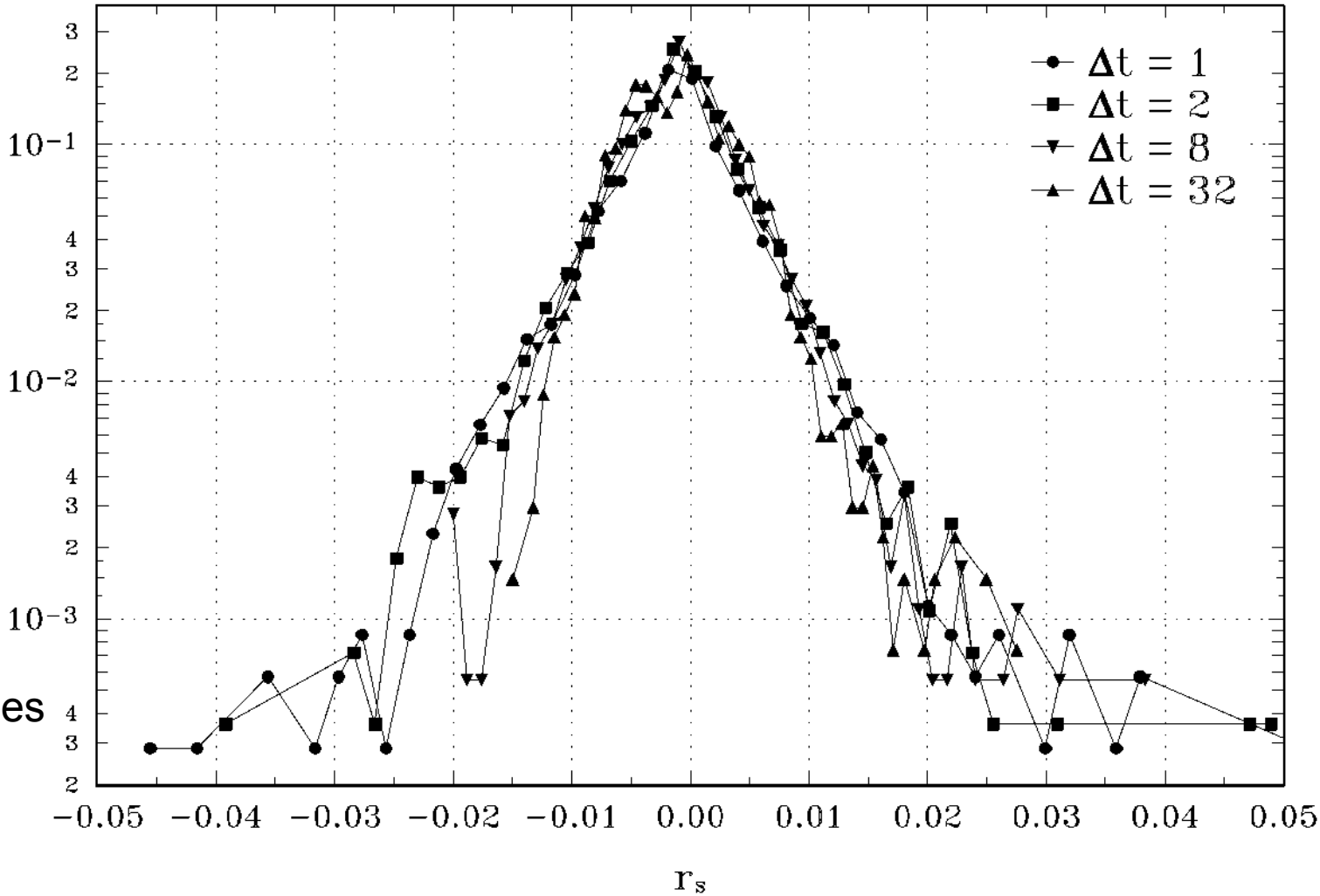


La evolución del tipo de cambio de la Libra Esterlina contra el Dólar Americano

Las distribuciones anteriores pueden hacerse coincidir si se utiliza el rendimiento renormalizado r_s y se utiliza la distribución corregida P_s

Con el exponente característico obtenido de la distribución $P(r, \Delta t)$ se puede corroborar que las distintas distribuciones son similares salvo un factor de escala. Para verificar esta propiedad, se debe re - escalar todas las distribuciones para diferentes Δt , utilizando la variable r_s

salvo eventos de gran amplitud, las distribuciones escaladas se superponen en una unica distribucion.



Mercado		Período - No. Datos		Exponente β	Mercado		Período - No. Datos		Exponente β
L.E. – D.A.		6/73-5/87 – 3511		0.60 ± 0.02	M.A. – D.A. (1)		6/73 – 6/87 –2409086		0.591 ± 0.004
Y.J. – D.A.		6/73-5/87 – 3511		0.59 ± 0.02	ORO – D.A. (1)		6/73 – 6/87 – 652194		0.580 ± 0.004
D.C. – D.A.		6/73-5/87 – 3511		0.59 ± 0.02	Dow Jones		1/00-6/93 – 25770		0.583 ± 0.007
M.A.– D.A.		6/73-5/87 – 3511		0.651 ± 0.03	S & P 500 (2)		7/62-12/95 – 8431		$0.588 \pm \text{¿?}$
F.S – D.A.		6/73-5/87 – 3511		0.60 ± 0.02	IBM (2)		7/62-12/95 – 8431		$0.561 \pm \text{¿?}$
L.E. – D.A. (1)		6/73 – 6/87 – 1,174,941		0.581 ± 0.004	AMPCO (2)		7/62-12/95 – 8431		$0.621 \pm \text{¿?}$
Y.J. – D.A. (1)		6/73 – 6/87 – 1,477,992		0.591 ± 0.003	DAX (3)		11/86-9/92 - 1452		0.54 ± 0.08
F.S. – D.A. (1)		6/73 – 6/87 –1,214,901		0.594 ± 0.004	DA:dólar americano; LE:libra esterlina; YJ: yen japonés MA marco alemán; DC Dólar canadiense;FS: F. suizo DAX: índice mercado bursátil de Frankfurt				

Nota: El exponente característico en mercados de divisas es ~.60 y en mercados de valores es menor ~.56.

MERCADO CAMBIARIO ARGENTINO

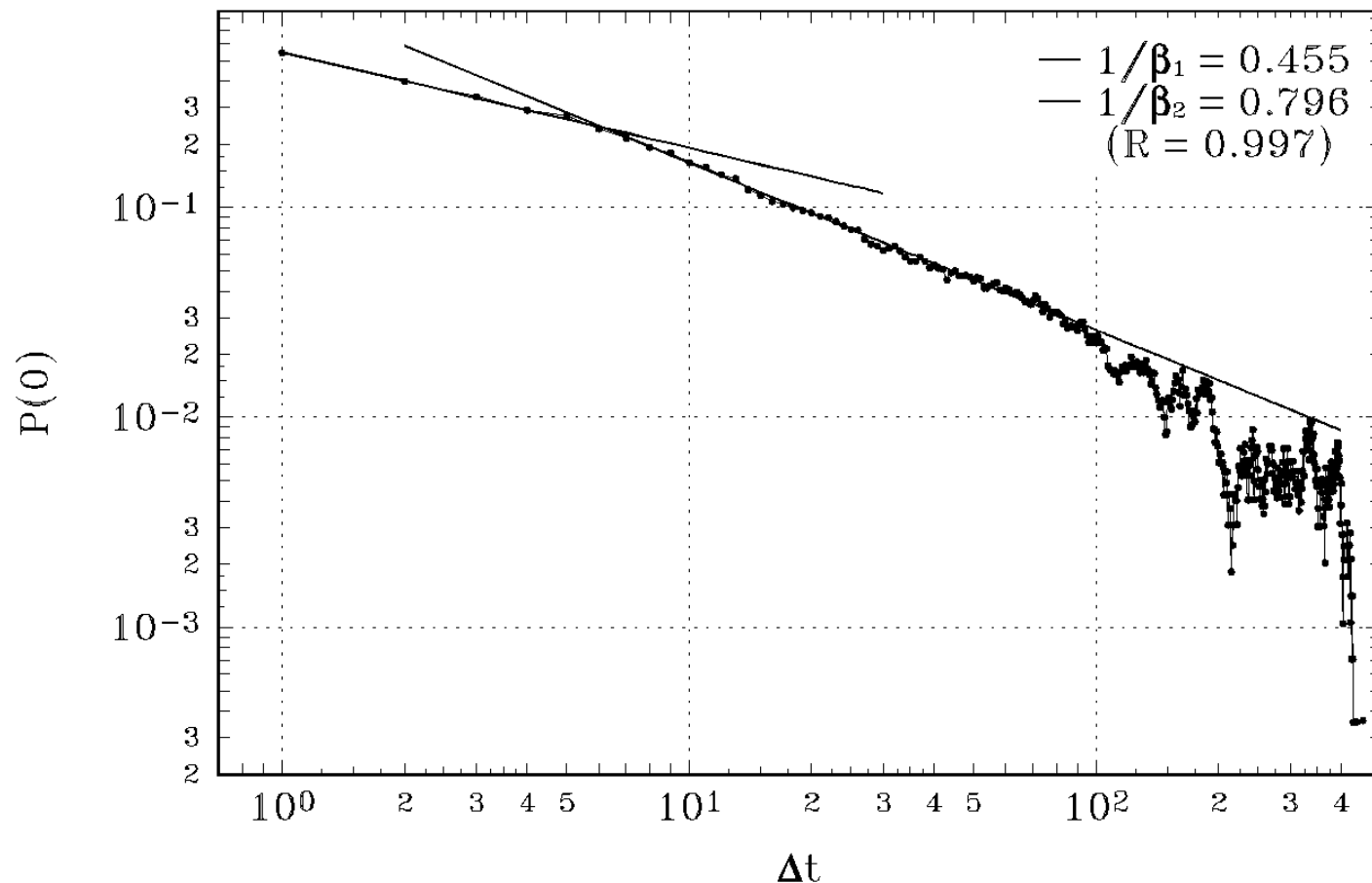
Se puede hacer la experiencia de aplicar este cálculo al mercado cambiario argentino en el período 1871-1981 con datos diarios, recopilado según la siguiente tabla

Paralelo	2/8/71-30/12/75;4/1/82-9/10/87;(19/5/89-14/12/89)
Libre	(7/1/76-17/6/81);15/10/87-18/5/89;15/12/89-30/3/91
Finan-ciero	18/6/81-23/12/81

La ley de potencias que sigue la probabilidad de retorno al origen se quiebra

MERCADO CAMBIARIO ARGENTINO

La probabilidad de retorno al origen.
No queda representada por una única ley de potencias sino que se requieren dos rectas de distintas pendientes para ajustar la distribución de probabilidad.
Una de ellas es válida para periodos breves, de $t < 8$ días y la otra abarca todos los intervalos de tiempo más prolongados.

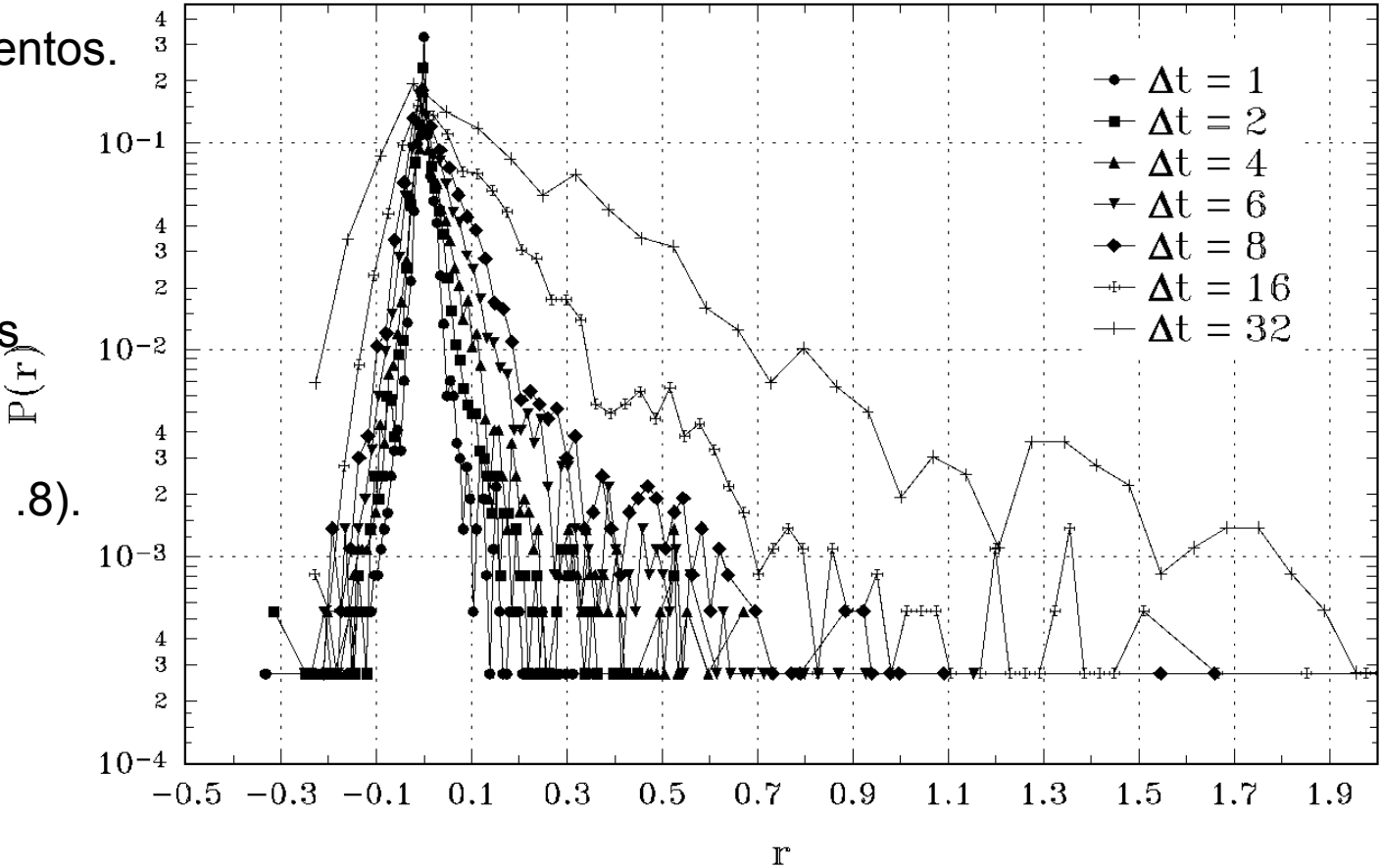


Es con todo llamativo que se pueda ajustar con dos rectas de distinta pendiente válidas para períodos breves ($\Delta t < 8$) y prolongados ($\Delta t > 8$)

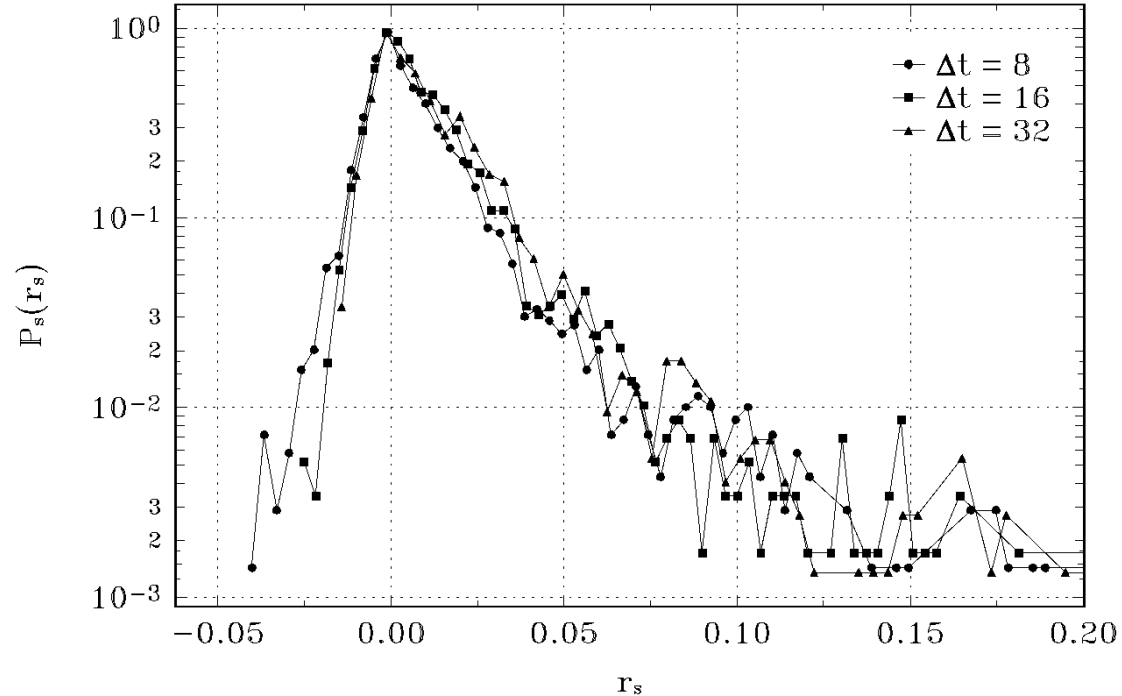
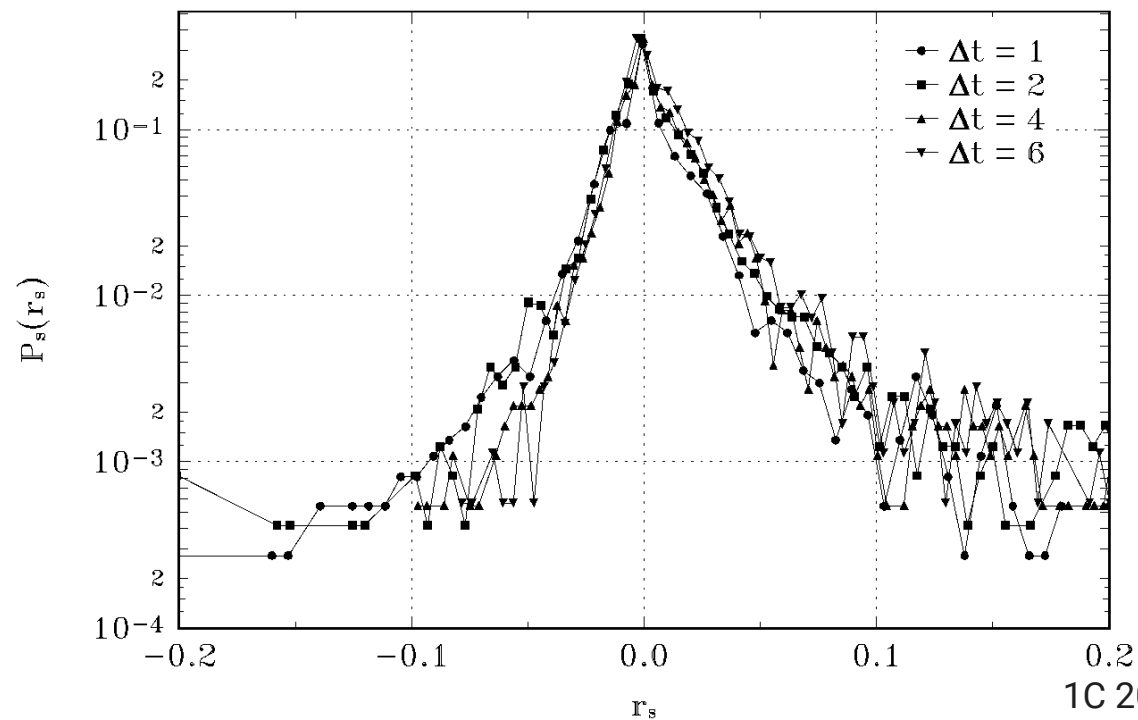
MERCADO CAMBIARIO ARGENTINO

Las correspondientes distribuciones de rendimientos.

Se obtienen distribuciones asimétricas por efectos de la inflación.
El régimen inflacionario se traduce en una probabilidad de retorno al origen que decae más rápidamente para períodos más prolongados (exponente crítico para el mercado dólar-libra que es .6 con el del mercado peso-dólar que es .8).



Cuando se trata de re-escalar estas distribuciones para comprobar su invariancia resulta imposible. Se utilizan las dos pendientes diferentes



Con todo es posible encontrar un alto grado de invariancia de escala en las distribuciones correspondientes a períodos prolongados y breves por separado



Zipf Distribution of U.S. Firm Sizes

Robert L. Axtell

Science **293**, 1818 (2001);

DOI: 10.1126/science.1062081

Zipf Distribution of U.S. Firm Sizes

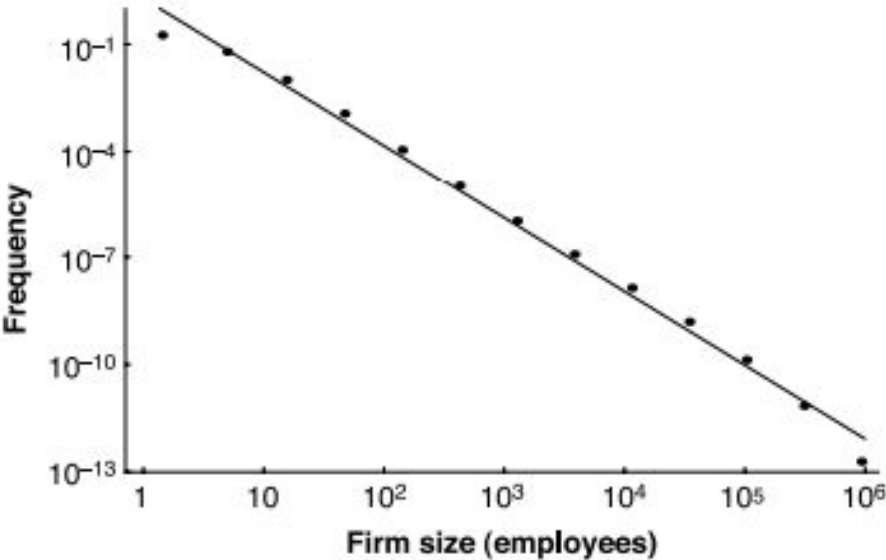
Robert L. Axtell

Analyses of firm sizes have historically used data that included limited samples of small firms, data typically described by lognormal distributions. Using data on the entire population of tax-paying firms in the United States, I show here that the Zipf distribution characterizes firm sizes: the probability a firm is larger than size s is inversely proportional to s . These results hold for data from multiple years and for various definitions of firm size.

- Oficina del Censo de EE.UU./El número total de empresas que tuvieron empleados en algún momento de 1997, aprox. 5,5 millones, de las cuales más de 16.000 tenían más de 500 empleados.
- Los datos del censo muestran un aumento monótono del número de empresas cada vez más pequeñas, una forma que la distribución lognormal no puede reproducir, y que sugiere que puede aplicarse una distribución de ley de potencia

Table 1. U.S. firm size distribution in 1997, compared across data sources. Number of firms in various size categories, with size defined as the number of employees, comparing COMPUSTAT and U.S. Census Bureau data for 1997. Note that there are monotonically decreasing numbers of progressively larger firms in the Census data, whereas this is not the case in the COMPUSTAT data (29).

Size class	COMPUSTAT	Census
0	2,576	719,978
1 to 4	123	2,638,070
5 to 9	149	1,006,897
10 to 19	251	593,696
20 to 99	1,287	487,491
100 to 499	2,123	79,707
500+	4,267	16,079
Total	10,776	5,541,918



REPORTS

Fig. 1. Histogram of U.S. firm sizes, by employees. Data are for 1997 from the U.S. Census Bureau, tabulated in bins having width increasing in powers of three (30). The solid line is the OLS regression line through the data, and it has a slope of 2.059 (SE = 0.054; adjusted $R^2 = 0.992$), meaning that $\alpha = 1.059$; maximum likelihood and nonparametric methods yield similar results. The data are slightly concave to the origin in log-log coordinates, reflecting finite size cutoffs at the limits of very small and very large firms.

Zipf's law for US Firms size

Table 3. Theoretical power law exponents for U.S. firms over a 10-year period. Note that even though the number of firms and total employees each increased over this period, as did the average firm size, the value of α was approximately unchanged.

Year	Firms	Employees	Mean firm size	α , from (4)
1997	5,541,918	105,299,123	19.00	0.9966
1996	5,478,047	102,187,297	18.65	0.9986
1995	5,369,068	100,314,946	18.68	0.9983
1994	5,276,964	96,721,594	18.33	1.0004
1993	5,193,642	94,773,913	18.25	1.0008
1992	5,095,356	92,825,797	18.22	1.0009
1991	5,051,025	92,307,559	18.28	1.0004
1990	5,073,795	93,469,275	18.42	0.9995
1989	5,021,315	91,626,094	18.25	1.0006
1988	4,954,645	87,844,303	17.73	1.0039

El patrón matemático que nos define cuando cruzamos un paso de cebra

- El ‘Vuelo de Lévy’ es un paseo aleatorio que utilizan desde los tiburones hasta los cazadores-recolectores de la tribu Hazda de Tanzania cuando salen a cazar
- ¿Qué tienen en común los tiburones con los cazadores-recolectores de la tribu Hazda de Tanzania y los peatones que cruzan un paso de cebra en cualquier ciudad del mundo? Resulta que todos ellos, además de muchos otros animales, utilizan un patrón matemático conocido como el “Vuelo de Lévy” para definir instintivamente sus trayectos.
- En los dos primeros casos, los científicos hace tiempo que averiguaron que aplican este sistema, que mezcla trayectorias largas con movimientos cortos al azar, durante sus jornadas de caza. En el caso de los aglomerados pasos de peatones, las personas tienden a caminar de forma predecible para poder evitarse entre sí, según un [estudio publicado en el Journal of the Royal Society Interface.](#)

El patrón matemático que nos define cuando cruzamos un paso de cebra

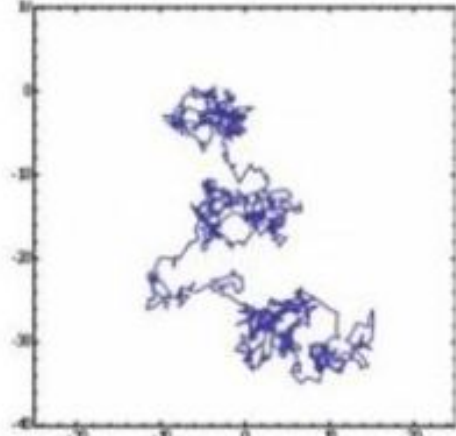
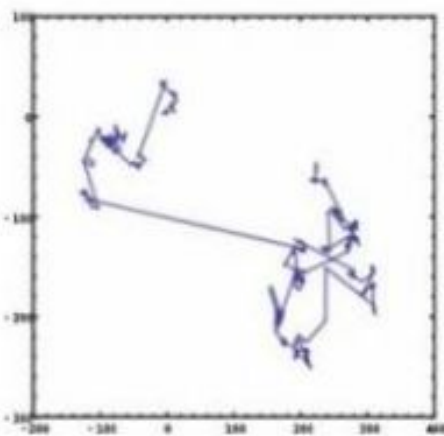
- El ‘Vuelo de Lévy’ es un patrón que mezcla trayectorias largas con movimientos cortos al azar
 - las “multitudes humanas exhiben varios patrones colectivos que emergen de la autoorganización” y en los que los individuos “anticipan los movimientos de sus vecinos para buscar sus caminos en el flujo peatonal dinámico”.
- Esta búsqueda de la ruta más funcional da como resultado que las personas se desvíen “de sus direcciones deseadas, de la ruta directa a su destino”, escriben. “Eso se hace a través de la estrategia de movimiento sin escala llamada Vuelo de Lévy, que podría facilitar la transición al comportamiento a nivel de grupo”



Para obtener estas conclusiones, los científicos organizaron un experimento en el que filmaron a un grupo de voluntarios mientras caminaban de forma aleatoria a través de unos corredores. Lo que observaron fue “cómo los flujos de peatones en movimiento opuesto se separan espontáneamente en varios carriles unidireccionales”.

“Descubrimos que antes (pero no después) de la formación de carriles, los peatones se desvían de la dirección deseada por el proceso de caminar de Lévy, que se considera óptimo cuando se buscan recursos distribuidos de forma impredecible”, añaden. Es decir, que cada individuo hizo ajustes en su camino asándose en las acciones de quienes viajaban en dirección puesta.

Así, en lugar de que las personas se encuentren continuamente cara a cara, simplemente seguirían a alguien que se mueve en la misma dirección, evitando la necesidad constante de cambiar de rumbo, aumentando de esta manera la eficiencia, tanto individual como colectiva.



Leyes de potencias en la ocurrencia de fenómenos críticos

Las leyes de potencias pueden ocurrir en la naturaleza en las proximidades de una *transición de fase*.

Los diagramas de fases ilustran las variaciones entre los estados de materia de elementos o compuestos en relación con la presión y las temperaturas. A continuación se muestra un ejemplo de un diagrama de fases para un sistema genérico de un solo componente:

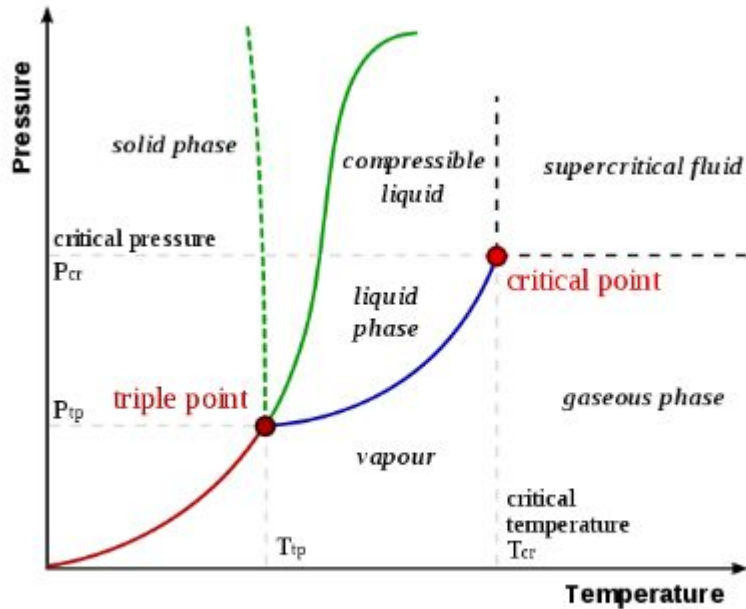


Figura 1. Diagrama general de fases

- **Punto triple:** el punto en un diagrama de fases en el que coexisten los tres estados de la materia: gas, líquido y sólido
- **Punto crítico:** el punto en un diagrama de fases en el que la sustancia es indistinguible entre los estados líquido y gaseoso
- **Curva de fusión (fusión) (o congelación):** la curva en un diagrama de fases que representa la transición entre los estados líquido y sólido
- **Curva de vaporización (o condensación):** la curva en un diagrama de fases que representa la transición entre los estados gaseoso y líquido
- **Curva de sublimación (o deposición):** la curva en un diagrama de fases que representa la transición entre los estados gaseoso y sólido

Las “*transiciones de fase*” se dan en sistemas físicos en los que se produce un cambio estructural como pueden ser

- Cambio de una sustancia cuando se funde, congela o evapora (*cambio de sólido \Leftrightarrow líquido \Leftrightarrow vapor*)
- *Magnetización* de un medio
- Percolación*: aparición de un camino que vincula partes distantes de un sistema o situación en la cual un cambio local se siente en todo el sistema

En un sistema social pueden darse fenómenos de esta naturaleza por ejemplo, circulación de rumores, adopción de modas, etc.

- Se da cuando dos agentes desvinculados entre si adoptan la misma conducta porque existe un conjunto de terceros agentes que actúen como intermediarios

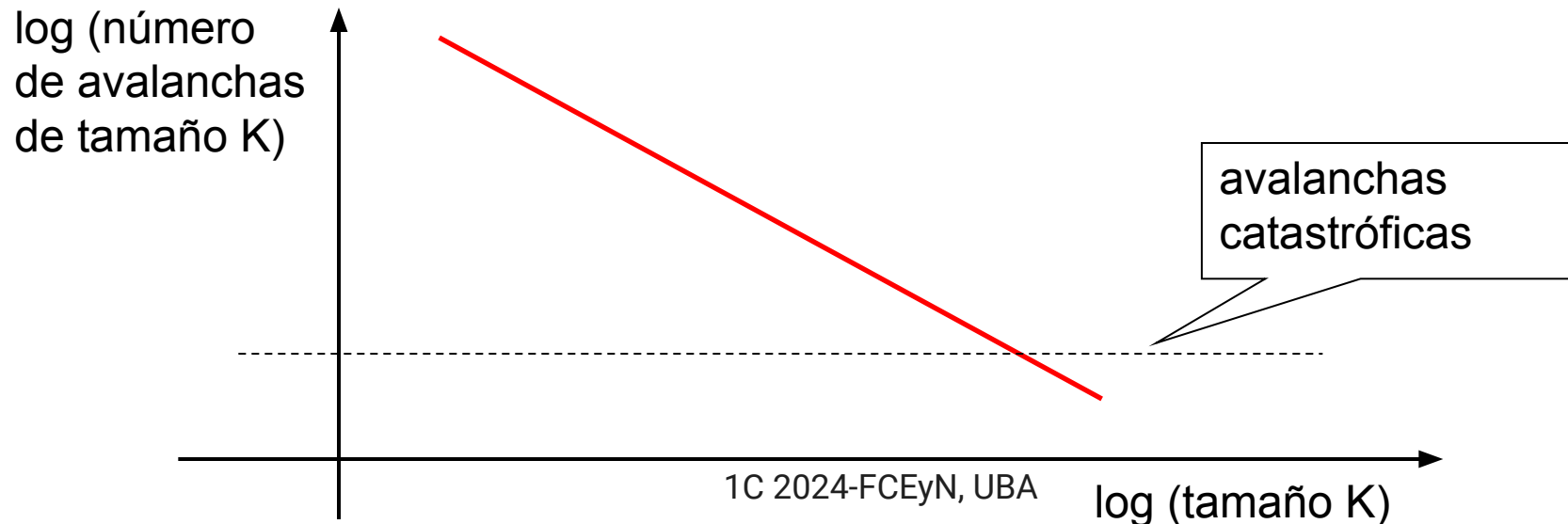
Por lo general las transiciones de fase se producen para un valor muy preciso de algún parámetro estructural o de control.

3. Leyes de potencia en el Modelo de Criticalidad Auto-organizada

El **Modelo de Criticalidad Auto-organizada (SOC)** fue propuesto como un mecanismo genérico para dar lugar a leyes de potencias

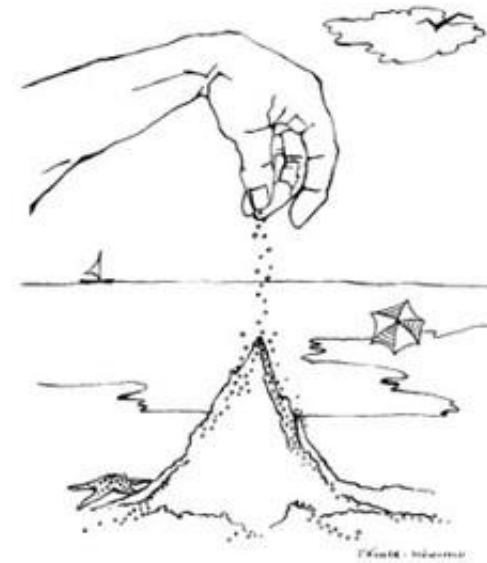
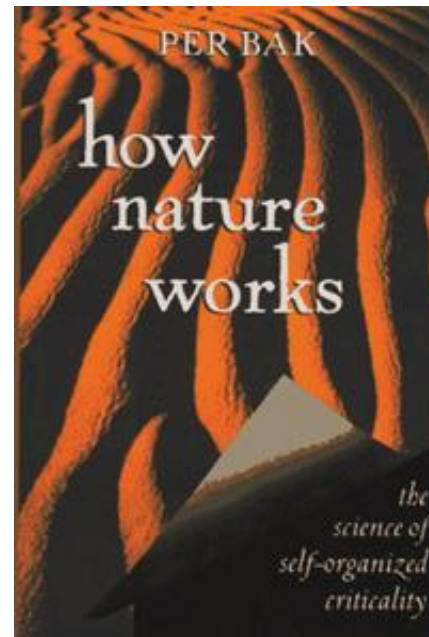
El modelo **SOC** da cuenta de la ocurrencia de leyes de potencias suponiendo que existe un mecanismo subyacente de “**todo o nada**” en la interacción entre los integrantes de un sistema abierto. Este mecanismo puede dar lugar a la concatenación de eventos (**avalanchas**)

El SOC se lo suele ejemplificar con avalanchas en una “pila de arena”. El sistema se **auto-organiza** en un estado crítico con una distribución estable en el tamaño de las avalanchas. En este modelo las catástrofes no requieren de explicaciones ad hoc



SOC: el modelo de pila de arena

- ¿Habrá algún mecanismo simple y único que genere la complejidad que observamos en la naturaleza?
 - Per Bak (1947) “How Nature Works”, Oxford Univ. Press



¿Cuál es la hipótesis de SOC?

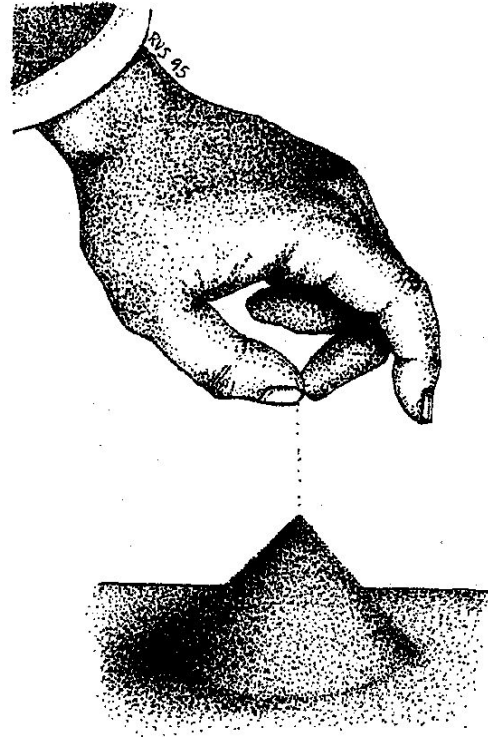
- La hipótesis de Per Bak sugiere que una gran cantidad de clases de sistemas se comporta como sistemas termodinámicos en el punto de transición de fase.
- Estos sistemas se mueven espontáneamente hacia ese estado (a diferencia de los sistemas en equilibrio termodinámico para los cuales hace falta sintonizar algún parámetro).

Qué es Criticalidad Auto-Organizada? (SOC)

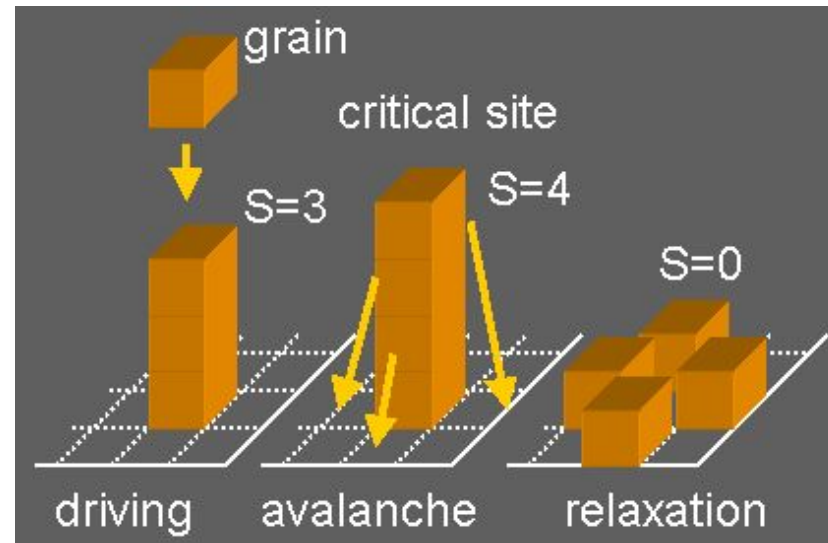
La idea es especificar un mecanismo simple que produzca una conducta típica compartida por una gran cantidad de sistemas sin depender de los detalles específicos del sistema en particular.

El “sistema” evoluciona en el tiempo bajo la influencia **Externa** y **Interacciones** internas.

El modelo



Las reglas



Simplicidad: Los granos interactúan y causan que otros se muevan

El modelo de la pila de arena de Per Bak se formaliza mediante un autómata celular que idealiza una pila a la que se le agregan granos desde el exterior:

Dinámica

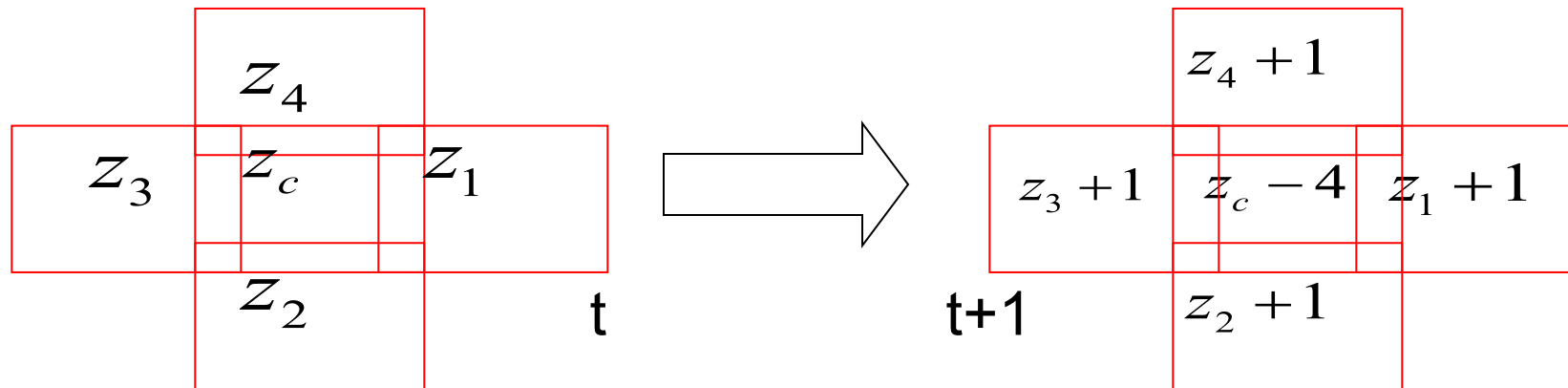
Cada sitio del autómata puede albergar Z_c “granos de arena” y cuando se supera ese umbral se descargan “granos de arena” a todos los sitios vecinos a él:

Si al agregar un grano resulta que $z_i > z_c$
en el instante subsiguiente, ese sitio descarga un grano en todos sus vecinos

$$z_i(t+1) = z_i(t) - n_v$$

$$z_{i \in N(i)} = z_{i \in N(i)} + 1$$

(**Nota:** los sitios en el borde del autómata descargan sus granos al exterior del mismo)

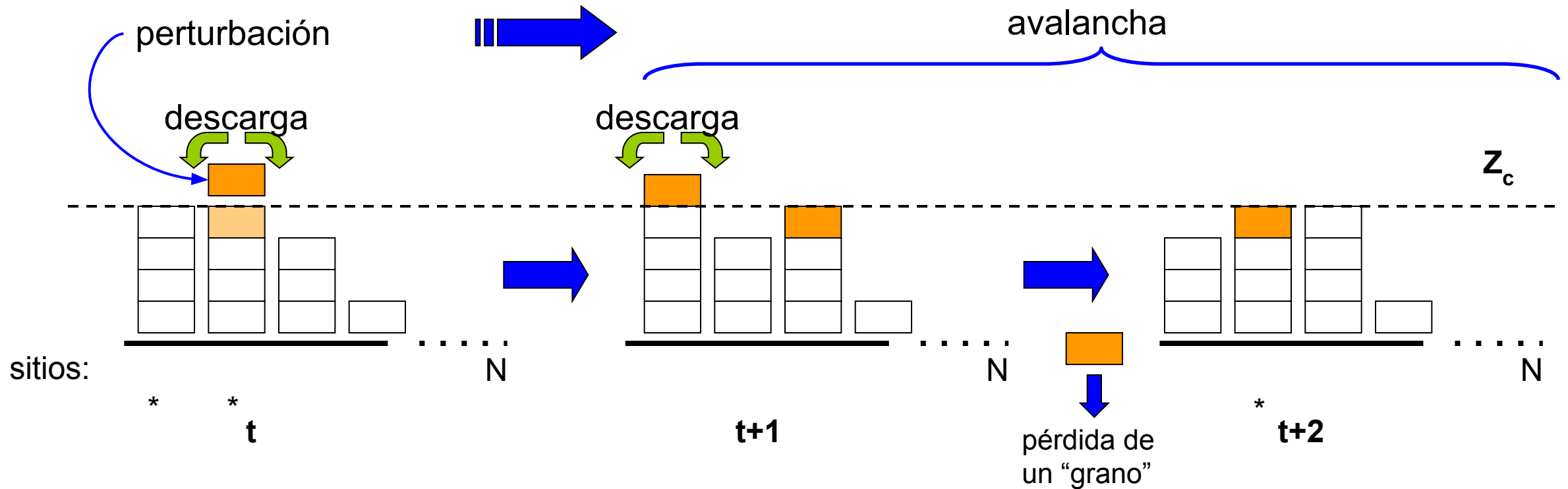


Estado inicial (estado crítico)

- Se llena el autómata de N celdas con una gran cantidad de granos de arena y se permite que todos los sitios descarguen sobre sus vecinos y sobre el exterior del sistema hasta que no haya más desplazamiento de granos. Ese es el estado crítico

Estadística

- Se agrega un grano de arena sobre cada uno de los sitios con $Z = Z_c$ del estado crítico.
- En cada oportunidad se permite que se produzcan avalanchas hasta que se recupera el reposo. Todos los sitios del autómata se actualizan sincrónicamente.
- Se contabiliza cuántos granos se movieron en todas las avalanchas. Esa es la “energía” liberada en ese evento.
- Se repite todo el proceso para obtener una estadística representativa del proceso



(*) sitios críticos

Ejemplo de una "pila de arena" dimensión 1 con una vecindad de dos sitios (der. e izq.)

Ilustración de una avalancha en dos dimensiones

Agregamos
un grano aqui

a

1	2	0	2	3
2	3	2	3	0
1	2	3	3	2
3	1	3	2	1
0	2	2	1	2

b

1	2	0	2	3
2	3	2	3	0
1	2	4	3	2
3	1	3	2	1
0	2	2	1	2

c

1	2	0	2	3
2	3	3	3	0
1	3	0	4	2
3	1	4	2	1
0	2	2	1	2

d

1	2	0	2	3
2	3	3	4	0
1	3	2	0	3
3	2	0	4	1
0	2	3	1	2

e

1	2	0	3	3
2	3	4	0	1
1	3	2	2	3
3	2	1	0	2
0	2	3	2	2

f

1	2	1	3	3
2	4	0	1	1
1	3	3	2	3
3	2	1	0	2
0	2	3	2	2

g

1	3	1	3	3
3	0	1	1	1
1	4	3	2	3
3	2	1	0	2
0	2	3	2	2

h

1	3	1	3	3
3	1	1	1	1
2	0	4	2	3
3	3	1	0	2
0	2	3	2	2

i

1	3	1	3	3
3	1	2	1	1
2	1	0	3	3
3	3	2	0	2
0	2	3	2	2

Fin de la
avalancha



a=ayer,
b=hoy,
c=mañana,
...etc

Cada perturbación puede generar avalanchas de tamaños muy desiguales

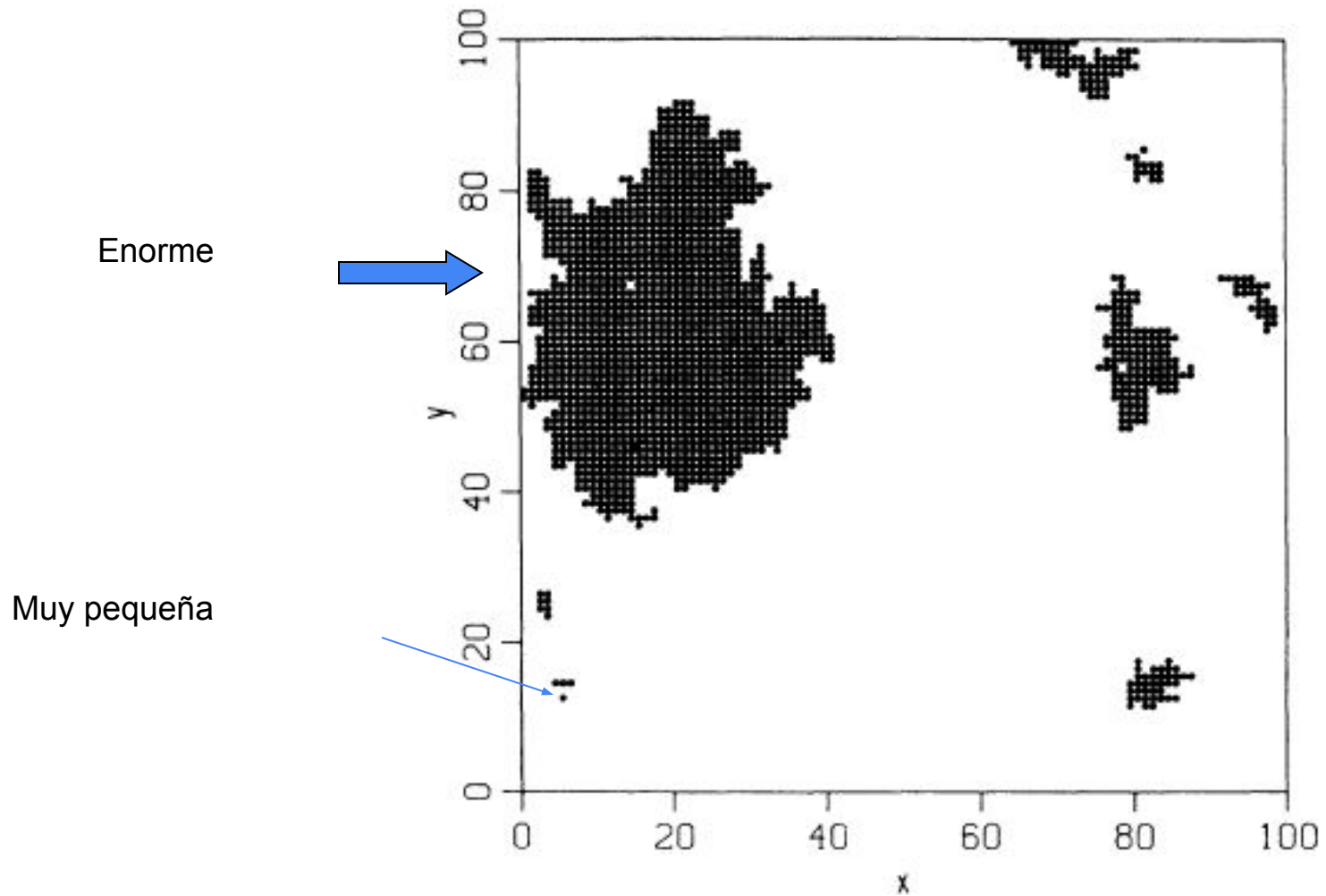
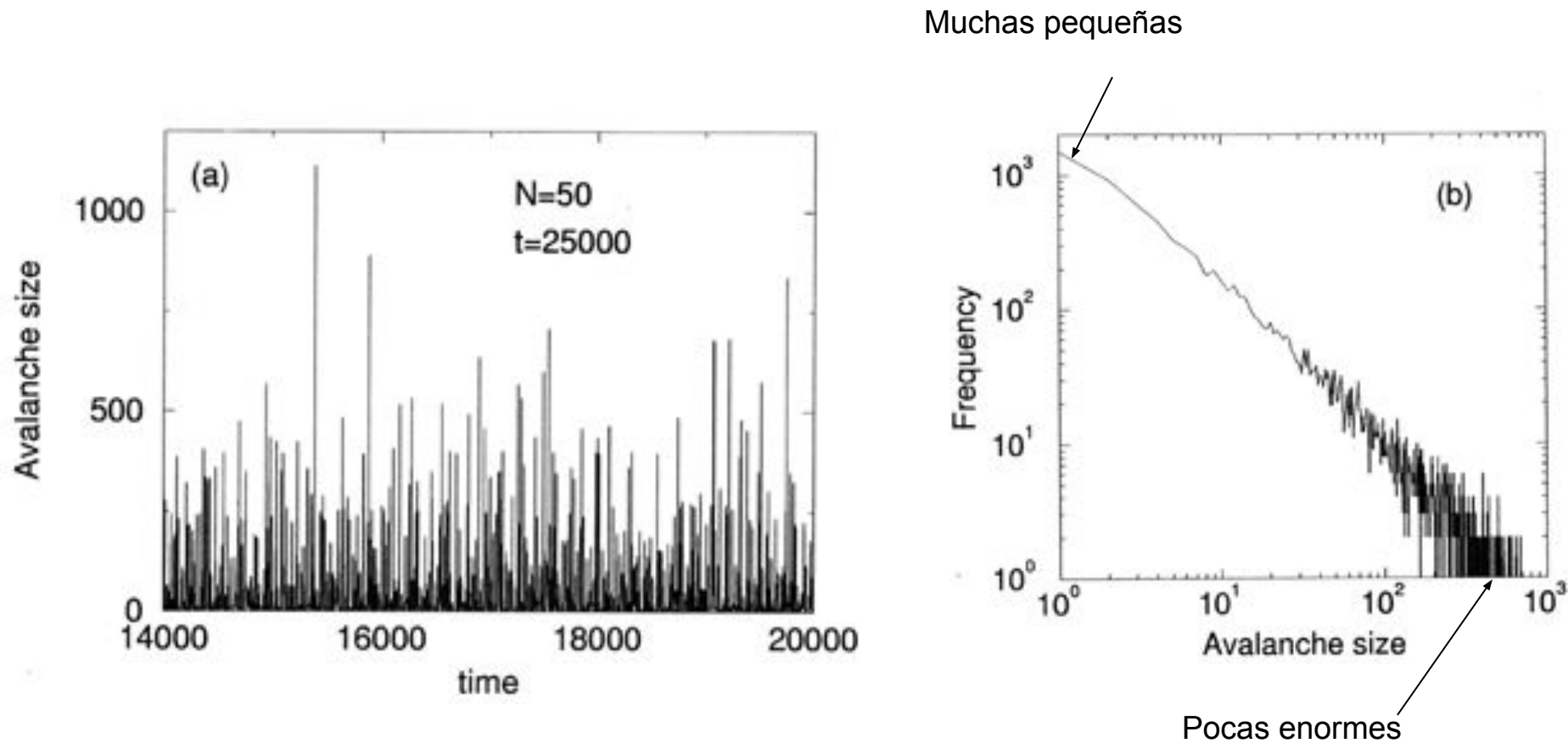
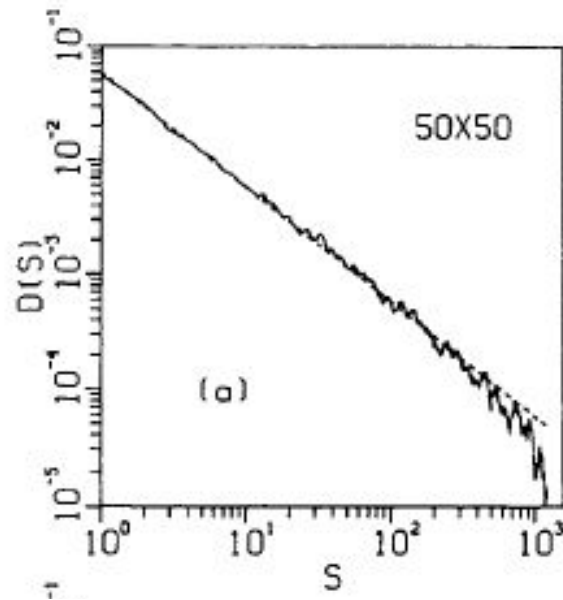


FIG. 2. Typical domain structures resulted from several local perturbations for a 100×100 array. Each cluster is triggered by a single perturbation.

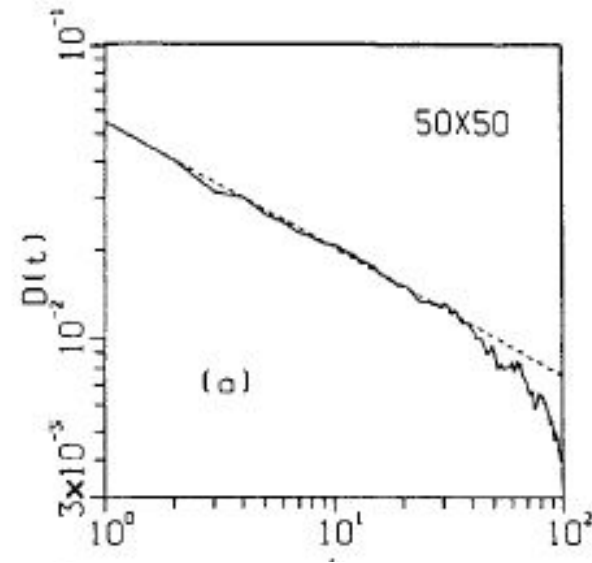
El sistema espontáneamente alcanza criticalidad



La distribución tanto de la duración como del tamaño de las avalanchas es libre de escala.



Tamaño



Duración

Sólo limitado por el tamaño del sistema...

Las avalanchas son un fenómeno

- emergente
- Complejo
- Inevitable
- determinístico

Criticalidad es el único estado en que al mismo tiempo

- Es el más inestable (un solo granos basta...)
- Es el más robusto (se vuelve siempre a el...)

Bak SOC: algunas referencias

<https://www.pawelmandera.com/2017/12/24/sandpile/>

<https://www.jasss.org/4/4/reviews/bak>

<https://www.sciencedirect.com/science/article/abs/pii/S0378437104005667>

<https://spiral.imperial.ac.uk/bitstream/10044/1/51489/2/25%20Years%20of%20Self-organized%20Criticality.pdf>

Características del modelo SOC

- Los experimentos numéricos realizados con este modelo dan distribuciones de intensidad de avalanchas que siguen leyes de potencia.
- El exponente crítico de las mismas **sólo** está determinado por la dimensionalidad (número de vecinos) del autómata celular y es independiente de la interacción que se suponga entre sitios **en tanto ésta sea de umbral**
- Se habla de criticalidad **auto-organizada** debido a que las distribuciones que siguen leyes de potencias se presentan en fenómenos críticos determinados por un **parámetro de control externo** (usualmente la temperatura) mientras que en este caso la organización se debe a la dinámica del propio sistema.
- Este modelo ha tenido particular éxito describiendo la ley de Gutenberg & Richter que determina que las intensidades de los terremotos siguen una distribución que se ajusta a una ley de potencias.

¡Nos vemos en la próxima clase!