



UNIVERSIDAD NACIONAL AUTÓNOMA  
DE MÉXICO

---

---

FACULTAD DE CIENCIAS

LEYES DE POTENCIAS DE DOS COLAS

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

F Í S I C O

P R E S E N T A :

OCTAVIO EDUARDO VIZCAYA XILOTL



FACULTAD DE CIENCIAS  
UNAM

TUTOR:  
DR. GERMINAL COCHO GIL

2009



Universidad Nacional  
Autónoma de México



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

1. Datos del alumno.

Vizcaya

Xilotl

Octavio Eduardo

51 12 61 98

Universidad Nacional Autonoma de Mexico

Facultad de Ciencias

Fisica

091165501

2. Datos del tutor.

Dr

Cocho

Gil

Germinal

3. Datos del sinodal 1

Dr

Carmona

Ruiz

Gerardo

4. Datos del sinodal 2

Dr

Stephens

Stevens

Christopher

5. Datos del sinodal 3

Dr

Mansilla

Corona

Ricardo Lino

6. Datos del sinodal 4

Dr

Miramontes

Vidal

Pedro Eduardo

7. Datos de la tesis.

Leyes de potencias de dos colas

77 p.

2009

# Agradecimientos

---

Las siguientes líneas, cargadas de *pensantimientos intersubjetivos*, sólo son una muestra de gratitud hacia aquellos seres humanos que han contribuido, según mi propio criterio, en la realización de este trabajo. En un ambiente de tremenda *transitoriedad*, intentan señalar su contrario: lo que permanece cambiando, en este caso, los nodos de un entramado reticular que se mueve, posiblemente, de acuerdo a la *topología* del espacio de las cercanías personales. Por ello, sugiero leer las cursivas así, de forma personal, y con amplitud interpretativa.

Va pues, lo que no pretende ser una letanía, sino tan sólo una pequeña muestra *reiterada* de afecto y agradecimiento:

A la *malcontenta*, donde quiera que se esté... por los dolorosos aprendizajes condensados en el tiempo, y por dejar *preguntas con semilla de respuesta... varada por la realidad...*

A *Ji-menina*, en su doble acepción, por la risa, y por hacer su tarea permanente de ser hija, de la cual aprendo lentamente...

A Germinal Cocho Gil, por las enseñanzas científicas y epistémicas, la paciencia rayana en lo infinito (y en tiempo real), por sus comentarios acerca del “Barbas”, así como por las teorizaciones sobre la *ayuda mutua*.

A Flavio Cocho Gil, por la imprescindible ayuda *metapocatastásica*, sobre todo por indicarme la lectura del poema *If* de Kipling; por la paternidad ideológica y por la sabrosa lectura de su caleidoscopio literario, que es *insoluble en Kapitalismo*.

A Moisés Villegas, por las enseñanzas pediátricas y médicas que me han hecho reconocermé (y mucho) en algo en lo que ya no me reconocía: la científicidad de la medicina y la lucha contra los tartufos y charlanes.

A Einar Adame, por la amistad solidaria en los *momentos depres*, a pesar de que nos encante “respirar diesel por la tarde” y cíclicamente den ganas de tirar la toalla

a la manera *del Sísifo*.

A Ricardo Mansilla, por la “asistencia en el camino”, por el aprendizaje basado en analogías, y por mostrar que, contrariamente a lo que me había convencido ya Richard P. Feynman, el análisis matemático es necesario y puede ser divertido.

A los sinodales, por dedicar un poco de su tiempo en la realización de comentarios y sugerencias, y por el planteamiento de nuevas e interesantes *betas* de trabajo.

A Anaesther, por los consejos *cibernáuticos*; los realistas-fríos, y los metodológicos, que me llevan a pensar en la existencia de microgeopolíticas *hegemónicas*...

A Fernando Zambrana, por las enseñanzas editoriales, las mermeladas y los comentarios de medio tiempo.

A mi madre, Soledad Xilotl, que, como su nombre lo indica, ha tenido que llevar esa carga (pero que con *escéptica esperanza* habrá de reducirse), por la maternidad bio-sociológica.

A mis hermanas, Lety, Miry y Vicky, por el apoyo fraterno que *se renace* desde sus cenizas.

A mi hermano, Javier, por el *ensordecedor silencio*, que a veces ha dicho más que las palabras, y por la ventana de oportunidad que representa la existencia.

Al año 1999, punto de bifurcación individual e inicio de algunos caminos sin retorno a bordo de una *barcaza rojinegra*; y otros, *muy otros*, con una cierta cantidad de ergodicidad manifiesta y modificable, siendo este trabajo una muestra.

A los *oximorones* y a los *chamizales*, (cuyos nombres intentan referirse a conocidos y amigos que si ahora no cito explícita y puntualmente no es por ausencia, sino pretextando mi mala memoria) por intentar transformar al mundo, transformándonos a nosotros mismos, a pesar de los resultados y haciendo votos por nuevas *emergencias* y *convergencias*.

# Índice

---

<b>Agradecimientos</b>	<b>iii</b>
<b>Introducción</b>	<b>vii</b>
<b>1 Aspectos generales</b>	<b>1</b>
1.1 Gigantes, escalas, leyes de potencia y colas . . . . .	1
1.2 Escalas típicas, jerarquías y descripciones . . . . .	4
1.3 Funciones homogéneas y funciones libres de escala . . . . .	5
1.4 Mecanismos que dan lugar a leyes de potencia . . . . .	10
1.4.1 Transiciones de fase de segundo orden: fenómenos críticos . . .	10
1.4.2 Objetos fractales . . . . .	15
<b>2 Distribuciones</b>	<b>23</b>
2.1 Funciones homogéneas como distribuciones . . . . .	23
2.2 Contexto sociohistórico . . . . .	24
2.3 Meteoritos y terremotos . . . . .	25
2.4 Distribución de ingresos. Ley de Pareto . . . . .	26
2.5 Los aportes de Zipf . . . . .	29
2.5.1 Estadística de ciudades . . . . .	29
2.5.2 De Pareto a Zipf y viceversa . . . . .	31
2.5.3 Estadística de palabras . . . . .	32
2.6 Modelos estocásticos . . . . .	36
<b>3 Distribuciones rango-orden</b>	<b>41</b>
3.1 Bibliometría, informetría, ... . . . .	41
3.2 Factores de impacto . . . . .	44
3.3 Ley de potencias de dos colas . . . . .	45
3.3.1 Resultados en el caso de los factores de impacto . . . . .	46
3.3.2 Conclusiones relativas a los factores de impacto . . . . .	47

---

3.3.3	Sobre la ubicuidad . . . . .	48
3.4	Ranking mundial de universidades . . . . .	48
3.5	Fluctuaciones y <i>rank-clocks</i> . . . . .	51
<b>4</b>	<b>Modelaje</b>	<b>57</b>
4.1	Ecuación maestra para comunidades ecológicas . . . . .	57
4.2	Generalización de la ecuación maestra . . . . .	58
4.3	Procesos estocásticos . . . . .	60
4.3.1	Ecuación de Langevin . . . . .	60
4.3.2	Construcción de la ecuación de Fokker-Planck . . . . .	62
4.3.3	De la ecuación de Langevin a la de Fokker-Planck . . . . .	62
4.3.4	Caso estacionario . . . . .	63
4.3.5	Observaciones . . . . .	64
<b>5</b>	<b>Discusión y conclusiones</b>	<b>65</b>
5.1	Conclusiones . . . . .	66
<b>6</b>	<b>Apéndice</b>	<b>67</b>
6.1	Desde una ecuación maestra generalizada . . . . .	67
6.2	Dentro del cálculo de Itô . . . . .	68
	<b>Bibliografía</b>	<b>71</b>

# Introducción

---

Desde hace poco tiempo, el estudio de fenómenos colectivos se ha visto beneficiado por los cambios tecnológicos ocurridos en el terreno de la computación. La existencia de grandes bases de datos, el *seguimiento capilar* de multitud de fenómenos digitalizables (es decir, el registro en tiempo real de una gran cantidad de observables), la popularización de las tecnologías afines, así como el incremento en la velocidad de cómputo, han incrementado las posibilidades de exploración y entendimiento de gran cantidad de fenómenos; entre las características que comparten se encuentra la gran cantidad de elementos interactuando y la notable existencia de distribuciones peculiares, que se pueden describir en términos de leyes de escalamiento o leyes de potencia, o, como en el caso de este trabajo, diversas generalizaciones de las mismas.

La investigación de leyes de potencia, en diversos campos, lleva ya más de un siglo. En el caso de la física, durante algún tiempo se centró en el estudio de los fenómenos críticos y las transiciones de fase; pero la aparición de estadísticas que se pueden ajustar mediante leyes de potencia, y en diversos contextos, ha motivado la recopilación de datos en los más diversos campos, provocando un efecto recíproco: las leyes de potencia (es decir, de un único exponente) han aparecido por doquier. Se observan en la distribución de meteoritos, ingresos, terremotos, guerras, llamadas telefónicas, citas de artículos científicos, palabras, ciudades, etc. Tal ubicuidad ha motivado la reflexión teórica pues siguiendo una de las largas tradiciones de trabajo en física, se considera que debe haber mecanismos generales que las expliquen, a pesar de que los datos no sean tradicionalmente procesados en física.

La ubicuidad de las leyes de potencia abarca las formulaciones rango-tamaño; desafortunadamente en muchos casos el ajuste falla en los extremos, es decir, el escalamiento sólo es válido para algún intervalo, que usualmente no incluye las colas. Es así que algunos autores han empleado desde la distribución lognormal hasta distribuciones de exponenciales truncadas, pasando por combinaciones de ellas (siendo el caso de la distribución doble de Pareto-lognormal [61]), esperando mejorar los ajustes. Lamentablemente, tales intentos no son capaces de ajustar al mismo tiempo las dos colas, y no se diga ya la descripción del intervalo completo de definición.



Diversos ordenamientos rango-tamaño, entre los que se encuentran: a) el factor de impacto de las revistas científicas [43], b) las secuencias genéticas de mapeos unimodales [11], y c) la estadística de las notas musicales [12], sugiere la existencia de distribuciones que se ajustan muy bien mediante el producto de dos leyes de potencia definidas sobre el conjunto completo de los datos, incluyendo las colas; una de ellas medida de “derecha a izquierda” y la otra de “izquierda a derecha”, y con coeficientes de correlación que rivalizan con los de las otras aproximaciones o que incluso los superan. La relación funcional de que hablamos, la ley de potencias de dos colas del título, y denotadas por  $\{a, b\}$ , en la literatura se denomina función beta de dos parámetros. La exposición central de este trabajo se da alrededor de la estadística del ranking mundial de universidades, para el que se encontró un muy buen ajuste en términos de la misma ley de potencias de dos colas: una función beta de dos parámetros.

En el capítulo 1 se hace una revisión de las primeras conceptualizaciones relativas a las leyes de potencia: los fenómenos críticos y los conjuntos fractales; además se presentan algunas propiedades matemáticas para el caso continuo, como es la homogeneidad y su equivalencia en términos de libertad de escala.

En el capítulo 2 se continúa con la exposición de algunas propiedades matemáticas, pero ahora en el caso de distribuciones; se presentan casos históricos representativos, asociados a los nombres de Pareto y Zipf (la distribución de ingresos, la estadística de ciudades y la estadística de palabras); finaliza con la descripción de un modelo estocástico (Reed [60]) para el caso de la distribución de los asentamientos humanos.

En el capítulo 3 se abordan los resultados que sugirieron la búsqueda en el ranking mundial de universidades: el caso de los factores de impacto de revistas científicas. Para el caso de las universidades, se citan las fuentes, sus metodologías, y se exponen los resultados obtenidos para tres indicadores bibliométricos durante los años 2003-2007. El capítulo finaliza con una exploración cualitativa, para el caso de las universidades, sobre las fluctuaciones, y se emplea una graficación debida a Michael Batty [9]: los *rank-clocks*, que intentan hacer una aproximación visual de la evolución temporal de los rangos.

En el capítulo 4 se exponen un par de modelos estocásticos que son fuertes candidatos para explicar la dinámica subyacente que genera fenómenos con leyes de potencias de dos colas. El primero de ellos se basa en algunas analogías, en términos de termalización, y se extrae de la dinámica de comunidades ecológicas, donde ha sido importante el trabajo de Hubbell [31]; después se generaliza y se establecen algunas condiciones para que se pueda obtener la representación  $\{a, b\}$ . Posteriormente, a partir de la ecuación de Langevin se pasa a una ecuación de Fokker-Planck, para ser resuelta en el caso estacionario, y sumado a condiciones específicas, dar lugar, nuevamente, a una ley de potencias de dos colas. Finalmente, en el capítulo 5 se elabora la discusión y algunas conclusiones.

# Aspectos generales

---

## 1.1 Gigantes, escalas, leyes de potencia y colas

*Los viajes de Gulliver* (1726), de Johnatan Swift, es referencia obligada cuando se trata de abordar el tema de los objetos escalados, ya sean miniaturas o gigantes. En la descripción que hace de Lilliput y Brodignag (la tierra de humanos miniatura y la tierra de los gigantes, respectivamente), logra un verdadero reportaje en tanto que sus descripciones concuerdan con las medidas exactas, poniendo gran atención a las proporciones, aunque ninguna a las posibilidades de existencia física. Y ello es correcto, pues no intentaba dar cuenta de eso, sino hacer fluir las letras. Dejando de lado las múltiples interpretaciones a las que se pueden someter las travesías gulliverianas (que van desde la crítica sociopolítica hasta el psicoanálisis), la idea de que seres vivos a escala existan, tanto gigantes como miniaturas, es atrayente por sí misma.

En sus *Diálogos referentes a dos ciencias nuevas* (1638), Galileo explora las posibilidades de un mundo habitado por gigantes, es decir, objetos (animales y estructuras) cuyos tamaños han sido aumentados varias veces. Las dos ciencias nuevas a las que Galileo se refiere son la ciencia del movimiento (cinemática) y la ciencia de resistencia de materiales. Una pregunta que surge de manera natural en este último campo es: ¿cómo debemos modificar un cierto objeto (animal, construcción, etc.) para que al hacer una escala<sup>1</sup> del mismo, su estructura no se desplome bajo su propio peso? Dicha cuestión es presentada en varios textos básicos de física y biología, con las siguientes variantes: ¿puede existir una hormiga tan grande como un elefante?, ¿es posible que exista una célula del tamaño de una hormiga?

Para responder a estas cuestiones, reproduzcamos una argumentación parecida a la dada por Galileo, y concentrémonos en que el ‘objeto’ a considerar es una vaca (no necesariamente esférica): Como el peso de la vaca es proporcional a su volumen, y la resistencia de sus huesos es proporcional a la sección transversal, si queremos una

---

<sup>1</sup>Aquí al ‘hacer una escala’ nos referimos a aumentar el tamaño, aunque muy bien podría uno pensar en disminuirlo.

vaca gigante, digamos del doble de tamaño de una vaca común, tendremos que la vaca gigante pesará 8 veces más que la normal, pero la resistencia de sus huesos deberá aumentarse más del doble para poder soportarla; por lo que si seguimos aumentando el tamaño de la vaca y conservando su forma, en algún momento sus huesos no la soportarán y entonces no podrá vivir. De manera que la respuesta a las preguntas planteadas es negativa.

Del mismo modo, imposible sería edificar las estructuras óseas de los hombres, caballos y demás animales de manera que quedasen trabadas y cumpliesen sus funciones ordinarias si dichos animales tuvieran que aumentar enormemente su estatura; porque este aumento de estatura no puede lograrse sino empleando un material más duro y fuerte que el de costumbre o agrandando el tamaño de los huesos, y cambiándoles así la figura . . .

(Galileo, *Diálogos referentes a dos ciencias nuevas*, reproducido en [52], p. 29.)

Pero no hay que pensar en fantasías que podrían rayar en la monstruosidad para darle la razón a la descripción galileana, pues nos ocurre en cierta medida a los humanos: en un bebé, su cabeza es mayor en relación a su cuerpo que en un adulto; a medida que crece, el cuerpo del bebé se va deformando para poder cumplir las funciones que su crecimiento demanda, por ejemplo, la cabeza va ocupando cada vez menos porcentaje de la masa corporal, y de manera paralela el tronco va aumentando su porcentaje relativo. En otras palabras, los bebés no son seres *a escala* de los adultos: la cabeza crece a un cierto ritmo, el tronco a otro y las extremidades a otro más.

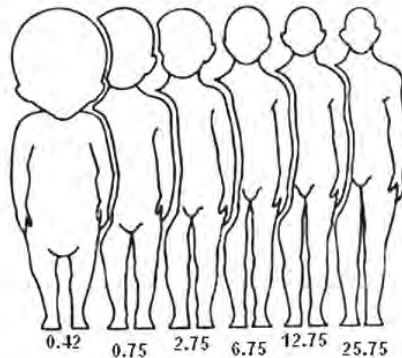


Figura 1.1: Comparativo del cuerpo humano a varias edades, indicadas por los números que están debajo. Nótese el cambio relativo para cada edad.

Matemáticamente se dice que dos objetos son *semejantes* cuando uno está hecho a escala del otro; en este sentido, los bebés no son nuestros semejantes, tan sólo son parecidos a nosotros. Ahora bien, para asegurarnos de que dos cuerpos,  $C$  y  $C'$ , son semejantes, debemos comparar sus lados correspondientes, y lo serán en el caso en que hallemos que existe una razón constante  $\lambda$ , en ese caso podremos decir que uno es la escala del otro.

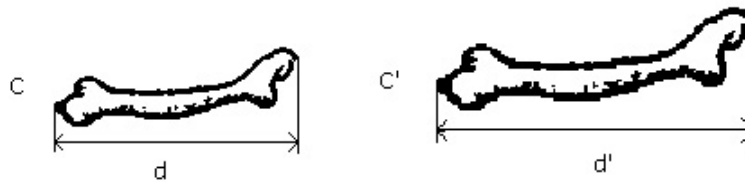


Figura 1.2: Huesos.

Si existe  $\lambda = d'/d$ , el factor de escala o de escalamiento, también se pueden considerar las razones de sus áreas y volúmenes correspondientes. Se tiene entonces que:

$$\frac{A'}{A} = \lambda^2 \quad \frac{V'}{V} = \lambda^3$$

De manera muy general, y pensando en el caso del desarrollo humano, se podría pensar en que el factor de escala sea una función de los puntos correspondientes a considerar, de forma tal que para lados correspondientes diferentes, factores de escala diferentes (es lo que le ocurre a los fractales afines). Si el factor de escala es constante, podemos construir tres funciones, que denotan las razones entre longitudes, áreas y volúmenes, tales que:

$$f_p(\lambda) = \lambda^p \quad \text{con } p = 1, 2, 3 \text{ para longitud, área y volumen respectivamente.}$$

Estas funciones así definidas pertenecen a un tipo general de funciones (sobre el campo de los números complejos) que se conocen como funciones potenciales, y que tienen la forma:

$$f(x) = ax^p \quad (a, x, p \in \mathbb{C}) \quad (1.1)$$

De estas funciones también se dice que siguen una *ley de potencia* (ley potencial o ley de escala), y para los casos de interés físico de este trabajo pensaremos en funciones reales de variable real. Cuando decimos que un fenómeno sigue una ley de potencia es suficiente con que la función que lo describe tenga un comportamiento

asintótico, lo que se denota por  $f(x) \sim x^p$ .

## 1.2 Escalas típicas, jerarquías y descripciones

Planteado en la terminología de la sección anterior, podemos decir que en la física abundan las leyes de potencia. En electromagnetismo se tiene la ley de Coulomb; en mecánica, la ley de gravitación universal. De hecho tienen la misma forma funcional, pero con significados e implicaciones físicas muy diferentes. Aparentemente no hay límites a su campo de acción, y decimos aparentemente, porque se sabe que la mecánica cuántica y la relatividad las limitan a ciertas escalas. Se tiene también que la energía elástica de un resorte crece como ley de potencia y que aparentemente no tiene límites, pero sabemos que ese crecimiento está acotado, ya que los resortes no son ideales sino que a veces se deforman permanentemente perdiendo sus cualidades elásticas o de plano se rompen. Todos estos ejemplos de leyes de potencia comparten el hecho de que se refieren a partículas puntuales, y que una vez fijada una de ellas para su estudio, se la observa en lo sucesivo. Comparten también el hecho de que tienen límites de aplicación dados por la física, es decir, se aplican a ciertas escalas.

Regresando al ejemplo de la vaca, podemos decir también que existen seres vivos tan grandes como la resistencia de sus huesos se los permite (en el caso de los vertebrados), y que ello limita las posibilidades de la biodiversidad, pues aunque la resistencia ósea no es el único factor involucrado –en última instancia condicionada por la gravedad [78]–, sí juega un papel central; además entran en juego factores dinámicos, las funciones de sus órganos, la relación entre el tamaño de su piel (área) y la disipación energética, la disponibilidad de alimento, etc. (ver [5] y [78]). Podemos también decir que la vida –al menos como la conocemos– tiene una escala definida, aunque hay que mencionar lo sorprendente que resulta el que abarque tantos órdenes de magnitud (desde  $10^{-6}$  m para los microbios hasta  $3 \times 10^1$  m en el caso de las ballenas, o incluso más en el caso de los extintos dinosaurios).

De manera general, observamos objetos a diferentes escalas: la nuestra, la de las galaxias, la de los microbios, la atómica y así. Y es el hecho de que la materia esté organizada jerárquicamente en niveles y escalas bien definidos (quarks, hadrones, átomos, moléculas, macromoléculas, células, órganos, seres vivos, etc.) lo que posibilita la existencia de la física, la química y la biología, pues sería imposible estudiar cualquier fenómeno natural si para hacerlo debiéramos describir todos sus componentes atómicos (ver [18]). El éxito de casi todas las teorías prácticas en física depende de aislar a los objetos de estudio en algún rango limitado de escalas. Si fuera necesario conocer la posición y la velocidad de todas y cada una de las moléculas en hidrodinámica, nos quedaríamos pasmados, pues no existen herramientas de cálculo

que manipulen semejante número de ecuaciones.

Pero no sólo en cuanto a la organización de la materia existen escalas definidas (y no sólo para la física, sino también para la biología), pues la mayoría de los fenómenos naturales tiene asociada una escala espacial y temporal (tienen un origen y un fin). Y si queremos entenderlos debemos elegir adecuadamente la escala de descripción. También vemos de forma general que los eventos que se distinguen por una gran disparidad en tamaño tienen poca influencia uno sobre el otro y que los fenómenos asociados con cada escala pueden ser tratados de forma independiente. Por ejemplo, las olas del mar pueden ser descritas como perturbaciones de un medio continuo, ignorando por completo la estructura molecular del líquido. Y lo mismo vale para la electricidad: su naturaleza granular se ignora en la descripción maxwelliana.

Incluso a nivel de física teórica hay escalas definidas, pues se piensa en la existencia de objetos físicos a niveles ultramicroscópicos y responsables de toda la materia: las supercuerdas, y que tienen asociadas una escala característica conocida como “escala de Planck”. Se supone que en esa escala ocurren fenómenos físicos muy interesantes y trascendentes, pero lo sorprendente aquí es que se puede calcular el orden de magnitud de las cosas de interés, i.e. las consideradas unidades fundamentales (longitud, masa y tiempo de Planck) empleando tan sólo análisis dimensional.

### 1.3 Funciones homogéneas y funciones libres de escala

Hasta ahora sólo nos hemos referido a fenómenos a los que se les asocia una escala típica; hemos hablado de fenómenos que están ‘atados’, por así decirlo, a ciertos tamaños. Pero qué pasaría si hubiesen fenómenos que no se asociaran a una escala o un tamaño en particular, sino que tuvieran libertad, que fueran *libres de escala*.

Puesto que hacemos nuestras descripciones físicas en lenguaje matemático (a la Galileo), podemos preguntarnos cómo describir los cambios de escala en este lenguaje, lo que equivale a pensar en términos de transformaciones, en este caso, transformaciones de cambio de escala. Es decir, estamos frente a la búsqueda de invariantes ante cambio de escala, lo que nos evoca inmediatamente a los famosos principios de conservación, los cuales surgen al considerar los invariantes ante cierto tipo de transformaciones (espaciales y temporales), y que se resumen en el teorema de Noether, y que además definen una cierta geometría. En nuestro caso también la hay: la de los conjuntos fractales, como veremos más adelante.

En termodinámica tenemos ejemplos de cantidades que no dependen del tamaño, como la temperatura, y se les denomina variables intensivas. La entropía y la energía, entre otras, dependen del tamaño del sistema, y variables como éstas se las llama variables extensivas. Lo interesante para nosotros es que esa dependencia del tamaño

no es cualquiera imaginable, sino que se puede poner en términos de un tamaño patrón. Es decir, las variables extensivas se reproducen a sí mismas cuando se aumentan los tamaños, y lo que las modifica es un factor multiplicativo que sólo depende del factor de escalamiento.

Supongamos que medimos la cantidad  $x$  de un objeto, si ahora denotamos con primas a un objeto semejante (aumentado o disminuido)  $\lambda$  veces, tenemos que:  $x' = \lambda x$  es la transformación de cambio de escala. Sea  $F(x)$  una función de  $x$ . De lo dicho en relación a las variables extensivas, podemos decir que  $F$  es libre de escala si ante la transformación anterior se reproduce a sí misma excepto por un factor multiplicativo que depende del factor de escala  $\lambda$ . Es decir,  $F(x)$  será libre de escala si es de la forma:

$$F(\lambda x) = g(\lambda)F(x) \quad (1.2)$$

Pero habiendo expresado lo anterior, estamos ya completamente dentro del terreno de las matemáticas, pues una función que satisface (1.2), se dice que es homogénea por definición.

Vemos claramente que las propiedades de las variables extensivas se expresan en términos de funciones homogéneas, pues una función homogénea  $F(x)$  tiene la propiedad de que si conocemos su valor en un punto,  $x = x_0$ , y conocemos la forma funcional de  $g(\lambda)$ , entonces conocemos a la función en todo su dominio. Ello es consecuencia de que cada valor de  $x$  puede ser escrito en la forma  $\lambda x_0$ , y de que

$$F(\lambda x_0) = g(\lambda)F(x_0)$$

lo que nos dice que el valor de la función  $F(x)$  en cualquier punto está relacionado con el valor de  $F(x)$  en el punto de referencia  $x = x_0$  por un simple cambio de escala. Por supuesto que el cambio de escala es, en general, no lineal (a menos de que  $g(\lambda)$  sea lineal).

La forma general de la función  $g(\lambda)$  se puede encontrar de la siguiente manera: Primero notemos que el producto de dos cambios sucesivos de escala es:

$$F(\lambda\mu x) = g(\lambda)g(\mu)F(x)$$

Así que:

$$g(\lambda\mu) = g(\lambda)g(\mu)$$

Supongamos que  $g(\lambda)$  es diferenciable.<sup>2</sup> Si tomamos la derivada con respecto a  $\mu$ , encontramos

$$\frac{\partial}{\partial \mu} g(\lambda\mu) = \lambda g'(\lambda\mu) = g(\lambda)g'(\mu)$$

---

<sup>2</sup>En el caso más general no se necesita esta hipótesis, pero la prueba se simplifica con ella.

donde  $g'(\mu) = (\partial/\partial\mu)g(\mu)$ . Ahora hagamos  $\mu = 1$  y definamos  $g'(\mu = 1) \equiv p$ .

Entonces

$$\lambda g'(\lambda) = pg(\lambda)$$

que se puede reescribir como

$$\frac{g'(\lambda)}{g(\lambda)} = \frac{d}{d\lambda} \ln g(\lambda) = \frac{p}{\lambda} \quad (1.3)$$

De acuerdo con (1.2), tenemos que  $g(1) = 1$ ; y si integramos esta última ecuación (1.3), de 1 a  $\lambda$ , encontramos finalmente que:

$$g(\lambda) = \lambda^p$$

Que, como decíamos, en general no es lineal a menos de que  $p = 1$ . Por tanto,

$$F(\lambda x) = \lambda^p F(x) \quad (1.4)$$

y se dice entonces que  $F(x)$  es una función homogénea cuyo orden o grado es el exponente  $p$ . Si queremos conocer la forma de  $F(x)$ , hagamos  $\lambda = x^{-1}$  en la ecuación (1.4), para obtener que:

$$F(x) = F(1)x^p = Ax^p \quad (1.5)$$

con  $A = cte.$ , esto es, *toda función homogénea sigue una ley de potencia*. Y viceversa, pues todas las funciones tipo ley de potencia satisfacen la definición de función homogénea. De hecho, muchos autores usan indistintamente los términos ley de potencia, ley de escala, función homogénea y escalamiento.

No es casual que hayamos llegado a la homogeneidad a partir de las variables extensivas, pues de hecho, cuando se presenta la termodinámica de forma axiomática se postula que lo que llamamos variables intensivas son funciones homogéneas de orden cero, y las llamadas variables extensivas son funciones homogéneas de primer orden (ver por ejemplo [15]).

Otra forma de ver por qué una ley de potencia es libre de escala es hacer lo siguiente: Comparemos funciones del tipo (1.5), con  $p$  arbitrario, con alguna otra función y que representen alguna cantidad física. Sea  $x_0$  la unidad patrón para medir a  $x$ , entonces:

$$F(x) = A' \left( \frac{x}{x_0} \right)^p$$

Consideremos  $g(x) = \exp\left(\frac{x}{x_0}\right)$ , donde el argumento sólo puede ser adimensional (lo mismo se podría hacer con funciones trigonométricas). ¿Cuál es la diferencia entre las funciones  $F(x)$  y  $g(x)$ ?



1) Tomemos un intervalo logarítmico centrado en  $x_0$ ,  $(\frac{x_0}{3}, 3x_0) = (a, b)$ , y sea  $\Delta_g \equiv \frac{g_{\text{máx}}}{g_{\text{mín}}}$ , donde  $g_{\text{máx}} = g(b)$ ,  $g_{\text{mín}} = g(a)$ , entonces, para este intervalo  $\Delta_g = e^{\frac{8}{3}}$ . Si ahora cambiamos el centro a  $x'_0 = 9x_0$ , en el intervalo  $(\frac{x'_0}{3}, 3x'_0)$ , tendremos que ahora  $\Delta_g = e^{24}$ . En general, si cambiamos el centro,  $\Delta_g$  variará, de manera que las gráficas de  $g$  NO son modelos a escala unas de otras; no es posible superponerlas mediante un simple cambio lineal (una contracción o una dilatación) de escala de los ejes. No son gráficas semejantes.

2) Si ahora hacemos lo mismo con nuestra función  $F$ , encontramos que  $\Delta_F = 9^p$  para todo intervalo; ahora sí será posible superponer todas las gráficas realizando un simple cambio lineal, lo que se puede ver en la figura 1.3.

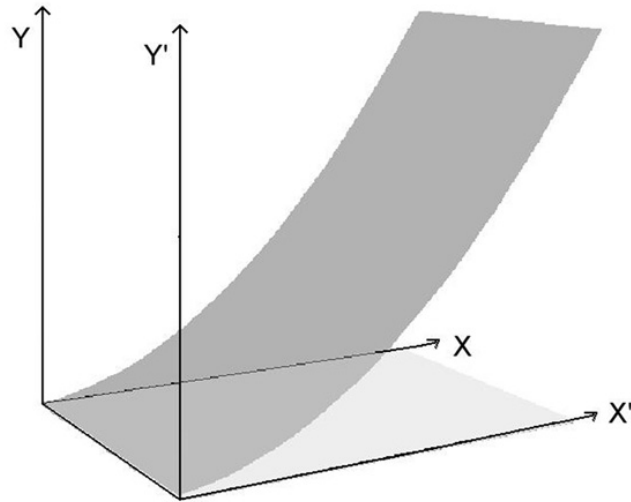


Figura 1.3: Superposicion de gráficas mediante un simple cambio lineal.

Por lo tanto, un fenómeno que se describe con una ley potencial, tiene el mismo aspecto, independientemente de la escala a la que se le mire; sus propiedades se hacen independientes de la escala de observación y por ello es libre de escala.

El concepto de homogeneidad se puede extender a funciones de un número arbitrario de variables. En el caso de  $n$  variables, se puede pensar que  $x$  es un vector  $n$ -dimensional, por lo que la ec. (1.1) se verá de la forma:

$$F(\lambda \vec{x}) = g(\lambda) F(\vec{x})$$

y por argumentos similares se puede probar nuevamente que  $g(\lambda) = \lambda^p$ .

Aun se puede generalizar un poco más, y tener funciones homogéneas generalizadas con  $n$  exponentes, en vez de uno solo. Si pensamos en el caso de una función homogénea generalizada de dos variables  $F(x, y)$ , ésta siempre se puede escribir de la forma

$$F(\lambda^p x, \lambda^q y) = \lambda F(x, y) \quad (1.6)$$

y así, la función está caracterizada por dos parámetros,  $p$  y  $q$ , que se conocen también como *exponentes de escalamiento*.

Se podría pensar que (1.6) tiene una generalización más, una expresión del tipo:

$$F(\lambda^p x, \lambda^q y) = \lambda^s F(x, y) \quad (1.7)$$

pero esto no es así, porque sin perder generalidad se puede tomar  $s = 1$  en (1.7), lo que la reduce a (1.6). Y también al revés, porque si se cumple (1.6), se cumple (1.7):

$$F(\lambda^p x, \lambda^q y) = F((\lambda^{p/s})^s x, (\lambda^{q/s})^s y) = F((\lambda^s)^{p/s} x, (\lambda^s)^{q/s} y) = \lambda^s F(x, y)$$

Vemos así que la forma más general es (1.6).

Una forma común y muy útil de usar a  $F(x, y)$  es la que se conoce como *representación de escalamiento*, para el caso de dos variables. Lo que se hace es tomar  $\lambda = y^{-1/q}$  en (1.6), entonces

$$F(x, y) = y^{1/q} F\left(\frac{x}{y^{p/q}}, 1\right)$$

de donde podemos definir  $F(z, 1) \equiv f(z)$  con  $z = \frac{x}{y^{p/q}}$  y  $F(x, y) = y^{1/q} f(z)$ , vemos que la función homogénea generalizada  $F(x, y)$ , depende de  $x, y$  sólo a través de la razón  $\frac{x}{y^{p/q}}$  (aparte del factor multiplicativo). Pero no es la única, ya que también podemos tomar  $\lambda = x^{-1/p}$  en (1.6) y tener:

$$F(x, y) = x^{1/p} F\left(1, \frac{y}{x^{q/p}}\right) = x^\alpha h\left(\frac{y}{x^\beta}\right) \quad (1.8)$$

con  $\alpha = 1/p$  y  $\beta = q/p$ . Como veremos, esta última expresión es muy útil en la descripción de las transiciones de fase de segundo orden.

## 1.4 Mecanismos que dan lugar a leyes de potencia

### 1.4.1 Transiciones de fase de segundo orden: fenómenos críticos

Hasta ahora no hemos hablado de ningún fenómeno físico que no se refiera a partículas puntuales y que dé lugar a leyes de potencia. Lo que ahora nos interesa exponer son una clase de fenómenos colectivos donde aparecen leyes de potencia y donde lo importante no son las partículas, sino que eventos de todas las escalas hacen pequeñas contribuciones de igual importancia; en otras palabras, donde lo importante es que una gran cantidad de elementos interactúan entre sí para producir un fenómeno en el que no se definen escalas típicas y que se les conoce como *fenómenos críticos*.

Supongamos que tenemos un sistema termodinámico constituido por un fluido, una sustancia pura, y que queremos describir la transición líquido-gas. Por definición, nuestro fluido deberá estar aislado de los alrededores, así que podemos pensarlo encerrado en un recipiente de volumen definido. Pensemos también que las paredes del recipiente son transparentes y que podemos ver hacia el interior. Algunas de las variables termodinámicas de este sistema son la presión, la temperatura, el volumen y la densidad media.

Una vez que tenemos confinado a nuestro sistema y que hemos fijado la presión, podemos empezar a variar la temperatura y ver lo que sucede (lo mismo se podría haber hecho fijando la temperatura y variando la presión). Desde luego que tal variación se realiza lo suficientemente lenta como para que la sustancia permanezca en equilibrio termodinámico. Para cada isobara, si variamos la temperatura y hacemos un registro cuidadoso de las mediciones, podemos hacer un diagrama de fases p-T, que será como el mostrado en la figura 1.4.

La región de nuestro interés es la región comprendida por arriba del punto triple. En este caso podemos olvidarnos de la transición sólido-gas. Si comenzamos con una temperatura mayor a la del punto triple  $T_{tr}$ , y trabajamos con isobaras, observamos que a ciertos valores de temperatura el fluido presenta dos fases bien definidas: una líquida y una gaseosa, separadas por una interface o menisco visible; en nuestro diagrama ello representa la que se conoce como curva de coexistencia, misma que separa las dos fases estables. Observamos también que cada fase tiene su propia densidad,  $\rho_l$ ,  $\rho_g$ , líquido y gas, respectivamente, y que conforme aumentamos la temperatura, para la isobara en la que nos movemos, éstas cambian. Para cada isobara vemos un comportamiento parecido, pero al llegar a la que podemos llamar isobara crítica, si variamos  $T$  como antes, al llegar al valor  $T_c$ , las densidades se igualan  $\rho_c = \rho_l = \rho_g$ , en ese momento se sigue dando la coexistencia entre las dos fases (estamos en el último punto sobre la curva de coexistencia) pero la interface desaparece. Es en ese momento que se dice que el sistema se encuentra en su *punto crítico*; a partir de ahí, para isobaras mayores, variando la temperatura se puede transitar de un estado a otro

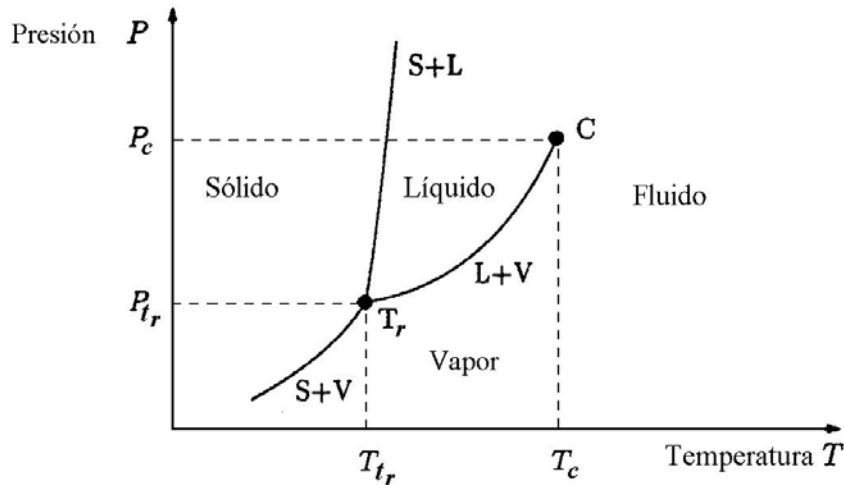


Figura 1.4: Diagrama p-T.

suavemente, sin distinción, se dice entonces que hay una continuidad de estado, es decir, no hay diferencias fundamentales entre ambas fases. Observamos también que en el punto crítico todo el sistema es invadido por una especie de niebla, conocida con el nombre de *opalescencia crítica* y explicable ópticamente mediante el fenómeno de dispersión óptica, lo que nos indica que existe una red cuyas distancias de separación son del orden de la longitud de onda de la luz visible. Pero el hecho de que invada todo el recipiente que contiene a nuestro fluido también nos indica que hay eventos de muchas escalas, todas ellas importantes. Se dice que la longitud de correlación se torna infinita, o sea, que no hay longitud característica del sistema.

Se puede decir que la fase fluida ( $T > T_c$ ) tiene una simetría espacial mayor, ya que no se observa una separación en líquido y gas. Por el contrario, cuando  $T < T_c$  sí hay una separación en líquido y gas, de manera que hay un rompimiento de simetría. Por ello, esas dos regiones ( $T > T_c$  y  $T < T_c$ ) deben ser descritas por diferentes variables termodinámicas, que no se pueden extender analíticamente a través del punto crítico (la idea orden-desorden es más clara si se piensa, por ejemplo, en un sistema ferromagnético, donde se forman dominios magnéticos a baja temperatura). Dada esa reducción de simetría, es necesario un parámetro extra para describir la termodinámica de la fase a baja temperatura, que mida la indistinguibilidad entre las fases, y que sea una especie de medida del orden existente en el sistema. Dicha cantidad se conoce como *parámetro de orden*. En nuestro caso es la diferencia de

densidades:

$$\Delta\rho = \rho_l - \rho_g$$

En general, un parámetro de orden se define como una cantidad no nula por debajo de un valor crítico y nula por arriba de ese mismo valor crítico y constituye una de las características comunes asociadas a los fenómenos críticos. Para el caso de sistemas ferromagnéticos la magnetización es el parámetro de orden, y existe el análogo a la opalescencia crítica antes descrita (para demostrar que se da el mismo fenómeno se recurre a la que dispersión de neutrones).

Cuando atravesamos la curva de coexistencia a través de una isobara, variando la temperatura, se dice que ocurre una transición de fase de primer orden. En tal caso, las variables del sistema son discontinuas y ocurre una reordenación radical de la estructura material. Por otro lado, cuando llegamos al punto crítico, se dice que ocurre una transición de fase de segundo orden; los ejemplos típicos de fenómenos críticos son estas transiciones en las que las cantidades que describen al estado macroscópico del sistema varían de forma continua pero presentan un punto particular en el cual las derivadas divergen y ocurre una reordenación suave de la estructura (la que da cuenta el parámetro de orden). Los puntos críticos son detectados por las singularidades en las variables termodinámicas, por la divergencia en la longitud de correlación y por la presencia de fluctuaciones estadísticas perceptibles en todas las escalas. Antes de pasar a la descripción de las leyes de potencia asociadas a los fenómenos críticos, señalemos un tercer ejemplo de ellos (además de la transición líquido-gas y la magnetización ferromagnética): el paso de un estado superconductor a conductor para ciertos metales a muy bajas temperaturas. Desde luego que no son los únicos, pero consideramos tener suficientes ejemplos.

Hemos llegado al tema de esta sección: cómo los fenómenos críticos dan lugar a leyes de potencia. Experimentalmente se encuentra que en una vecindad del punto crítico (la zona crítica), muchas de las cantidades interesantes desde el punto de vista termodinámico (entre ellas el parámetro de orden) siguen una ley de potencia,<sup>3</sup> en función de la cantidad adimensional  $\varepsilon \equiv \frac{T-T_c}{T_c}$ , denominada temperatura reducida y que mide las diferencias de temperatura respecto a la temperatura crítica:

$$\Delta\rho \sim |\varepsilon|^\beta \quad \text{si } T - T_c \longrightarrow 0$$

En términos generales, se tiene:

$$f(\varepsilon) = f_0 |\varepsilon|^\beta + \tilde{f}(\varepsilon)$$

donde  $\tilde{f}$  es la parte regular de  $f$  (si la hay), y no es necesariamente continua.

---

<sup>3</sup>Aunque no todas divergen como una ley de potencia; algunas tienden a cero y otras más son del tipo  $\ln|\varepsilon|$ .

A la cantidad  $\beta$  se le denomina *exponente crítico*. Este índice es tan importante que se tiene una definición general: supóngase que  $f(\varepsilon)$  es una función general importante, positiva y continua para  $\varepsilon$  suficientemente pequeño y positivo, y supóngase que el límite siguiente existe

$$\lim_{\varepsilon \rightarrow 0} \frac{\ln f(\varepsilon)}{\ln \varepsilon} \equiv \sigma \quad (1.9)$$

se dice entonces que  $\sigma$  es el exponente crítico asociado a la función  $f(\varepsilon)$ , lo que se denota por  $f(\varepsilon) \sim \varepsilon^\sigma$  (en el caso general se deben hacer correcciones en la vecindad de la singularidad).

Debemos resaltar que históricamente los fenómenos críticos se centraron en el estudio de los exponentes críticos, dado que:

- a) los eventos físicamente ricos que se observan en la zona crítica no aparecen lejos de ella;
- b) el término asociado al exponente crítico (la parte singular de  $f(\varepsilon)$ ) es el que experimentalmente domina en la zona crítica, por lo que el problema de describir fenómenos críticos se transforma entonces en determinar qué potencia domina;
- c) existen ciertas relaciones (desigualdades) entre los exponentes críticos (conocidas como *relaciones de escalamiento*) que trascienden al sistema particular en consideración, que se pueden deducir partiendo de termodinámica elemental junto con hipótesis adicionales, y que experimentalmente se cumplen como igualdades; esas relaciones son famosas porque se han asociado a los nombres de Fisher, Rushbrooke, Widom y Josephson.<sup>4</sup>
- d) las relaciones de escalamiento implican que no todos los exponentes críticos son independientes; y
- e) para sistemas físicamente diferentes (como un fluido y un ferromagneto) se observa un comportamiento idéntico; temperaturas críticas que difieren por varios órdenes de magnitud comparten los mismos exponentes; es decir, los exponentes críticos son independientes de la naturaleza física detallada del sistema: son *universales*, y se les considera ser parte de las propiedades *protegidas* por la materia.

En términos de leyes de potencia, las cantidades termodinámicas en la zona crítica, se ven así:

- el parámetro de orden (diferencia de densidades en fluidos y diferencia de magnetización en sistemas ferromagnéticos)  $M \sim |\varepsilon|^\beta$  (con  $H = 0$ );

<sup>4</sup>Para revisar la deducción, consultar [64].

- la longitud de correlación  $\xi \sim |\varepsilon|^{-\nu}$ ;
- $\chi \sim \frac{1}{|\varepsilon|^\gamma}$  que representa la compresibilidad isotérmica en fluidos, y la susceptibilidad isotérmica en ferromagnetos;
- capacidad calorífica  $C \sim |\varepsilon|^{-\alpha}$  (con  $H = 0$ );
- ecuación de estado ( $\varepsilon = 0$ )  $M \sim H^{-1/\delta}$

Es importante mencionar que la característica del inciso e) antes descrita también ha llevado a los físicos a plantear la hipótesis de universalidad del punto crítico, que supone que sólo dos cantidades determinan el comportamiento crítico de muchos sistemas: la dimensión del espacio,  $d$ , y la dimensión del parámetro de orden,  $n$ . Así, todos los sistemas que tengan los mismos  $d$  y  $n$  tienen los mismos exponentes críticos y se dice que son miembros de la misma *clase de universalidad*.

El comportamiento tipo ley de potencia de la zona crítica no se entendía completamente hasta hace relativamente poco, pues las teorías de campo medio si bien suministran valores numéricos, discrepan de los obtenidos experimentalmente. Así, en 1965, a Widom se le ocurrió pensar que dado que en la zona crítica existen singularidades, habría una contribución singular que es la físicamente relevante (habría además una contribución regular) y que podría describirse por medio de las funciones homogéneas. Formalmente planteó la hipótesis de escalamiento estático (HEE) o escalamiento de Widom, aseverando que tanto la energía libre de Gibbs<sup>5</sup> como la función de correlación, son funciones homogéneas generalizadas, es decir, (en el caso de Gibbs) que existen dos parámetros,  $a_\varepsilon$ ,  $a_H$ , tales que:

$$G(\lambda^{a_\varepsilon} \varepsilon, \lambda^{a_H} H) = \lambda G(\varepsilon, H)$$

Propiedad que inmediatamente se extiende a los otros potenciales, ya que la homogeneidad se conserva ante las transformaciones de Legendre.<sup>6</sup> A partir de lo anterior, todos los exponentes críticos pudieron ser expresados en términos de los exponentes de escalamiento  $a_\varepsilon$ ,  $a_H$ . O si se quiere, en términos de los exponentes críticos antes descritos, y en la representación de escalamiento, se tiene:

$$G(\varepsilon, H) = |\varepsilon|^{2\beta+\gamma} G\left(1, \frac{H}{\varepsilon^{\beta+\gamma}}\right)$$

Lo que implica que una de las cantidades medibles, digamos  $M(\varepsilon, H)$ , se vea de la forma  $M(\varepsilon, H) = |\varepsilon|^\beta M\left(\varepsilon/|\varepsilon|, |\varepsilon|^{-\beta-\gamma} H\right)$ . Experimentalmente eso significa que si graficamos el producto  $|\varepsilon|^\beta$  contra  $|\varepsilon|^{-\beta-\gamma}$ , todos los datos se ajustan a una sola curva.

<sup>5</sup>Se escoge el potencial de Gibbs, dado que el cálculo de los exponentes se simplifica.

<sup>6</sup>En [15] se hace un recuento de las propiedades de las transformaciones de Legendre.

La HEE significó dar un paso adelante en la comprensión de los fenómenos críticos, pues aunque por sí sola no suministró valores numéricos para ninguno de los exponentes críticos, restringe el número de exponentes independientes a tan sólo dos; reproduce las relaciones de escalamiento y hace predicciones específicas concernientes a la forma de la ecuación de estado. Fue hasta que se introdujo el Grupo de Renormalización que se pudo explicar por qué la energía libre (y por ende los demás potenciales termodinámicos) tiene las propiedades de escalamiento, además, entonces se determinaron los valores numéricos de los exponentes que concuerdan de manera sorprendente con los hallados experimentalmente.

### 1.4.2 Objetos fractales

La geometría de los objetos naturales es central en los modelos que hacemos de ellos para comprenderlos, tanto la que asociamos a los que existen en la escala atómica como la de los que son del tamaño del universo. Anteriormente hablamos de las invarianzas, y dijimos que una consecuencia de las simetrías del espacio era la existencia de cantidades que se conservan. Para el caso de transformaciones de escala los objetos geométricos asociados son los famosos fractales (debemos decir que hay todo un debate acerca del tipo de simetrías asociadas, que se puede ver en [20]).

No haremos un recuento del surgimiento de la geometría fractal, ya que existe una gran producción literaria al respecto; tan sólo diremos que la geometrización euclidiana es insuficiente para describir las invarianzas ante cambio de escala. Y eso se debe a que si una estructura geométrica es invariante ante un cambio de escala, entonces tiene la propiedad de *verse igual* “de cerca que de lejos”. En otras palabras, si comparamos una parte de esa estructura (o también de un proceso) con el todo y *parecen* ser iguales, decimos que son autosemejantes (o autosimilares), “la parte es *semejante* al todo”, y por ese hecho, son invariantes ante cambios de escala.<sup>7</sup> En general, las formas clásicas de la geometría no son autosimilares; si un círculo es aumentado lo suficiente, su perímetro se verá como una línea recta; fue por ello que la gente creía que el mundo era plano, pues la Tierra se ve así a la escala humana.

El concepto de autosemejanza ha surgido por doquier, como lo constatan los siguientes ejemplos:

Primer ejemplo. Sorprendentemente en un poema de la autoría de Swift, se hace alusión a la idea de la autosemejanza. Describe pulgas plagadas por pulguitas (ver la figura 1.5), que a su vez están plagadas por pulguitas más pequeñas, en un proceso *ad infinitum*. Contemporáneo del descubrimiento de la vida microscópica, Swift

---

<sup>7</sup>A la invarianza de escala también se le denomina *invarianza bajo similaridad interna* o algunas veces invarianza de dilatación (*dilation invariance*). Para más detalles, consultar [27].



probablemente encontró en ello una fuente de inspiración para canalizar sus motivaciones político-sociales y dar rienda suelta a sus dotes literarios, mismos que se ven plasmados en su poesía.<sup>8</sup>

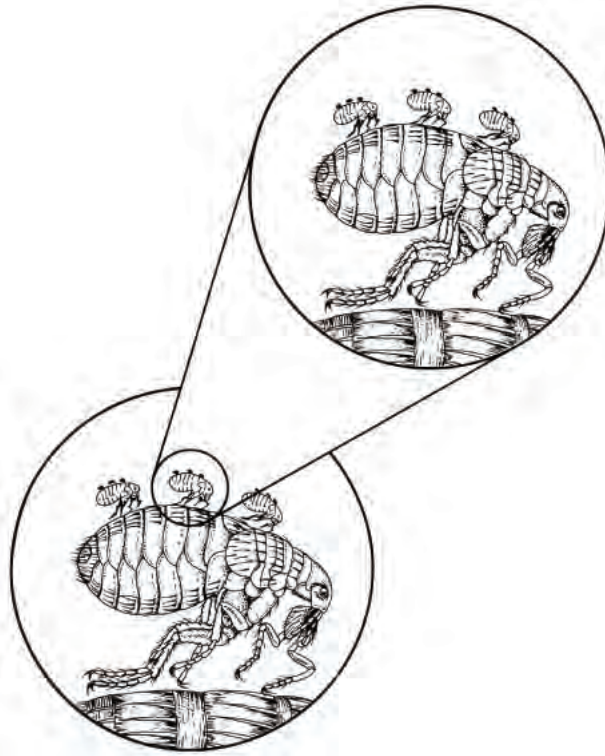


Figura 1.5: Las pulgas descritas por Swift.

Segundo ejemplo. Considérese un conjunto cerrado, digamos, un segmento de recta  $S$ . Tal segmento no tiene simetría de traslación, ya que si lo movemos obtenemos otro conjunto de puntos. Sin embargo, si cambiamos su longitud por un factor  $r$  menor que 1 (en tal caso miniaturizamos, encogemos, o escalamos hacia abajo), podemos generar un nuevo conjunto de puntos  $S' = r(S)$  que será un ‘segmentito’, el cual puede ser duplicado hasta cubrir todo el segmento original  $S$ . Si elegimos apropiadamente el factor de escala  $r$ , podremos cubrir al segmento original con  $N$  segmentitos que no se traslapen. Decimos entonces que el conjunto  $S$  es autosimilar con respecto al factor

<sup>8</sup>*So, naturalists observe, a flea/ Has smaller fleas that on him prey;/ And these have smaller still to bite'em;/ And so proceed ad infinitum. On Poetry, a Rhapsody.* J. Swift, 1733. Esta descripción después la reprodujo Lewis Richardson en 1922 cuando describió ‘cascadas de energía’ mientras estudiaba turbulencia.

de escala  $r$ . Podemos usar como factor de escala  $r(N) = (\frac{1}{N})^{1/d}$  para construir rectas, cuadrados o cubos autosemejantes (con  $d = 1, 2, 3$ , respectivamente). Esto se puede ver en la figura 1.6.

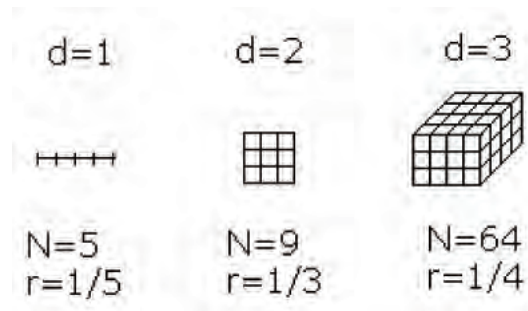


Figura 1.6: Autosemejanza para entes euclidianos.

Los ejemplos anteriores dan cuenta de una de las propiedades más importantes de los fractales: la autosemejanza, que de hecho es la propiedad que muchos autores emplean para definirlos. Ello implica que un fractal no tiene ninguna escala característica: todas las escalas son “buenas”, pues posee detalle a *todas* las escalas de observación. Y para el caso de fractales físicos, las escalas deben limitarse a un cierto rango (una escala mínima, límite microscópico, y otra máxima, tamaño de la muestra, longitud de correlación) que depende del objeto en consideración. Se argumenta que no es necesario “que la forma sea la misma para escalas de diferente longitud, sino sólo que el grado de irregularidad o de fragmentación es idéntico en todas las escalas”. Esto es lo que ocurre usualmente, porque el que la estructura (o proceso) se repita a sí misma exactamente no ocurre en la naturaleza. Si la autosemejanza sólo se repite estadísticamente, se habla de fractales aleatorios o estadísticos como el de la figura 1.7. Conviene aclarar que sólo cuando la autosimilaridad se presenta en muchos órdenes de magnitud (haciendo *zoom-in* o *zoom-out*), es cuando se habla propiamente de fractales; en otro caso tendremos sólo autosimilaridad en cierto grado, lo que recuerda un cierto tipo de recurrencia.

Una vez que se ha proyectado intuitivamente lo que son los objetos fractales lo que se requiere es poder caracterizarlos cuantitativamente y, para hacerlo, hay que asociarles algún valor numérico a cada uno, que describa su estructura y que permita diferenciarlos. En otras palabras, habrá que medirlos de alguna forma, lo que matemáticamente implica una cierta métrica y algunas relaciones topológicas.

Un método común de medir la longitud, área o volumen de un objeto consiste en cubrirlo con segmentitos (como los descritos en el segundo ejemplo anterior) o cajas cuya longitud, área o volumen, son tomados como unidad de medida.

Si  $\varepsilon$  es el lado de la caja unitaria,  $N$  es el número de cajas que cubren al objeto,

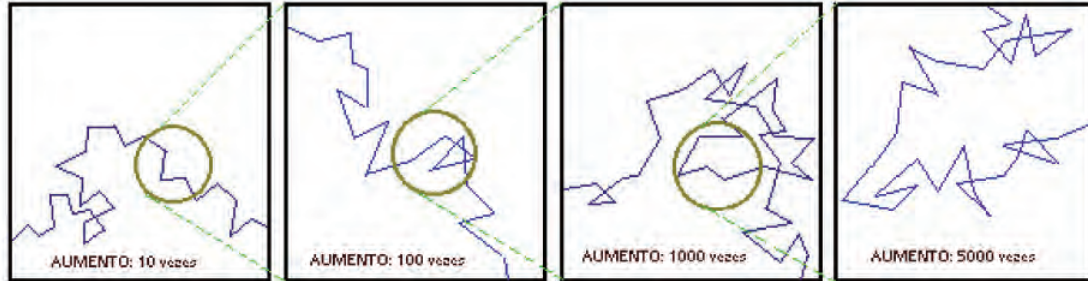


Figura 1.7: Fractal aleatorio.

y  $d$  es la dimensión euclidiana entonces la *medida*  $M$  del cuerpo es  $M = N\varepsilon^d = N\mu$ , donde  $\mu$  es la medida de un segmentito, cuadrado o cubito unidad, según sea el caso.

Muchos matemáticos (Cantor, Carathéodory, Peano, Lebesgue, etc.) demostraron que existen objetos *patológicos* para los que el método anterior falla y cuyo desarrollo generó un área muy interesante de las matemáticas, la teoría de la medida, pero tales desarrollos están fuera del alcance de este trabajo.

Un ejemplo muy citado de esos objetos patológicos, y que es fractal, es el que surge cuando se intenta medir la longitud de la costa de una isla: No es difícil convencerse de que a cada escala de observación aparecen cada vez más detalles, lo que implica una medición cada vez más refinada. Para una regla fija cuya unidad es  $\varepsilon_1$ , se tiene una cierta longitud  $L(\varepsilon_1)$ , pero si se hace que la unidad de medida (nuestra regla) se vaya haciendo cada vez más pequeña, evaluando sucesivamente la longitud, cada vez de forma más precisa, lo que se encuentra es que la longitud estimada no sólo es extremadamente grande, sino que de hecho es mejor considerarla infinita. Se tiene la siguiente sucesión:

$$\text{Si } \varepsilon_1 < \varepsilon_2 < \varepsilon_3 < \dots < \varepsilon_n \rightarrow 0 \text{ con } L(\varepsilon_1) < L(\varepsilon_2) < L(\varepsilon_3) < \dots < L(\varepsilon_n) \rightarrow \infty$$

De tal manera que prácticamente la longitud de la costa depende del tamaño de la unidad de medida, y si uno quisiera comparar diferentes costas desde el punto de vista de su ‘extensión’, la longitud será inadecuada.

Por lo anterior, se hace necesario el empleo de otra cantidad para caracterizar a los fractales. Se necesita entonces encontrar un sustituto. Históricamente fue esa búsqueda lo que motivó las diferentes definiciones<sup>9</sup> que existen para la *dimensión fractal*. Todas buscan caracterizar cuantitativamente las propiedades fractales, pero no

<sup>9</sup>Encontramos el método de conteo de cajas (*box-counting*), el método de Hausdorff-Besicovitch, el método del embutido, el de cubiertas y el de bolas disjuntas, por citar algunos.

todas son equivalentes entre sí, y tampoco caracterizan de forma única a un fractal. Suponen la noción de distancia entre dos puntos y un paso al límite, lo que desde el punto de vista físico muchas veces es complicado,<sup>10</sup> o de plano, algunas de ellas no tienen significado físico; otras son más precisas, o más apropiadas. Pero en lo que a nosotros respecta, varias de ellas se relacionan de alguna forma con las leyes de potencia, como se muestra en los siguientes ejemplos:

I. Cuando hablamos de la autosemejanza presentamos como ejemplo objetos que se miniaturizaban por un factor  $r(N) = (\frac{1}{N})^{1/d}$ , y equivalentemente se tiene  $Nr^d = 1$  para objetos clásicos, lo que implica que la dimensión de tales objetos es:  $d = \frac{\ln N}{\ln(1/r)}$ . Podemos generalizar la idea y ahora considerar que se tiene un objeto fractal, por lo que le podemos asignar una dimensión fractal  $D$  igual a:

$$D = \frac{\ln N}{\ln(1/r)} = \frac{\ln N}{\ln(\lambda)}$$

donde  $\lambda$  es el factor de escala tal como lo definimos al inicio de la sección 1.1 (longitudes del objeto grande entre las del objeto pequeño) y  $N$  es el número de ‘segmentos’ necesarios para cubrir al objeto escalado. Lo anterior también es equivalente a pedir que:

$$\lim_{N \rightarrow \infty} Nr^d = 1$$

Tres casos se observan en la figura 1.8:

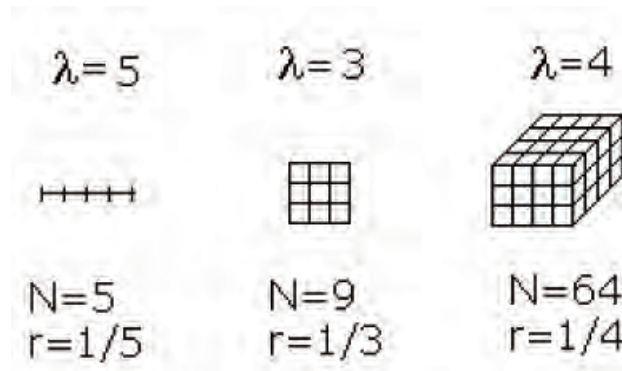


Figura 1.8: Primer ejemplo.

II. Si para objetos geométricos se considera la masa como el número de puntos, entonces se puede definir la densidad lineal (superficial o volumétrica) como el número

<sup>10</sup>Una observación curiosa es que la definición del exponente crítico (1.9), al menos por la forma, es casi idéntica a una de las dimensiones fractales, la del conteo de cajas (*box-counting*).

de puntos por unidad de longitud (área o volumen). En el caso de objetos geométricos clásicos la densidad está uniformemente distribuida, por lo que si amplificamos las dimensiones lineales de un volumen por un factor  $b$ , la masa contenida en este volumen se multiplica por el factor  $\gamma$ , teniendo una relación de escalamiento

$$M(bL) = \gamma M(L)$$

Para superficies o volúmenes ordinarios  $\gamma = b^d$ , donde  $d$  es la dimensión del objeto, y de esa forma se logra hacer que la densidad sea constante. En otras palabras, como cuando hablamos de Galileo y la vaca, si queremos que su densidad sea constante, entonces su masa deberá crecer como la dimensión del ‘volumen’ de la vaca, ya que ésta puede vivir en una recta, una superficie o un volumen.

Si estas ideas se generalizan para objetos fractales se tendrá la relación de escalamiento:

$$M(bL) = b^D M(L)$$

donde  $D$  es otra dimensión fractal, y que define un método directo para calcularla. Además  $D < d$ , donde  $d$  es la dimensión del espacio que contiene al objeto, en cuyo caso la densidad de un conjunto fractal decrece de manera indefinida.

Debemos aclarar que en el caso general las dos dimensiones que se definieron en los últimos dos ejemplos NO darán el mismo resultado, aunque para el caso de objetos matemáticos exactos sí pueden llegar a coincidir.

Apliquemos lo anterior al cálculo de la dimensión fractal de la curva de Koch (1906), que apareció como contraejemplo a las curvas rectificables, pues es una curva no diferenciable en todo punto y por ello considerada patológica.

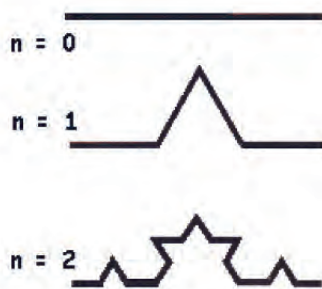


Figura 1.9: Algunos pasos en la construcción de la curva de Koch.

Se construye de la siguiente manera: Tomemos un segmento de recta de longitud  $L$ , que corresponderá al valor  $n = 0$ ; se le denomina iniciador. Dividámosla en tres partes

y sustituyamos su tercio central por dos segmentos de longitud  $1/3$  de  $L$ , haciendo que formen un triángulo equilátero (pero sin la base, como se ve en la figura 1.9); este paso, asociado con  $n = 1$ , proporciona lo que se conoce como generador. El siguiente paso consiste en repetir el proceso anterior en cada uno de los segmentos, es decir, hacer una copia a escala del generador en cada uno de los segmentos en que se divide la etapa en particular. Repitiendo el proceso obtendremos una sucesión de objetos, que se les asocia un número ( $n$ ), dependiendo de en qué etapa nos encontremos; se les denominan *prefractales*, y en este caso son autosemejantes discretamente. Finalmente, el objeto fractal será el que se obtiene cuando el proceso se realiza infinitas veces.

Calculemos ahora sí sus dimensiones fractales: A partir del primer método vemos que se necesitan 4 generadores para cubrir la iteración  $n = 2$ , y  $N = 4$ , y en ese caso el factor de escala es 3, por lo que  $D = \frac{\ln 4}{\ln 3}$ . Usando el segundo método, tenemos que  $M(3L) = 4M(L) = 3^D M(L)$  de donde podemos extraer la dimensión fractal  $1 < D = \frac{\ln 4}{\ln 3} = 1.26... < 2$ . En este caso (un fractal exacto) sí se obtuvo el mismo valor de la dimensión fractal, pero en general, lo recalamos, no lo son.

Además esta curva se puede considerar como un ejemplo simplificado de la costa de una isla, por lo que se puede intentar calcular su longitud. Se tienen las siguientes sucesiones (donde  $L$  es la longitud del iniciador):

$$\begin{aligned} \varepsilon_1 &= L/3 \rightarrow L_1 = 4\varepsilon_1 \\ \varepsilon_2 &= L/3^2 \rightarrow L_2 = 16\varepsilon_2 \\ &\quad \dots \\ \varepsilon_n &= L/3^n \rightarrow L_n = 4^n \varepsilon_n \end{aligned}$$

si eliminamos a  $n$  de las últimas dos expresiones, tendremos que:

$$L_n = L^D \varepsilon_n^{1-D}$$

con  $D = \frac{\ln 4}{\ln 3}$ , lo que quiere decir que para una regla de longitud  $\varepsilon_n$  fija, la longitud de la curva crece como ley de potencia de la longitud inicial  $L^D$ , y diverge si  $\varepsilon_n \rightarrow 0$ . Vemos también que si dividimos por  $\varepsilon_n$  lo que obtenemos es también una relación de escalamiento

$$\frac{L_n}{\varepsilon_n} = f\left(\frac{L}{\varepsilon_n}\right) = \left(\frac{L}{\varepsilon_n}\right)^D$$

Si se hace la gráfica log - log de la variable  $A(= \frac{L_n}{\varepsilon_n})$  contra  $z(= \frac{L}{\varepsilon_n})$  se encuentra una relación lineal, cuya pendiente es  $D$ , la dimensión fractal. Este último procedimiento de regresión lineal es lo que se hace muchas veces cuando se tiene un conjunto de datos experimentales y se tienen fuertes sospechas de que existe una ley de potencia;

se considera que la aparición de una línea recta en la gráfica log-log es la firma característica de una ley de potencia, pero para asegurar que un cierto conjunto de datos se ajusta de manera confiable a una ley de potencia deben aplicarse algunos métodos poderosos, como el de *maximum likelihood* o el *remuestreo* (para una descripción detallada, se sugiere revisar [51] y [60]).

# Distribuciones

---

... En el presente caso –y realmente en todos los casos de escritura secreta– la primera cuestión se refiere al lenguaje de la cifra, pues los principios de solución, en particular tratándose de las cifras más sencillas, dependen del genio peculiar de cada idioma y pueden ser modificadas por éste. En general, no hay otro medio para conseguir la solución que ensayar (guiándose por las *probabilidades*) todas las lenguas que os sean conocidas, hasta encontrar la verdadera ...

Edgar Allan Poe, *El escarabajo de oro*.

En el capítulo anterior nos aproximamos a algunos de los campos más explorados donde surgen leyes de potencia: los fenómenos críticos y los fractales; aún así, no hemos abarcado todos los puntos de vista desde los que se pueden mirar a las leyes de potencia. Las exploramos desde la visión de la termodinámica de equilibrio (fenómenos críticos) y desde el punto de vista geométrico-espacial (fractales). Vimos varias de sus analogías, pero no nos adentramos en el manejo de las herramientas matemáticas con que se les trabaja. En este capítulo haremos un recuento de diversos fenómenos que usualmente se les aproxima mediante un enfoque aleatorio, lo que permite aplicar consideraciones del tipo de las que usualmente se hacen en física estadística: ante un gran número de elementos, y a pesar de las diferencias internas de cada uno, sus regularidades colectivas no dependen de ellos.

## 2.1 Funciones homogéneas como distribuciones

En el capítulo anterior definimos lo que es una ley de potencia y describimos varias de sus propiedades matemáticas, la más importante: la libertad de escala. Supongamos ahora que tenemos una función homogénea de la variable  $x$ , definida en el intervalo



$I$ , y que pretendemos represente una densidad de probabilidad, entonces se deben satisfacer las siguientes condiciones (tanto para el caso continuo como para el discreto):

a)  $f(x) \geq 0, \forall x \in I$

b)  $\int_I f(x)dx = 1$

De la primera condición se sigue que las funciones potenciales con exponente impar no pueden usarse en todo el eje real, sino que el intervalo de definición se debe restringir a un subconjunto apropiado de los números reales; y de la segunda condición, conocida como la condición de normalización, para evitar divergencias, se debe restringir el intervalo dependiendo del valor del exponente. Una vez fijado el exponente, éste condiciona las características de los momentos de la función, que bien pueden no existir o hacerse infinitos.

## 2.2 Contexto sociohistórico

Fue durante el siglo XIX que se logró la *domesticación del azar*,<sup>1</sup> implicando por tanto que la probabilidad y estadística se unieran en maridaje y que de forma dialéctica, pero con mucho afán (lo que se constata en múltiples sucesos y trabajos de la época), se buscaran regularidades estadísticas que fueran susceptibles de ser explicadas a partir de teorías probabilísticas.

Acostumbrados a las estadísticas, se nos hace muy normal hablar de ellas en relación a casi todo lo que se nos antoje, pero ello no siempre ha sido así. Los habitantes decimonónicos gradualmente se fueron enfrentando a una cantidad cada vez mayor de este tipo de datos. El levantamiento de censos de población se hizo periódico pues se buscaba información acerca del número de habitantes, la distribución de éstos en el país, su ocupación, sus ingresos económicos, edades, el tipo de enfermedades contagiosas más frecuentes y su propagación, la cantidad de enfermos mentales y delincuentes de una población dada, la tasa de natalidad y mortandad, etc. pues todo ello representaba –representa– información valiosa a considerar en la determinación de políticas estatales. Dos razones principales motivaron esa búsqueda sistemática: primera, para el siglo XIX la física ya se había convertido en el paradigma a seguir de todo lo que se considerase científico, implicando que las disciplinas que aspirasen a considerarse científicas habrían de introducir en sus descripciones el lenguaje propio de la física, es decir, deberían matematizarse y medir. Segundo: dado que es la época de consolidación de los estados nacionales europeos, se dio toda una teorización acerca de las formas de consolidar a un Estado fuerte y moderno, y una manera de lograrlo, se decía, era conociendo con el mayor detalle posible las características de una nación, fueran sus riquezas naturales o su población. Fue así que de manera frenética,

<sup>1</sup>Se recomienda la sabrosa lectura del libro de Ian Hacking, con ese título, referido en [29].

sistemática e institucional, se comenzaron a recopilar datos que describieran lo mejor posible las características de una nación.

A su vez, la gran cantidad de datos recopilados y sistematizados institucionalmente motivó la investigación teórica tanto en probabilidad como en estadística. Una manera de percatarnos acerca de la extrapolación y uso social de ese conocimiento aún hoy es a partir de las palabras que empleamos. Por ejemplo, decimos que el desarrollo de un niño es ‘normal’ si cumple con ciertas características de peso, estatura, movilidad, agudeza visual, etc. y no reparamos en el hecho de que al hacerlo estamos empleando implícitamente ciertas ideas estadísticas, a saber, las ideas gaussianas acerca de la distribución normal; de hecho, la palabra *estadística* refleja la estrecha relación que había entre la recopilación de datos y su principal usuario y recolector: el Estado.

Bajo la anterior perspectiva histórica, se comprende la aparición de las regularidades de las que ahora queremos hablar. Podemos decir de forma vaga que fenomenológicamente comparten la característica de que existen eventos raros que aparecen muy de vez en cuando y otros que son muy comunes y que se presentan muy seguido; precisando: si consideramos sus ‘tamaños’ y vemos su distribución de acuerdo a su frecuencia en orden decreciente, encontramos que de entre todas las funciones monótonamente decrecientes que los pueden describir, son justamente las leyes de potencia las que las aproximan muy bien en un cierto intervalo, que abarca sorprendentemente varios órdenes de magnitud. A esta última propiedad (su gran extensión en órdenes de magnitud) se le denomina *robustez*, y se entiende mejor si se compara con lo que ocurre con los fenómenos críticos, donde la región que se puede describir con escalamiento es tan sólo alrededor de *un punto*, el punto crítico. Veremos aquí que este tipo de distribuciones aparece en fenómenos donde cabría esperar una ausencia de regularidades y un dominio del azar, dada la multiplicidad de factores involucrados.

Como hemos visto, la firma característica de este tipo de escalamiento se busca en las graficaciones log-log entre las variables involucradas, pero ahora nos centraremos en bases de datos estadísticas, a las que se les puede asociar conceptos tales como tamaño y frecuencia, y si se encuentra un buen ajuste lineal que los relacione, habremos reconocido dicha firma. Varios nombres se asocian a este tipo de descripciones, los más famosos son, sin duda, los de Pareto y Zipf, pero no son los únicos, como veremos en el capítulo siguiente. En lo que sigue iremos presentando ejemplos de leyes de potencia fenomenológicas y a la par expondremos algunas de las maneras que se han intentado para modelarlas.

### 2.3 Meteoritos y terremotos

Nuestro primer ejemplo de este tipo de distribuciones que siguen una ley de potencia viene de información importante para nuestra especie: la distribución de los meteoritos

que impactan la Tierra. Afortunadamente, la mayor cantidad de meteoritos que logran tocar la superficie de nuestro planeta son pequeños, del orden de micras; una cantidad menor de los del tamaño de centímetros, y los grandes meteoritos, como el que se cree que causó la extinción de los dinosaurios, tardan mucho tiempo en impactar la superficie terrestre.

Otra distribución que se puede relacionar con catástrofes y pérdidas humanas es la llamada ley de Gutenberg-Richter, usada en sismografía, y que constituye la observación empírica de que los terremotos ocurren con una distribución tipo ley de potencia. El núcleo de la Tierra, combándose bajo los esfuerzos de las placas circundantes produce terremotos, lo que diariamente genera pequeños temblores en alguna parte del planeta que son muy débiles para ser detectados sin instrumentos; los temblores más grandes, que pueden hacer vibrar los trastes o abrir las puertas son menos comunes y los grandes terremotos, que causan grandes tragedias, ocurren muy de vez en cuando.

## **2.4 Distribución de ingresos. Ley de Pareto**

La naturaleza, causas y consecuencias de las desigualdades económicas han preocupado a los economistas al menos desde Adam Smith, pero los economistas clásicos discutían en una completa ignorancia de los hechos. Hubo un cambio dramático hacia el final del siglo XIX, pues constituida en paradigma, la física inspiró la medición exhaustiva en muchas áreas del conocimiento; entre las que más pronto adoptaron sus lineamientos se encuentra la economía.

El éxito de la estadística en el estudio de un amplio rango de fenómenos en biología y psicología sentó las bases para un tratamiento inductivo de la economía, en general, y de la distribución de ingresos, en particular. El largo debate sobre la distribución de ingresos fue un punto de partida natural para un tratamiento estadístico serio, pero el debate se esculpió con el surgimiento de los movimientos socialistas y las uniones de trabajadores en Europa Occidental y Estados Unidos, y la pasión se encendió por todos lados: un tratamiento estadístico inductivo prometía respuestas científicas a las cuestiones centrales de entonces. La economía comenzaba a desarrollarse, pero su estado era el de una ciencia sin validación empírica, ya que muchos economistas defendían sólo en el discurso la necesidad de verificar sus teorías con la investigación empírica. Muy pocos se aventuraban a coleccionar datos relevantes. Fue el economista italiano Vilfredo Pareto (1848-1923), con sus investigaciones sobre la distribución de ingresos, quien primero mostró lo que sería una economía inductiva. Consideró datos de Inglaterra, un buen número de ciudades italianas, varios estados germanos, etc. y graficando las distribuciones de ingresos acumuladas para estas geografías en un papel doble logaritmo, encontró que en cada caso el resultado era una línea recta

con aproximadamente la misma pendiente. Resumió sus resultados mediante la curva asociada a su nombre.

En su *Cours d'économie politique* (1896), refiriéndose a la distribución del ingreso, presenta la que hoy se conoce como ley de Pareto: se preguntó por la cantidad de hogares cuyo ingreso anual excede o es igual a un valor dado  $x$ , y encontró que del ajuste de sus gráficas se cumplían expresiones lineales del tipo  $\log N = A - B \log x$ , donde  $A$  es un parámetro y  $B$  es (el valor absoluto de) la pendiente, cuyos valores eran cercanos a 1.5.

Para varios conjuntos de datos encontró que las líneas eran paralelas. Y aunque no empleó medidas sobre la bondad de los ajustes, la inspección visual sugiere que las líneas son buenas aproximaciones. Considerando sus estudios comparativos para diferentes sociedades, la constancia de la pendiente y la calidad del ajuste, afirmó que existe una 'ley de ingresos' de la distribución.

Puesto en términos de leyes de potencia, podemos decir que Pareto se preguntó por la cantidad de personas (u hogares) cuyo ingreso anual excede o es igual a un valor dado  $x$ , y encontró una descripción a partir de la función de distribución siguiente (ahora denominada paretiana o de Pareto):

$$P[X \geq x] = \left(\frac{x}{k}\right)^{-\alpha} \quad (2.1)$$

para algún  $\alpha > 0$  y  $k > 0$ ; y como ya decíamos líneas más arriba, en este caso se requiere restringir el intervalo de los números reales a valores tales que  $X \geq k$ , y donde  $k$  se interpreta como un ingreso mínimo o umbral. Como veremos, algunos modelos se concentran en darle una interpretación a esta cantidad umbral.

La densidad de probabilidad para la distribución de Pareto, y que representa cuánta gente tiene un ingreso de exactamente  $x$ , es la derivada de (2.1):

$$P[X = x] = f(x) = \alpha k^\alpha x^{-(\alpha+1)} \sim x^\beta$$

y usualmente  $0 < \alpha \leq 2$ , en cuyo caso la variable aleatoria  $X$  tiene varianza infinita, y si además  $\alpha \geq 1$ , entonces  $X$  también tiene media infinita.

A la distribución de frecuencias correspondiente a su ley, Pareto la llamó 'pirámide social'. Empieza en un ingreso mínimo y declina monótonamente en adelante, como se puede ver en la gráfica de la figura 2.1. Desde que la encontró, Pareto enfatizó el gran carácter asimétrico y por tanto, su diferencia fundamental con una curva normal.

No hay evidencia sobre exactamente cómo fue que Pareto llegó a su formulación. Algunos sugieren que el papel logarítmico doble era de uso común entre los ingenieros de la época, y dado que Pareto había practicado dicha profesión por algunos años, quizás ello le haya motivado a elegir tal papel de graficación. Lo que sí es un hecho es que la belleza matemática de la física newtoniana tuvo sobre él una influencia

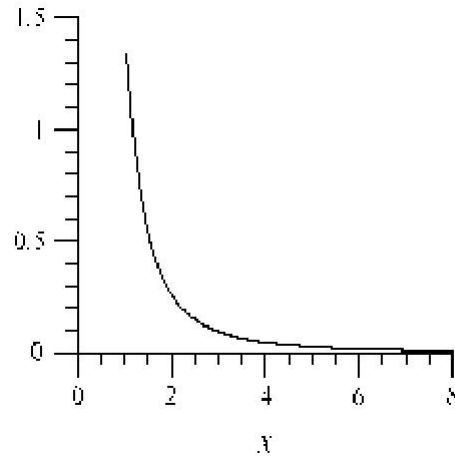


Figura 2.1: Distribución paretiana.

tremenda, y que deseaba convertir a la economía en una ciencia exacta, que se pudiera describir por leyes comparables en su universalidad a las formuladas por Newton.

Emocionado por su nueva ley, Pareto trató de explicarla. Según él, las posibles fuentes de desigualdad de ingresos las constituyen las oportunidades diferentes, las instituciones sociales y la naturaleza humana. Argumentó en su *Cours* que su ley prevalece aún después de profundos cambios en la organización económica, por lo que debería depender, en primer lugar, de la naturaleza humana, así que intentó dilucidar sobre qué es lo que en ésta determina tal distribución. A veces argumentaba que correspondía a la distribución subyacente de las capacidades humanas, truncadas por las necesidades de sobrevivencia; aquellos con capacidades considerablemente menores que la media debían morir o ser mantenidos por otros. Por tanto, la distribución observada debía tener una cola truncada. . . Dichas reflexiones le llevaron a la construcción de una sociología que posteriormente, en el siglo XX, sería empleada por los fascistas italianos para fundamentar sus proceder.

Desde el momento en que sus ideas en torno a la distribución de ingresos se hicieron públicas, fueron atacadas desde diversos frentes, y a sus investigaciones se les quiso restar importancia. Se adujo que lo que había encontrado no podía considerarse una ley puesto que estaba basado únicamente en datos empíricos; que sus ajustes lineales carecían de pruebas de ajuste y que los errores se ocultaban por el tipo de graficación empleada. Pero el tiempo revivió sus contribuciones, pues empezando por Zipf, en los treinta del tormentoso siglo XX, ha habido toda una serie de reinterpretaciones acerca de su trabajo. Han surgido aproximaciones y modelos que tratan de explicar y extender la ley de Pareto a partir de diversas técnicas. Como ejemplos, podemos

citar los siguientes: en 1953, aplicando procesos estocásticos a la evolución del ingreso individual, Champernowne, por un lado, y Wold y Whittle, por otro, y de forma independiente, diseñaron unas cadenas de Markov que convergen a distribuciones paretianas; en 1960, Mandelbrot exploró de qué forma la familia de distribuciones de Pareto-Levy generaliza las propiedades de estabilidad de la curva normal; en 1972, Samuelson y Chipman regresaron a la cuestión de cómo los cambios en el exponente afectan el bienestar social.<sup>2</sup>

## 2.5 Los aportes de Zipf

La búsqueda de regularidades estadísticas análogas a las de la física, llevó al lingüista de Harvard, George Kingsley Zipf (1902-1950), a encontrar varias distribuciones en las ciencias sociales que siguen leyes de potencia. Las dos más representativas son las que a continuación se presentan: la estadística de ciudades y la estadística de palabras; describiremos brevemente algunas de las modelaciones que se han planteado para explicarlas y veremos también algunas de sus limitaciones más importantes.

### 2.5.1 Estadística de ciudades

En este caso lo que buscamos es la distribución de los tamaños de las ciudades, expresado en términos del número de habitantes. Supongamos que tenemos una lista de los asentamientos humanos en el mundo, clasificados por continente, regiones económicas, países, estados, provincias o alguna otra etiqueta similar, y que además disponemos de los datos sobre el número de habitantes de cada una. Supongamos además que elegimos una de tales listas y la ordenamos de forma decreciente, para luego etiquetarla, asignándole a cada ciudad un número de acuerdo al lugar que ocupa en nuestra lista ordenada de forma decreciente. A la ciudad más poblada le asignamos el rango 1, a la segunda mayor aglomeración le ponemos rango 2, y así sucesivamente. Una vez hecho lo anterior y para percibir ‘de golpe’ el comportamiento de nuestro ejercicio de etiquetación, graficamos en papel log-log el tamaño de las poblaciones contra su rango, y lo que encontramos es una línea recta que cubre varios órdenes de magnitud. Tal ejercicio fue realizado por primera vez por Zipf, quien haciendo regresión lineal encontró que la pendiente de tal recta es aproximadamente  $-1$ , es decir, encontró que la distribución de las ciudades sigue una relación de proporcionalidad inversa entre el tamaño de las mismas y sus rangos. A tal distribución se le acostumbra llamar *distribución rango-tamaño*, para diferenciarla de la *distribución de Zipf*, a la que se le asocia un exponente cercano a 1. Si denotamos con  $y$  al tamaño de ocurrencia de un

---

<sup>2</sup>Para consultar tales materiales, se sugiere revisar los trabajos de William J. Reed, especialmente [60].

evento relativo a su rango,  $r$ , y si la pendiente de la regresión lineal es  $b$ , tendremos que

$$y \sim r^{-b}$$

que es la expresión clásica de Zipf cuando el exponente  $b$  es muy cercano a 1, y como ya dijimos, para una pendiente negativa general se usa el término distribución rango-tamaño.

El hecho de que entre más pequeña es una aglomeración es más frecuente encontrarla no es una observación muy aguda, pues ello se debe a que sería muy difícil que hubieran muchas ciudades como la de México, pues demandarían una cantidad tal de recursos que pondrían en riesgo la viabilidad de un país. Por eso, el número de megalópolis es muchísimo menor que el número de pueblos y pueblitos, y entre ambos números no existe una proporción lineal. Pero lo que sí es un hecho notable es el que la distribución de Zipf (general) exhiba un alto grado de estabilidad a través del tiempo y de diversos países, pues aparece usando datos de regiones geográficas con economías, historias y sociedades diversas. Las excepciones vienen marcadas por un *forzamiento* económico y político, como en su momento lo fue la Unión Soviética o, en su caso, la República Popular China.

Antes de seguir adelante, debemos hacer varias observaciones. Al momento de llevar a cabo nuestro listado ordenado por rangos, supusimos que las listas de las poblaciones ya estaban dadas, pero no dijimos nada acerca de cómo se construyeron. El problema radica en lo que se entiende por ‘población’, ‘aglomeración urbana’, ‘asentamientos humanos’, ‘conglomerado urbano’ o ‘ciudad’, pues para asignarles rangos y tamaños a cada uno hay que conocer cuál es el número de habitantes, y ello nos lleva a un problema de resolución, pues no podemos conocer el número exacto de habitantes, ya que éste cambia debido a los nacimientos, defunciones, migraciones, etc., pero además, hay agregados ciudadanos donde no es muy claro que se tenga una única ciudad, ya que hay fronteras borrosas (por ejemplo, el caso de una población rural transformándose en urbana); conglomerados como la Ciudad de México, Guadalajara o Monterrey, y sus respectivas zonas metropolitanas, muestran este problema, pero el mismo aparece cuando hay fronteras nacionales, estatales o municipales. Por todo ello, la manera de construir el listado de ‘ciudades’ es un tanto arbitraria, pues depende de la manera en que se cuente el número de habitantes y se consideren sus fronteras.

Se puede decir que todos los intentos por dar una explicación a esta regularidad en la estadística de ciudades se pueden clasificar en dos grandes clases: (i) modelos jerárquicos basados en hipótesis microeconómicas y (ii) modelos estocásticos que buscan explicar la distribución observada como una consecuencia de hipótesis probabilísticas simples referentes a la formación y crecimiento de las ciudades. Sin embargo, ninguno ha sido completamente exitoso; tan es así, que algunos investigadores han

llegado a expresar en tiempos recientes que la distribución rango-tamaño es una regularidad empírica en búsqueda de una teoría.<sup>3</sup>

En el caso que nos ocupa, se ha observado empíricamente que la distribución de los tamaños de los asentamientos humanos (*settlements*) (término más general que el de ‘ciudades’) tiene lugar sobre un amplio rango observado. Así, casi todos los modelos construidos son insatisfactorios por lo siguiente: unos se centran en la explicación del exponente igualado a uno (es decir, se restringen al caso clásico de Zipf, a pesar de la evidencia empírica en contra); otros sólo son capaces de reproducir la distribución rango-tamaño para un cierto intervalo, la región que tiene un buen ajuste, o sea la cola; otros más, sólo pueden explicar la distribución para *ciudades* por arriba de un cierto tamaño umbral, centrándose incluso en investigar la magnitud de ese umbral, lo que deja de lado nuevamente la explicación de la distribución completa. Pero ya veremos más adelante algunos modelos que respondan ante tales deficiencias.

### 2.5.2 De Pareto a Zipf y viceversa

La ley de Zipf usualmente se refiere al tamaño  $y$  de ocurrencia de un evento relativo a su rango, como el que vimos para el caso de la estadística de ciudades, pero también podríamos graficar en papel log-log el número de asentamientos humanos que tienen cierto número de habitantes (la frecuencia de aparición de una cantidad dada) contra el tamaño de la población. En ambos casos obtenemos un comportamiento que satisface una ley de potencia y los exponentes no son los mismos. Pero dado que se refieren a los mismos datos, ambas descripciones deben ser equivalentes. En este segundo arreglo (donde no empleamos los rangos), podemos hacer los ajustes necesarios para tener una densidad de probabilidad de Pareto que nos diga el número de ciudades con una población de exactamente  $n$  habitantes, que es:

$$P[X = n] \sim n^{-a} = n^{-(d+1)} \quad (2.2)$$

Para pasar a la primera forma (la de los rangos), podemos hacer la siguiente observación en juego verbal: decir que “la  $r$ -ésima ciudad tiene  $n$  habitantes” es equivalente a decir que “ $r$  ciudades tienen  $n$  o más habitantes”, y en este caso tenemos una formulación del tipo describible con la distribución de Pareto. Considerando nuestras gráficas, ese parafraseo casi equivale a intercambiar los ejes coordenados. En el caso de que  $b$  sea el exponente de la recta en términos de rangos, esto es  $n \sim r^{-b}$ , si invertimos los ejes, ahora tendremos que  $r \sim n^{-1/b}$  representará la expresión en términos de densidad descrita más arriba en (2.2). De manera que si deseamos expresar esta inversión de ejes en términos de la ley de Pareto, es decir, en términos de la distribución acumulada, sólo tenemos que comparar los exponentes. Por eso

---

<sup>3</sup>Por ejemplo, ver [60].



tenemos que el exponente de la distribución de Pareto es igual al de Zipf más uno, pues al derivar a Pareto obtenemos la expresión anterior.

También debemos señalar que a partir de la expresión de la estadística de ciudades en términos de Pareto, se puede llegar a interpretar la existencia del umbral como parte de la descripción, y entonces, por coherencia, hay que buscarle una explicación, pero eso no corresponde con lo empíricamente observado, pues como ya dijimos, la diferenciación entre los diferentes asentamientos se hace de manera arbitraria.

### 2.5.3 Estadística de palabras

Para fijar ideas, así como introducir terminología y luego poder hacer comparaciones, vamos a presentar el siguiente ejemplo-experimento, al que nos referiremos en lo sucesivo como EE:

Por razones que no vienen al caso, elegimos como objeto de estudio un ejemplar de una publicación universitaria: la *Revista Chiapas* No. 13, correspondiente al año 2002, que de ahora en adelante será nuestro *corpus* de trabajo. Dado que en todo texto hay palabras que se repiten, lo que buscamos es caracterizar cómo se distribuyen las palabras que se repiten, cuáles son las que se usan más veces y cuántas, y cuáles aparecen sólo una vez, o dos, o tres. Lo primero que se hace es contar el número de veces que aparece cada palabra diferente y también contar la cantidad de palabras que hay en el *corpus* particular elegido (el tamaño de nuestra muestra). Luego se ordenan los datos obtenidos en forma decreciente, empezando por la palabra que aparece más veces, a la que se le asigna el rango 1. El segundo rango se le asigna a la palabra que tiene el segundo valor más frecuente, y así sucesivamente; si encontramos con que dos o más palabras tienen la misma frecuencia, usamos el orden alfabético para asignar rangos. Una vez que se han obtenido todos esos datos lo que procede es graficarlos, pero para poder compararlos en términos estrictos de frecuencias debemos normalizarlos; hablamos entonces de frecuencias relativas, aunque por simplicidad, en el caso de una palabra cualquiera, uno se refiere a la frecuencia de aparición de tal palabra y no a su frecuencia relativa. Una vez normalizados, si hacemos una gráfica log-log observamos la firma característica de las leyes de potencia: una línea.

La figura 2.2 corresponde a nuestro EE, con los valores de la pendiente de la recta y del coeficiente de correlación.

Con el procedimiento del EE hemos relacionado el *rango* de una palabra con la *frecuencia* de la misma. Eso fue lo que hizo Zipf de manera exhaustiva, encontrando que esos resultados tienen un comportamiento uniforme bajo una notable variedad de circunstancias: para diferentes tópicos, autores y lenguajes naturales, y dedicó gran parte de su vida intelectual a explorar y tratar de explicar lo que tal regularidad podría significar.

Con todo y lo que hemos descrito para el EE, hay varias maneras de enunciar la ley

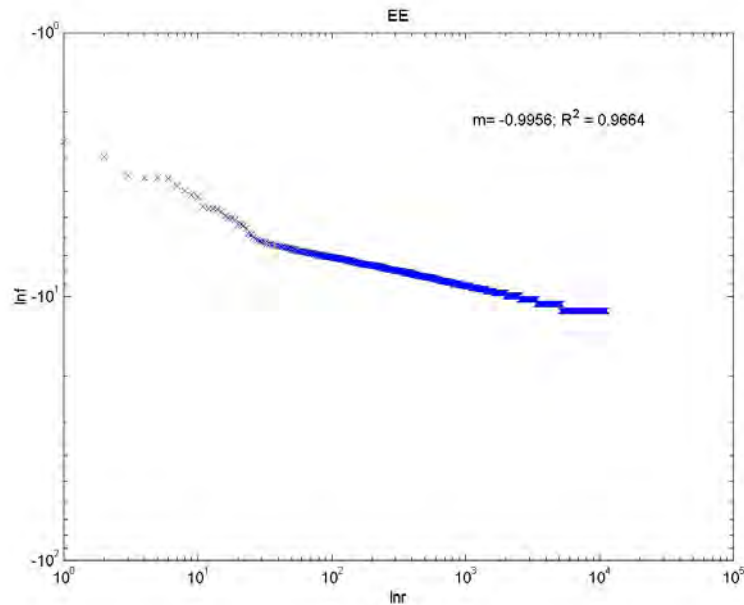


Figura 2.2: Gráfica log - log para nuestro EE.

de Zipf en lingüística. A veces se cita diciendo que establece con muy buena aproximación que la frecuencia relativa,  $f$ , de una palabra en un texto dado es inversamente proporcional a su rango, y se dan estos números:

$$f(r) \approx \frac{1}{r \ln(1.78R)}$$

Donde  $R$  es el número de palabras diferentes, o sea, el tamaño del *corpus*.

Otras veces sólo se presenta de la forma más simple (sin  $R$ ) y se dice que la ley de Zipf asegura que la frecuencia de aparición de una palabra es inversamente proporcional al rango de dicha palabra, es decir, que  $f(r) \sim 1/r$ ; a leyes de tal tipo se les llama ‘leyes hiperbólicas’ por obvias razones.

Hay varios resultados interesantes a partir del trabajo de Zipf. Por ejemplo, él encontró que en inglés escrito, para un texto cuya  $R = 12,000$ , las frecuencias relativas para las palabras más usadas, *the*, *of*, *and*, *to* son aproximadamente, 0.1, 0.05, 0.033, 0.025, respectivamente. La figura 2.3 muestra la correspondencia entre la ley de Zipf (una ley de potencia homogénea) y los datos descritos anteriormente.

La ley de Zipf no sólo es aplicable a idiomas escritos,<sup>4</sup> sino también a escritores

<sup>4</sup>Hay que tener muy presente que nos referimos a lenguajes escritos, no hablados, donde las palabras

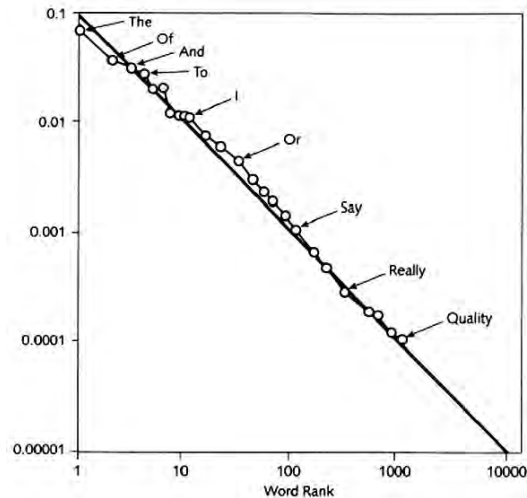


Figura 2.3: Ley de Zipf para el idioma inglés.

individualmente, y entre algunas consecuencias curiosas para este caso tenemos que, un buen escritor, con un vocabulario activo de  $R = 100,000$  palabras, las 10 más usadas ocupan alrededor del 24% del texto, mientras que para un léxico básico (digamos el de un periódico de tiraje nacional), con el 10% de su vocabulario ( $R = 10,000$ ), este porcentaje apenas se incrementa a casi el 30%. Por supuesto, cualquier escritor encontrará difícil evitar usar palabras como los artículos ‘la’, ‘el’, y los conectivos ‘y’, ‘de’ y ‘a’ en español (en inglés: *the*, *and*, *of*, y *to*, respectivamente). En nuestro EE ( $R \sim 83,000$ ) las 5 palabras mejor *rankeadas* alcanzan más del 20% del texto; las 10 primeras más del 30%, las 15 primeras más del 35% y las primeras 20 casi el 40%, o sea que ¡el resto ( $83,000 - 20$ ) ocupa el 60%! lo que nos indica que sólo unos pocos elementos dominan.

En lingüística se han propuesto dos alternativas generales de explicación para el comportamiento ley de potencia:

1. Se asume que la regularidad refleja alguna propiedad universal de la mente humana.
2. Se asume que representa alguna consecuencia necesaria de las leyes de la probabilidad.

Zipf optó por la primera y buscó un *principio de mínimo esfuerzo* que explicara el equilibrio aparente entre uniformidad y diversidad en nuestro uso de las palabras. Su mejor rankeadas son otras.

tratado de 1949, *Comportamiento humano y el principio del mínimo esfuerzo*, constituye un ambicioso trabajo, donde describe leyes hiperbólicas en psicología, semántica, sociología y geografía. Ahí presenta un cúmulo de datos estadísticos que ocurren frecuentemente en ciencias sociales; la explicación de esos datos ha representado un reto atractivo para muchos de sus sucesores, quienes han elegido el segundo camino y han logrado explicaciones convincentes.

Por otro lado, uno de los primeros tratamientos probabilísticos fue hecho por Benoit Mandelbrot. Consideró a un mono tecleando al azar una máquina de escribir y encontró un resultado sorprendente: si el mono tecléa hasta producir una larga y aleatoria cadena de caracteres, y si definimos una ‘palabra’ como una sucesión de letras entre espacios en blanco, entonces el ‘lenguaje’ del mono también sigue la ley de Zipf. Una manera de verlo es asumir un texto como una colección aleatoria de símbolos con espacios en blanco distribuidos al azar a través del texto, entonces existen necesariamente más ocurrencias de ‘palabras’ cortas que largas. Y como la variedad de palabras disponibles se incrementa exponencialmente con la longitud de la palabra, entonces el fenómeno de Zipf se hace inexorable, pues pocas palabras serán usadas un enorme número de veces, mientras que una vasta cantidad de palabras largas se emplearán muy pocas veces o nunca.

Pero los resultados de Mandelbrot van más allá (Ver [71]):

- Un análisis detallado muestra que si la máquina de escribir del mono tiene  $N$  teclas equiprobables y una barra espaciadora (con probabilidad  $p_0$ ), entonces sus palabras tendrán frecuencias relativas en función del rango dadas por:

$$f(r) \sim r^{-1+\log(1-p_0)/\log N}$$

Para fijar ideas, consideremos un alfabeto occidental que no incluya a la ñ, es decir,  $N = 26$ , y tomemos  $p_0 = 1/5$ , entonces, el exponente que resulta es  $b = -1.068$ , sólo ligeramente menor que  $-1$ . Si nuestro alfabeto ahora es de  $N = 9$  y  $p_0 = 1/10$ , entonces se puede modelar con todas las fracciones decimales entre 0 y 1, en las cuales el ‘0’ nunca aparece. En general, las “palabras” del mono podrán ser modeladas como un conjunto de Cantor con dimensión fractal  $D$  que iguala al exponente recíproco de  $1/r$ . Para el polvo de Cantor  $D \approx 0.954$ , nuevamente muy cercano a 1.

- Para el vocabulario generado por el mono, es necesaria una cantidad enorme de palabras (1,895,761) para alcanzar la probabilidad de un medio. En contraste, el rango medio en textos en inglés está entre 100 para textos medios y 500 para grandes escritores. De esta manera, el mono, apegándose estrictamente a la ley de Zipf, produce un lenguaje con muchísimas palabras, causando la imposibilidad de construir un diccionario, pues sus palabras forman un incontable conjunto de Cantor.

- Asociado a su dimensión fractal, el lenguaje del mono es autosemejante, ya que si tomamos una palabra y eliminamos la letra de la extrema izquierda, obtendremos otra palabra de ese mismo lenguaje. Si seguimos ese proceso de ‘generación de palabras’ en árboles autosimilares, obtendremos al final una copia de todo el vocabulario. El punto importante es que los lenguajes naturales NO crecen como el descrito, ya que, en efecto, muchas combinaciones de letras son no-palabras.

De esta forma, la aproximación de Mandelbrot deja sin explicar la ley de Zipf, porque no logra distinguir entre un comportamiento aleatorio (el del mono) y uno racional (humano). Ya que si nos centramos tan sólo en la longitud de las palabras, no podemos generar una explicación únicamente a través de ello. De esta forma es como Zipf abordó el problema de la relación entre significado y la extensión de las palabras (longitud), y argumentó que existen sucesiones de letras que son impronunciables (no-palabras), que aunque sean más cortas, tienen la dificultad de su pronunciación; por el contrario, existen palabras largas pero que sí se pueden pronunciar y ello las hace más útiles. Por ejemplo, la no-palabra ‘zvk’ es impronunciable pero ‘indudablemente’ no.

Zipf debió considerar como ‘palabras’ a todas aquellas sucesiones de letras que empleamos en nuestros lenguajes, no sólo las que según el diccionario son equivalentes o están relacionadas, por ejemplo, las conjugaciones de los verbos (en un diccionario sólo aparece el verbo en infinitivo); consideró así las palabras pertenecientes a un mismo conjunto de raíces.<sup>5</sup> Así construimos también nosotros nuestro EE.

A partir de las observaciones de Zipf, varios usos se han planteado, pues para estudiar diversas propiedades de las palabras es obvio emplear sus propiedades estadísticas. Ello mismo impone ciertas restricciones, pues un número pequeño de palabras ocupa el mayor volumen del *corpus*, mientras que la inmensa mayoría de palabras tiene un número de ocurrencias estadísticamente insuficiente. Una manera de resolver el problema de la poca aparición de ciertas palabras es aumentando el tamaño del *corpus*, o de plano recurrir al *corpus* más grande que existe: Internet. Y de ahí, una de las aplicaciones principales es proporcionar conocimiento léxico a diferentes niveles y tipos: co-ocurrencias, colocaciones, etc.

## 2.6 Modelos estocásticos

Ya que ha quedado establecida la equivalencia entre la descripción de Pareto y la de Zipf, podemos proceder a presentar uno de los modelos que se han construido

---

<sup>5</sup>En este caso se tiene ‘una palabra, un rango’, definiendo palabra como una sucesión de letras con significado, y casi siempre, legible.

para explicar tales regularidades, y que responde ante varios de los retos no resueltos. Describiremos un modelo construido para los asentamientos humanos, pero que se podría adaptar al caso de ingresos o a la lingüística según sea el caso, y que reproduce muy bien los datos observados tanto en la cola superior (la de los primeros rangos o de grandes poblaciones), donde se cumple la propiedad rango-tamaño, como en la cola inferior (de rangos muy altos o de poblaciones pequeñas) donde empíricamente aparece una propiedad inversa a la de rango-tamaño. El reto, en general, es encontrar un modelo que no tenga las fallas de los anteriores y que reproduzca lo mejor posible la distribución observada en un amplio espectro. El modelo que presentamos lo hace muy bien.

Este modelo está basado en una hipótesis probabilística muy simple sobre la formación y crecimiento de los poblados (asentamientos o ‘ciudades’, que a partir de aquí en adelante consideramos como sinónimos y que empleamos indistintamente), misma que refleja su variabilidad inherente, tanto de uno a otro poblado como en el tiempo. Nos referimos al trabajo de W. Reed citado en [60], pero también desarrollado en [61], [62] y [63]. La explicación es entonces esencialmente matemática y es una consecuencia de procesos estocásticos con ciertas características, lo que no implica que los factores demográficos, económicos, geográficos, políticos o de otro tipo no sean importantes en la fundación y desarrollo de las ciudades; por el contrario, significa que cuando consideramos la distribución de los asentamientos humanos sobre una región, los efectos de las variaciones en esos factores pueden ser modelados efectivamente por un proceso estocástico. Siendo matemática, la explicación no está confinada a la distribución de los tamaños de los asentamientos humanos, como ya decíamos, sino que es relevante para otros fenómenos con una estructura subyacente similar.

Se adopta un punto de vista ‘macro’, implicando que las diferencias entre asentamientos serán consideradas como esencialmente aleatorias, y las componentes del modelo estocástico son dos, una para la fundación y otra para el crecimiento o evolución subsecuente. A partir de estas componentes derivaremos una distribución para el tamaño actual de los asentamientos y discutiremos sus propiedades.

Consideremos primero la evolución de los poblados después de su fundación. Los asentamientos de los individuos humanos crecen (a veces se contraen) de diferentes y variadas formas. La tasa de crecimiento (proporcional) en un año dado varía de poblado a poblado y para un poblado dado probablemente variará año con año (o década con década, etc.) dependiendo de factores económicos, demográficos, políticos, etc. A un nivel macro tal variabilidad puede ser modelada matemáticamente suponiendo que el logaritmo del tamaño de la población para cualquier asentamiento constituye una realización de una caminata aleatoria. Esta es la ley de Gibrat de los efectos proporcionales (1931). Por conveniencia analítica se adopta una versión continua en el tiempo, es decir, el tamaño (población o número de habitantes)  $X(t)$  de un asentamiento se supone que sigue un movimiento browniano geométrico (GBM -

*geometric brownian motion*) gobernado por la ecuación diferencial estocástica de Itô:

$$dX = \mu X dt + \sigma X dw$$

donde  $dw$  es ruido blanco (i.e. el incremento aleatorio de un proceso de Wiener en el tiempo  $dt$ ). El parámetro  $\mu$  es la razón media de crecimiento calculada sobre todos los asentamientos y tiempos, y  $\sigma$  es el parámetro que refleja la variabilidad en esta razón de crecimiento. Entonces el crecimiento proporcional  $dX/X$  en el tiempo  $dt$  para cualquier asentamiento comprenderá una componente sistemática  $\mu dt$  (que refleja la razón de crecimiento promedio sobre todos los asentamientos, a todos los tiempos) y una componente aleatoria  $\sigma dw$  (que refleja lo que ocurre para un asentamiento particular a un tiempo particular).

Este es el modelo de crecimiento de asentamientos usado por Gabaix [24], quien cita estudios empíricos para justificar la ley de Gibrat, pero de forma contraria a éste último, en el modelo de Reed no se buscan las condiciones de equilibrio, por lo que no es necesario asumir la existencia de un tamaño mínimo de poblado o umbral que actúe como una frontera reflejante.

Si el tamaño inicial del asentamiento (al tiempo de su fundación) es  $X_0$ , entonces bajo un GBM, el tamaño  $\tilde{X}_T$  del asentamiento,  $T$  unidades de tiempo después, será una variable aleatoria lognormal con:

$$\ln \tilde{X}_T \sim N(X_0 + (\mu - \sigma^2/2)T, \sigma^2 T)$$

Pero en realidad los poblados fueron fundados a diferentes tiempos y sin duda tenían diferentes poblaciones iniciales. Por tanto, para cualquier asentamiento dado en el país o región considerada, las variables  $X_0$  y  $T$  deberían ser manejadas como variables aleatorias. Una especificación simple para la distribución del tamaño inicial  $X_0$ , sería que tuviera una forma lognormal con:

$$\ln X_0 \sim N(\mu_0, \sigma_0^2)$$

Esta expresión tiene dos propiedades deseables: a) que el tamaño inicial siempre sea no-negativo, y b) que la varianza se incremente con la media. Es posible que la distribución de los tamaños iniciales haya cambiado con el tiempo (los asentamientos agrícolas probablemente fueron inicialmente más pequeños que los industriales o que los urbanos). Así, se puede ajustar esto suponiendo que  $X_0$  también evoluciona como un GBM. No hay diferencias esenciales en el desarrollo si se incluye lo anterior, pero por simplicidad eso se hace en el apéndice.

Consideremos ahora la fundación de los asentamientos. El tiempo desde la fundación variará de poblado a poblado, y su distribución sobre todos los asentamientos en la región reflejará el desarrollo histórico de la región, que, por supuesto, dependerá de múltiples factores. Como con la evolución de los asentamientos, ignoraremos

los detalles y modelaremos la fundación de asentamientos de forma estocástica. Lo más simple que se puede considerar es que las fundaciones ocurran en un proceso de Poisson sobre las últimas  $\tau$  unidades de tiempo (es decir, ocurren aleatoria e independientemente a una tasa promedio constante). Tal modelo, sin embargo, está limitado en que no permite un crecimiento sostenido y general para toda la región. Un modelo más realista resulta de asumir que en el intervalo de tiempo  $(t, t + dt)$ , cualquier asentamiento existente puede formar un nuevo poblado satélite con una probabilidad  $\lambda dt$ . Esta suposición constituye un proceso de Yule, quien lo propuso en 1924 para modelar la creación de nuevas especies biológicas, y que no está de más decir, es un proceso similar en muchos aspectos a la fundación de nuevos asentamientos humanos. Para tal proceso, el número de poblados esperados  $t$  unidades de tiempo después de la fundación del primer poblado es  $e^{\lambda t}$ . En otras palabras, el número de poblados está creciendo, en promedio, a la tasa proporcional  $\lambda$ . Debe notarse, sin embargo, que la evolución efectiva del número de poblados es un proceso aleatorio. Para este modelo se puede mostrar que la distribución al tiempo  $T$  desde la fundación de un asentamiento (que bien puede ser el primero) es de la forma de una distribución exponencial truncada en  $\tau$  (la edad del primer asentamiento). En muchos casos es razonable asumir que  $\tau$  es grande, y entonces es mejor considerar la distribución límite cuando  $\tau \rightarrow \infty$ , resultando una distribución exponencial con densidad  $\lambda e^{-\lambda t}$ , para  $t > 0$ .

Bajo la suposición del modelo de Yule para la fundación de poblados y el modelo de GBM para su crecimiento subsecuente, la distribución del tamaño en curso,  $\bar{X}$ , sobre todos los asentamientos, puede ser obtenida integrando la densidad de la distribución lognormal de  $\tilde{X}_T$  con respecto a la distribución exponencial de  $T$ . Esto puede hacerse analíticamente dando como resultado una densidad de probabilidad (de  $\bar{X}$ ) de la forma:

$$f_{\bar{X}}(x) = \frac{\alpha\beta}{\alpha + \beta} \left[ x^{-\alpha-1} \exp \{ \alpha\mu_0 + \alpha^2\sigma_0^2/2 \} \Phi \left( \frac{\ln x - \mu_0 - \alpha\sigma_0^2}{\sigma_0} \right) + x^{\beta-1} \exp \{ -\beta\mu_0 + \beta^2\sigma_0^2/2 \} \Phi^c \left( \frac{\ln x - \mu_0 + \beta\sigma_0^2}{\sigma_0} \right) \right] \quad (2.3)$$

con  $x > 0$ , donde  $\Phi$  es la función de distribución acumulada de la distribución normal estándar;  $\Phi^c = 1 - \Phi$ ; y  $\alpha, -\beta$  ( $\alpha, \beta > 0$ ) son las raíces de una ecuación característica cuadrática. Reed bautizó a esta distribución como distribución doble de Pareto-lognormal (dPIN - *double Pareto-lognormal*), y tiene la propiedad de que sigue leyes de potencia en las dos colas:

$$f_{\bar{X}}(x) \sim x^{-\alpha-1} \quad (x \rightarrow \infty); \quad f_{\bar{X}}(x) \sim x^{\beta-1} \quad (x \rightarrow 0)$$

El primer resultado indica que para valores grandes de  $x$ , la distribución rango-tamaño sigue una ley de Pareto (es decir,  $\ln(\text{Pr}(X \geq x))$  está relacionada linealmente



con  $\ln(x)$  y tiene pendiente  $-\alpha < 0$ ); y el segundo resultado indica una ley de Pareto inversa para  $x$  pequeños (es decir,  $\ln(\Pr(X \leq x))$  está relacionada linealmente con  $\ln(x)$  y tiene pendiente  $\beta > 0$ ). La cola superior paretiana ha sido ampliamente verificada empíricamente (es la famosa ley rango-tamaño), pero aparentemente nadie ha buscado una ley en la cola inferior (inversa), que existe, como lo muestra la figura 2.4. Incluso a partir de los datos presentados por Reed se ve que el grado de linealidad en la gráfica de la cola inferior es aún más notable que en la gráfica de la cola superior, confirmando empíricamente la presencia de la propiedad rango-tamaño en la cola inferior.

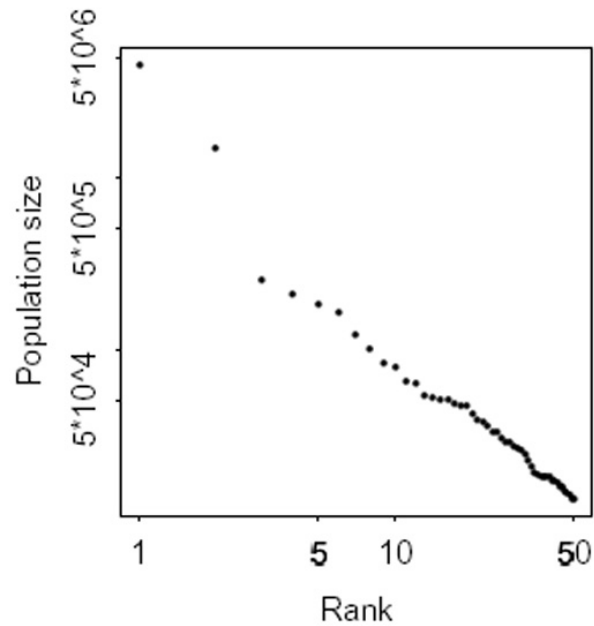


Figura 2.4: La cola de los rangos altos para los poblados de la provincia de Barcelona.

Posteriormente, Reed lleva a cabo varias pruebas de bondad de ajuste, logrando un ajuste muy bueno; y como decíamos, el modelo se puede aplicar a fenómenos que tengan, en principio, comportamientos similares, como es el caso de la estadística de palabras, pero habría que tener cuidado de cómo hacer la correspondencia. Quizás una falla de este modelo podría encontrarse en el intento de explicar todo el dominio de definición de los rangos y no sólo las regiones asintóticas.

# Distribuciones rango-orden

---

[...] When we say that the apparatus of probability theory is best suited to the study of molecular phenomena, we do not in the least wish to say that the philosophical premises for applying the theory of probability in natural science spring from the insufficiency of our knowledge. The basic principle lies in the fact that when studying mass-scale phenomena, a set of *new and peculiar regularities* come to light. When studying phenomena caused by the action of large numbers of molecules, it is not necessary to take into consideration all the properties of every molecule.

B. V. Gnedenko, *The Theory of Probability*.

En el capítulo anterior se desarrolló una distribución asociada a los nombres de Zipf y Pareto. Ahora veremos otra área donde surgen leyes de potencia asociadas al nombre de Lotka, pero que matemáticamente equivalen a la formulación de Zipf. Veremos una descripción que la generaliza mediante el empleo de dos exponentes, en la representación denotada por  $\{a, b\}$ , y que será la que se modelará en el siguiente capítulo.

## 3.1 Bibliometría, informetría, ...

Sin hacer siquiera un breve recuento histórico sobre el origen de los nombres que encabezan esta sección,<sup>1</sup> podemos decir que al paso del tiempo se han desarrollado varias disciplinas estrechamente relacionadas con las denominadas ciencias de la información, y que comparten el empleo del sufijo ‘metría’ (del griego *metron*), que significa tanto ‘medir’ como ‘métrica’: la bibliometría, la informetría, la cienciometría y la webometría.

---

<sup>1</sup>Se pueden consultar algunos avances en ese sentido, en los trabajos [4], [32], [53], [77], [57] y los materiales citados en [69].

Originalmente, la aplicación de diversos métodos matemáticos al estudio de la literatura científica, se hizo en el trabajo de los bibliotecarios, quienes deseaban conocer cómo se estructuraban las referencias bibliográficas y la sistematización de algunos aspectos de la productividad científica; por ello el nombre que se acuñó refleja parcialmente su origen: bibliometría. Como se señala en [32], lo que convierte a tales metodologías en bibliometría “no es tanto el recuento que en sí mismo, que no pasaría de ser estadística bibliotecaria, cuanto la idea de que existe una relación sustancial entre cantidad de libros y cantidad de conocimiento”, es decir, “la bibliometría ha estado históricamente vinculada a la idea de que es posible representar el conocimiento humano a través de la cuantificación de los documentos en los que éste se expresa y de los elementos que componen a éstos”. Este enfoque se extendió a otras áreas para conformar los campos de informetría, ciencimetría y webometría, modelando las aplicaciones concretas de los campos a los que pretender servir de instrumento. Así, se entiende a la ciencimetría como aquellos estudios dedicados a los aspectos cuantitativos de la ciencia como disciplina o actividad económica; se considera una parte de la sociología de la ciencia y estaría enfocada al establecimiento de políticas científicas (hacer pronósticos y tomar decisiones de planeación), incluyendo las políticas editoriales y de difusión (traslapándose así con la bibliometría). De manera similar, la informetría abarca el estudio de los aspectos cuantitativos de la información, independientemente de la forma en que aparezca registrada y del modo en que se genere, almacene o recupere; y otra vez se traslapa con la ciencimetría y la bibliometría. Tales traslapes han generado ambigüedad, tanto en la interpretación de los términos como en su uso, pero, resaltamos, hay un trasfondo común: la búsqueda de relaciones matemáticas que describan estructuras y procesos que superan la mera recopilación de datos.

Por ejemplo, de un tiempo a la fecha, varios indicadores inicialmente bibliométricos se han usado para medir los resultados de las ciencias en un país, nación u organización. Pero como son un tanto reducidos, diversos actores han propuesto que se les correlacione con otros indicadores, desde los demográficos y económicos, hasta los de tipo social, para que ello redunde en un enfoque más amplio de los análisis realizados. Por ejemplo, la UNESCO ha propuesto los siguientes indicadores: el crecimiento del producto interno bruto (PIB), el gasto en educación, la cantidad de profesionales y publicaciones científicas por número de habitantes, y el número de profesionales en I+D (investigación y desarrollo), lo que ha impactado tanto la evaluación del quehacer científico como la definición de sus políticas asociadas.

Actualmente, y de acuerdo a [4], pueden considerarse como objetos de estudio y como delimitadores de la informetría, las siguientes áreas :

- Los aspectos estadísticos del lenguaje y la frecuencia del uso de las palabras y frases.

- Las características de la productividad de los autores, medida por la cantidad de documentos publicados en un tiempo determinado o por su grado de colaboración.
- Las características de las fuentes donde se publican los documentos, incluida su distribución por disciplinas.
- Los análisis de citas, según distribución por autores, tipo de documento, instituciones o países.
- El uso de la información registrada a partir de su demanda y circulación.
- La obsolescencia de la literatura mediante la medición de su uso y de la frecuencia con que se cita.
- El incremento de la literatura por temas.
- La distribución idiomática según la disciplina o el área estudiada.

Como se ve, los estudios en lingüística que dieron lugar a la ley de Zipf, caen dentro de esta definición. Vayamos a estudios que se consideran específicos de la informetría, pero teniendo en cuenta que las delimitaciones no son todavía claras, sino más bien flexibles.

Se considera que los primeros estudios sobre la productividad científica corresponden al trabajo de Arnold Dresden, *A Report on the Scientific Work of the Chicago Section, 1897-1922*, (ver [21]) y al de Alfred J. Lotka, *The Frequency Distribution of Scientific Productivity*, (ver [40]). En el trabajo de éste último aparece lo que se conoce como la ley de Lotka, y que, como ya decíamos, tiene la misma forma matemática que la ley de Zipf,<sup>2</sup> pero expresada en términos de las variables informétricas, a saber: a  $r$  se le considera como colaboradores o autores de un trabajo (*paper*) dado y a  $f(r)$  como artículos (*papers*) en sí. Y desde entonces es común llamar “fuente” a la variable independiente e “item” a la variable dependiente. De tal suerte, que la ley de Lotka establece que el número de items es una ley de potencia de las fuentes. Originalmente, Lotka enunció el resultado como sigue: El número de autores,  $A_n$ , que publican  $n$  trabajos sobre una materia es inversamente proporcional a  $n^2$ ; por ello, a veces también se da el caso de referirse a la ley de Lotka únicamente para el caso del exponente dos, aunque esto se da regularmente en la literatura asociada a la biblioteconomía. La tendencia actual es romper con lo anterior y plantear desarrollos generales, así, se ha desarrollado una rama dedicada al estudio de las leyes de potencia, denominada: *informetría lotkaiana*. Por tanto, se puede decir que las relaciones entre las fuentes

---

<sup>2</sup>De hecho, algunos autores la refieren como la ley de Zipf-Lotka, por ejemplo [37].

y los items son objeto de estudio de la informetría; y que en la literatura es normal encontrar las parejas autor-journal o journal-bibliografía referidas como fuentes e items.

Otro personaje estrechamente asociado con el desarrollo de la informetría es Samuel C. Bradford, quien investigó la productividad de las revistas (ver [4]), haciendo varias contribuciones importantes; la más conocida es la ley que lleva su nombre, “en la que se plantea una forma longitudinal acumulativa de distribución de los documentos por disciplinas en las publicaciones seriadas”.

A continuación se hace una descripción de los resultados obtenidos recientemente en exploraciones acerca del parámetro de impacto de las revistas científicas (*Journal Impact Factors* - JIFs), o simplemente factores de impacto (FI), y que extienden la informetría lotkaiana, pues muestran que las gráficas rango-orden (análogas a las de rango-tamaño de ciudades) relativas a diversos datos de JIFs se desvían de la tradicional ecuación lotkaiana.

## 3.2 Factores de impacto

En esta sección nos estamos refiriendo a los resultados publicados en el artículo de Mansilla *et al.*, [43].

Un factor de impacto es una medida de la frecuencia con la que un “artículo promedio” en una revista ha sido citado en un año en particular o en cierto un periodo  $T$ ; es calculado “dividiendo el número de veces que una revista ha sido citada por el número de artículos que ha publicado durante algún periodo de tiempo dado. El factor de impacto de una revista reflejará entonces una tasa promedio de citación por artículo publicado”. El factor de impacto de una revista es un intento por evaluar la producción de conocimiento publicado entre diferentes revistas de un campo dado. Los resultados son centralizados en la base de datos del *Science Citation Index* (SCI), que se publica anualmente, desde 1975, en los *Journal Citation Reports*.

Un índice de citación, índice de citas o análisis de citas de una revista científica, es aquel que mantiene un seguimiento sobre qué artículos son citados en otros artículos. El más usado y mejor conocido es el *Web of Science* publicado por el Institute for Scientific Information (ISI). El SCI es una base de datos documental donde se recogen todas las contribuciones (artículos, editoriales, cartas, revisiones, discusiones, etc.) que se puedan publicar en las revistas de ciencia y tecnología indexadas por Thomson Scientific. A este índice de citación también se le conoce como ISI ya que en un principio la institución que producía el índice era el ya mencionado Institute for Scientific Information, fundado por Eugene Garfield, en 1960.<sup>3</sup>

---

<sup>3</sup>Se puede seguir el desarrollo histórico de las ideas de Garfield revisando sus primeros trabajos, entre los que destaca [26].

El FI ha sido objeto de muchas críticas y existe un debate sobre su uso como herramienta de evaluación de la investigación. Incluso la influyente revista *Nature*, en 2005, afirmó que las cifras del JIFs deben ser tomadas cuidadosamente.<sup>4</sup> A pesar de los pros y contras, el hecho es que es ya una medida cotidiana de la importancia de una revista y se usa a nivel internacional.

Aun manteniendo una actitud crítica hacia el uso de FIs para evaluar la investigación científica, se debe reconocer que es resultado del proceso de publicación y que se ha convertido por sí mismo en objeto de estudio científico. D. Lavalette se encuentra entre los autores cuya atención ha sido atraída hacia la distribución rango-orden del JIFs, pues Popescu lo señala en 2003 (ver [56]) como el autor de la propuesta de ley siguiente:

$$f(r) = K \left[ \frac{N + 1 - r}{r} \right]^b \quad (3.1)$$

Donde  $N$  es el número de revistas,  $r$  es el rango,  $f(r)$  es el factor de impacto,  $b$  es un parámetro por ajustar, y  $K$  una constante de normalización.

En la siguiente sección se expone una ley que generaliza la de Lavalette y que ajusta mejor para el caso de los factores de impacto.

### 3.3 Ley de potencias de dos colas

En diversas gráficas log-log, relativas a un buen número de datos empíricos, entre los que se incluyen los de los JIFs, se muestran evidencias que apuntan en contra de una ley de potencia con un único exponente, pues aparece una curvatura en la cola derecha de la gráfica, la correspondiente a los rangos altos (o frecuencias bajas), y muy parecida al decaimiento observado para el caso de las ciudades. En [43], Mansilla *et al.*, proponen una función tipo Beta con dos parámetros, como una extensión de lo que podría ser llamado *leyes lotkaianas de dos exponentes*, y que tiene una representación  $\{a, b\}$  de la forma:

$$f(r) = K \frac{(N + 1 - r)^b}{r^a} \quad (3.2)$$

Donde  $f(r)$  representa el factor de impacto rango-orden,  $r = 1 \dots, N$ ;  $a$  y  $b$  son dos parámetros a ajustar, y  $K$  es un factor de normalización sin significado. A la ecuación 3.2 se le refiere en la literatura como una función beta de dos parámetros.

<sup>4</sup>Un trabajo relativamente detallado acerca de los usos de los factores de impacto, lo constituye [81], abarcando aspectos de productividad por especialidad, por autores, etc. y cuestionando su uso en premiaciones; propone su uso junto al de otros indicadores, como el índice de inmediatez y el índice de vida media.

Campo científico	$K$	$b$	$a$	$R^2$
Física	0.0273	0.991	0.4058	0.9999
Matemáticas	0.0437	0.676	0.2622	0.9999
Ciencias de la Computación	0.0066	1.0626	0.2840	0.9999
Agrocencias	0.0070	0.9597	0.2210	0.9999
Ciencias Ambientales	0.0358	0.9357	0.2781	0.9800
Biociencias	0.0304	1.0161	0.5140	0.9999
Química	0.0549	0.9733	0.4560	0.9999
Ingeniería	0.0033	1.0472	0.3522	0.9999
Geociencias	0.0463	0.8739	0.3505	0.9999
Ciencias de Materiales	0.0408	0.9072	0.4477	0.9999
Medicina	0.0819	0.7735	0.4307	0.9999
Educación	0.0819	0.7735	0.4307	0.9999

Tabla 3.1: Valores encontrados para los exponentes y el coeficiente de correlación.

Nótese que cuando  $b = a$ , la ecuación 3.2 se convierte en la que podríamos llamar ley de Zipf-Lotka-Lavalette, y que representa, por tanto, una generalización de la misma.

### 3.3.1 Resultados en el caso de los factores de impacto

Una vez que han expresado su propuesta, los autores de [43] hacen lo siguiente: Para cada conjunto de datos, encuentran los valores de los parámetros usando el método de mínimos cuadrados en la variable logarítmica, aplicada a la ec. (3.2):

$$\log(f(r)) = \log(K) + b \log(N + 1 - r) - a \log(r) \quad (3.3)$$

La tabla 3.1 muestra los valores obtenidos de  $K$ ,  $b$  y  $a$ , además del coeficiente de regresión  $R^2$  para los factores de impacto de 12 disciplinas. Obsérvese que los valores de  $a$  y  $b$  casi no varían para disciplinas distintas, sino que más bien parecen concentrarse alrededor de algún valor promedio.

En el artículo mencionado se presenta también una serie de gráficas con los datos de los parámetros de impacto, así como los ajustes teóricos para los campos de las ciencias físicas, matemáticas y ambientales. Ejemplo de ello es la gráfica 3.1, en el caso del campo de la física. Conviene aclarar que la manera más natural de presentar los datos es mediante el empleo de gráficas semilogarítmicas (y no log-log), pues permiten ubicar el cambio en el orden de magnitud en una misma gráfica; y ello se hará también

más adelante en los otros casos tratados; la firma característica será una curva tipo sigmoide como la de la figura 3.1.

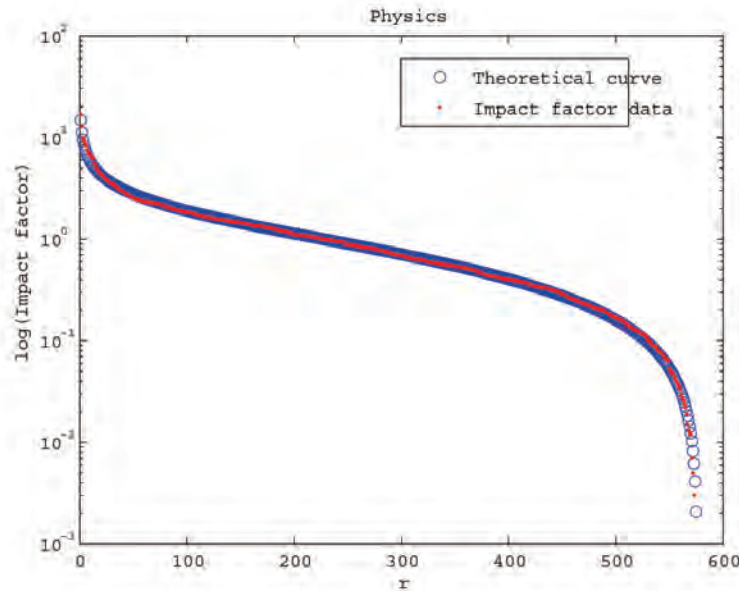


Figura 3.1: Gráfica del factor de impacto en el campo de la física

### 3.3.2 Conclusiones relativas a los factores de impacto

Los autores de [43] concluyen mostrando la excelente concordancia de los datos con su modelo, pues la calidad del ajuste es superior a la propuesta de Lavalette, a la que incluyen como caso particular haciendo  $a = b$ . A partir de ello reflexionan sobre los posibles ingredientes que deberían tener los modelos que intenten reproducir la representación  $\{a, b\}$ . Mencionan que “el mecanismo subyacente propuesto que da lugar al comportamiento anterior a menudo asume una forma de evolución biológica”, donde se incluyen los trabajos de G. Yule (1924) y los sistemas de expansión-modificación, propuestos por Wentian Li en 1991, [39], que toman en cuenta las características básicas del proceso de mutación del ADN, y cuyos resultados han sido ampliados en [44].

Cuando se discuten los factores de impacto de una revista, señalan, un balance entre la importancia para los investigadores de publicar un trabajo en una revista de alto rango, las dificultades asociadas con hacer esto y el incremento del impacto recibido por las revistas con un factor de impacto alto, parecen darle sustento al



“efecto San Mateo”,<sup>5</sup> que también ha sido observado en sitios tales como las redes complejas.

### 3.3.3 Sobre la ubicuidad

Considerando que la ec. (3.2) se le conoce en la literatura como función beta de dos parámetros (FBDP), nos referiremos a ella de esa forma. Así, además del trabajo [43], tenemos que la FBDP aparece tanto en las secuencias genéticas de mapeos unimodales [11], como en la estadística de las notas musicales. En [12] se aborda esto último, considerando las notas como secuencias de símbolos, y permitiendo un ordenamiento rango-tamaño; dicho trabajo abarca alrededor de 1800 composiciones musicales de música clásica, jazz y rock, y se muestra que los parámetros  $a$  y  $b$  obtenidos se agrupan según las tonalidades.

Por otro lado, considerando “estudios lingüísticos” no sólo aquellos relacionados con los lenguajes naturales sino también los relacionados con lenguajes arbitrarios sobre alfabetos finitos abstractos, se ha encontrado el mismo comportamiento. Se arguye que cuando el número de posibles ‘palabras’ es grande, como es el caso de los lenguajes naturales, se espera tener un buen ajuste con una ley de potencia de un parámetro (ley de Zipf-Lotka); sin embargo, cuando el número de palabras es más bien pequeño, como en el caso de los lenguajes de programación, una ley de potencia de un exponente falla completamente y más parámetros son necesarios para tener un ajuste conveniente. Siguiendo esta línea, en [48] se presentan algunos resultados, producto de exploraciones de tipo general sobre las propiedades algebraicas de alfabetos finitos.

Además, en [49] se avanza un poco más, pues se considera razonable asociarlas al producto de distribuciones de probabilidad correlacionadas; más aún, en el límite de muchas variables aleatorias, se tiene un ajuste cualitativo y cuantitativo con la FBDP, lo que es probado transformando el problema en uno algebraico, es decir, encontrando el rango de productos sucesivos de un conjunto de números aleatorios.

Así pues, las FBDP parecen ser ubicuas en física, biología, geografía, economía, música, lingüística, etc.

## 3.4 Ranking mundial de universidades

De forma análoga a los factores de impacto de las revistas científicas, de un tiempo a la fecha se ha buscado establecer cierto tipo de indicadores asociados a la labor que realizan las universidades a lo largo del mundo. A la fecha se tienen ya bases de datos,

---

<sup>5</sup>Robert K. Merton bautizó con ese nombre al hecho de que los investigadores científicos eminentes reciben mayores reconocimientos que otros investigadores, menos conocidos, por contribuciones equivalentes, haciendo referencia al Evangelio según San Mateo 25:29, que dice: “Porque al que tiene, le será dado, y tendrá más; y al que no tiene, aun lo que tiene le será quitado.”

metodologías, y como era de esperarse, partidarios y críticos de lo que se ha dado en llamar: ranking mundial de universidades.

Partiendo de lo realizado en el caso inmediato anterior, y omitiendo los acalorados debates que se han generado, se procedió a buscar en las bases de datos existentes, distribuciones tipo Beta de dos parámetros. Se consultaron los dos sitios siguientes, que son los más citados:

1. Ranking Mundial de Universidades en la Web  
([http://www.webometrics.info/index\\_es.html](http://www.webometrics.info/index_es.html))
2. Academic Ranking of World Universities (ARWU)  
(<http://www.arwu.org/>)

El primero es una iniciativa del Laboratorio de Cibermetría, dependiente del Consejo Superior de Investigaciones Científicas (CSIC) de España, dedicado al análisis cuantitativo de Internet y de los contenidos de la Red. Sus bases de datos fueron descartadas por varias razones: en primer lugar, la información publicada se refiere a la presencia web de las universidades, lo que en la literatura se considera como un criterio objetivo no-bibliométrico, es decir, la base de datos se construye a partir de las informaciones que proporcionan las universidades a discreción, lo cual es potencialmente manipulable y no completamente verificable; en segundo lugar, los datos se presentan de una forma no amigable para su manipulación, incluso de forma trunca, lo que representa una labor un tanto engorrosa.

El segundo sitio, que depende del Instituto de Educación Superior de la Universidad de Shanghai Jiao Tong, de China, tiene una base de datos muy completa, y fue la que se empleó. Desde que empezó a publicar información, en 2003, cataloga (*rank*) el lugar de una universidad mediante una evaluación global que es calculada a partir de cinco indicadores del desempeño académico y de investigación, considerando tanto a su alumnado como a su personal. Para cada indicador, la máxima calificación alcanzada es igualada a un puntaje de 100, y calificaciones menores son calculadas como un porcentaje del máximo. En el año 2003, para asignar el rango respectivo, le otorgó un 20 por ciento a cada uno de los indicadores siguientes: Nobel, HiCi, NS, SCI y PF. Como explican, cada uno señala lo siguiente: Nobel, indica el número de premios Nobel en física, química, medicina y economía, entre 1911 y 2002; HiCi, indica el número de los investigadores más citados (*Highly Cited*) en 21 categorías de las ciencias de la vida, medicina, ciencias físicas, ingeniería y ciencias sociales, en el periodo 1981-1999; NS, indica el número promedio de artículos publicados en las revistas *Nature* y *Science*, en el periodo 2000-2002; SCI, indica el número de artículos citados en el *Science Citation Index-expanded* y en el *Social Science Citation Index*; PF (*Performance per Faculty*), indica el desempeño académico por facultad.

A partir del año 2004 introducen cambios en la forma de hacer la evaluación total y la presentan anualmente. El indicador Nobel se transforma en Award, para incluir, además de los premios Nobel ya mencionados, las medallas Fields, y se introducen diferentes pesos según el año en que se obtuvo el premio, dándosele mayor peso a los premios más recientes. Además, el indicador PF desaparece, siendo sustituido por los indicadores Alumni y Size, pesados con un 10 por ciento cada uno; Alumni, indica el número de alumnos de una institución que ganó un premio (Nobel o Fields), y también se otorga un peso mayor a fechas más recientes; Size, indica los puntajes pesados de los otros cinco indicadores, divididos por el número de personal académico de tiempo completo. Dados estos cambios, para hacer los ajustes, se eligieron sólo tres indicadores: NS, SCI y Size, y se encontraron ajustes que alcanzan a los de los factores de impacto, y eso para todos los años en que han publicado información, o sea, del año 2003 al 2007.

La figura 3.2 muestra un histograma con la forma en que se ven los datos sin introducir una gráfica semilog, para el caso del indicador NS 2007, y representa la forma (o firma) característica de este comportamiento: la caída en la cola superior (la de los rangos altos). Después se muestra un ajuste para cada indicador: NS 2006, SCI 2005, y Size 2006, en las figuras 3.3, 3.4 y 3.5, respectivamente.

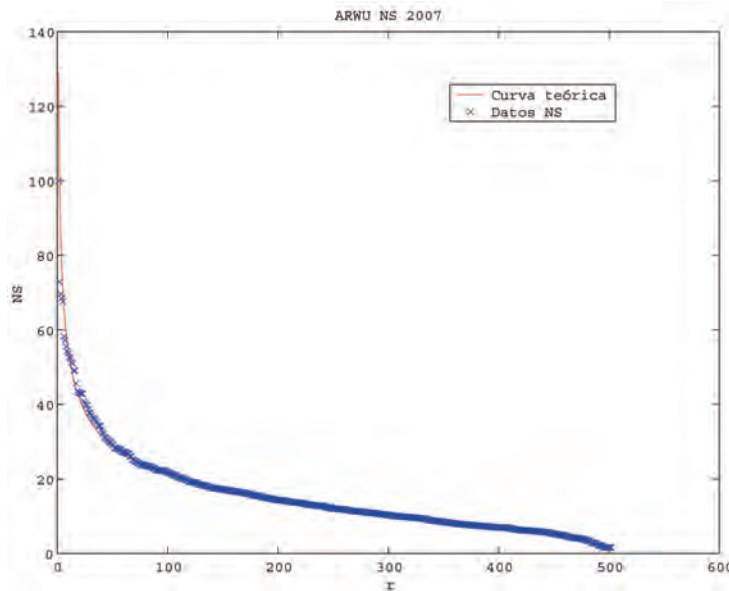


Figura 3.2: Histograma de NS para el año 2007.

Las tablas 3.2, 3.3 y 3.4 muestran los valores obtenidos de  $K$ ,  $b$  y  $a$ , y el coeficiente de regresión  $R^2$  para los indicadores NS, SCI y Size, en los años 2003-2007.

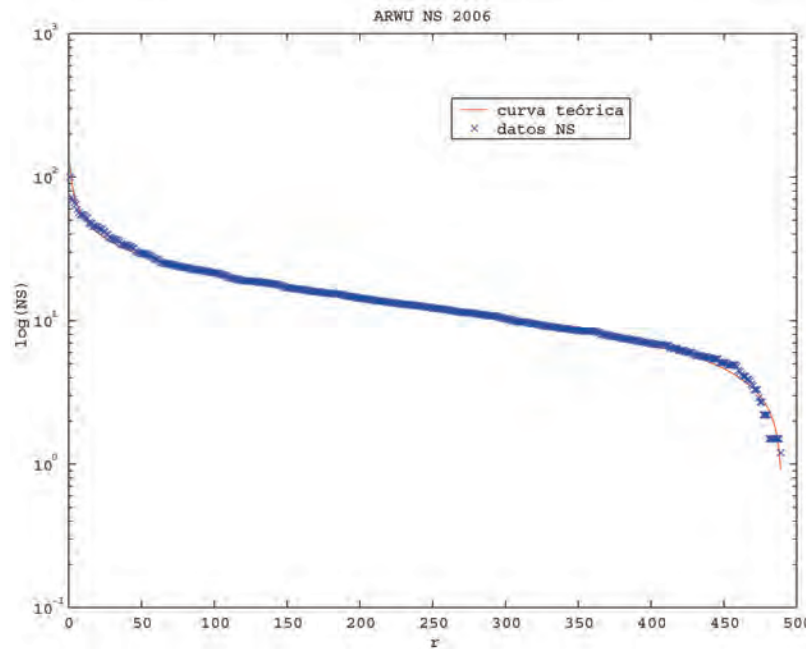


Figura 3.3: Gráfica semilog de NS para el año 2006.

Como una última observación de este apartado, hay que mencionar que los criterios y la forma de construcción de la mayoría de los indicadores para ARWU forman parte de los que se llaman *criterios objetivos bibliométricos*, es decir, sus datos no son proporcionados directamente por las instituciones y por tanto se consideran objetivamente confiables. Se dice que los análisis basados en ellos son reproducibles y rigurosos, siendo los más comunes los que corresponden a los indicadores Nobel, HiCi, NS, SCI, aunque con variaciones en los periodos de tiempo considerados. Los indicadores PF y Size caen dentro de los criterios objetivos no-bibliométricos, que se basan en la información que proporcionan las universidades, y que generalmente no estiman el impacto y calidad de las instituciones, sino más bien la infraestructura y el presupuesto que disponen.

### 3.5 Fluctuaciones y *rank-clocks*

Una manera de explorar el comportamiento observado en torno a la ley de potencias de dos colas encontrada (la distribución Beta de dos parámetros), es considerar los aspectos que atañen a las fluctuaciones involucradas. Es decir, algo que comparten los sitios donde ha aparecido esta ley de potencias es el hecho de que se presentan

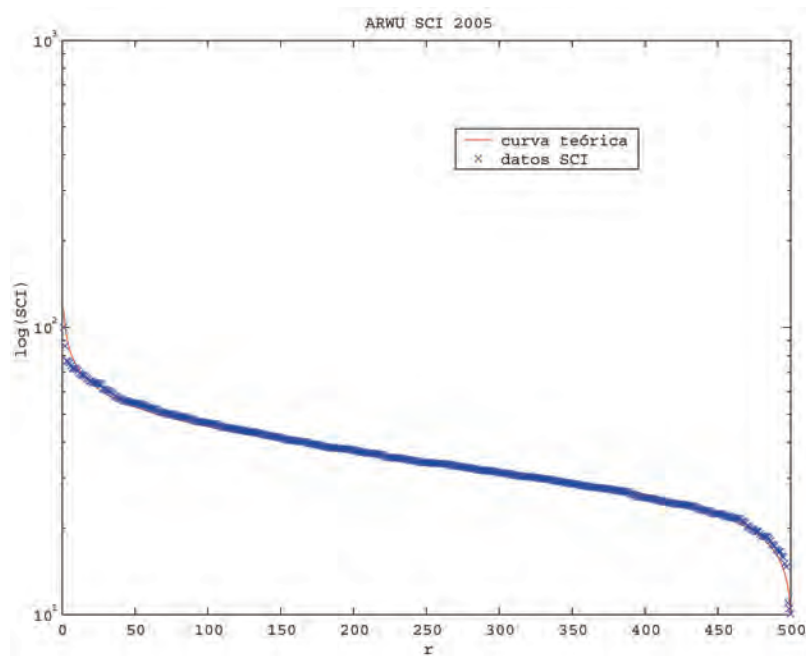


Figura 3.4: Gráfica semilog de SCI para el año 2005.

NS/ año	$K$	$b$	$a$	$R^2$
2003	8.9800	0.4270	0.3527	0.9921
2004	9.1358	0.4273	0.3581	0.9944
2005	10.2022	0.4095	0.3681	0.9936
2006	8.3103	0.4337	0.3579	0.9932
2007	9.0241	0.4278	0.3727	0.9948

Tabla 3.2: Valores encontrados para el indicador NS.

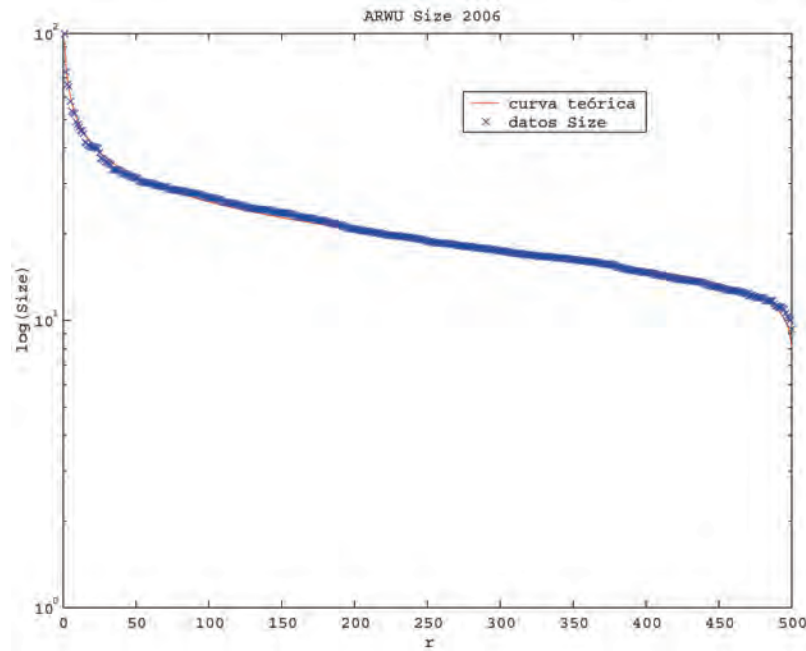


Figura 3.5: Gráfica semilog de Size para el año 2006.

SCI/ año	$K$	$b$	$a$	$R^2$
2003	38.4708	0.1839	0.2016	0.9906
2004	33.4616	0.2030	0.1940	0.9947
2005	33.8690	0.1980	0.1947	0.9932
2006	34.6674	0.1956	0.1943	0.9925
2007	35.2653	0.1910	0.1972	0.9903

Tabla 3.3: Valores encontrados para el indicador SCI.

Size/ año	$K$	$b$	$a$	$R^2$
2003	No data	--	--	--
2004	46.3305	0.1658	0.3848	0.9904
2005	50.7646	0.1059	0.3009	0.9970
2006	44.5896	0.1196	0.2736	0.9948
2007	44.5361	0.1207	0.2799	0.9947

Tabla 3.4: Valores encontrados para el indicador Size.

un gran número de fluctuaciones entre los datos de un año y otro. Así, en el caso de las ciudades del capítulo anterior, se tiene *grosso modo* que una ciudad que ocupa un primer lugar hoy, en el pasado ni siquiera existía y en el futuro seguramente desaparecerá.

Con esta idea en mente, Michael Batty, propone en [9], la construcción de lo que califica como *rank clocks*, que son formas gráficas que intentan condensar la información relativa a la volatilidad de los datos que dan lugar a las distribuciones observadas, mismas que exhiben una estabilidad a nivel macroscópico (como la ley de Zipf-Lotka para la estadística de ciudades, o los dos casos anteriores -FIS e indicadores de ARWU), pero que en una escala micro exhiben una dinámica volátil.

Un *rank clock* es una gráfica tipo coordenadas polares, donde una escala temporal se ubica girando en el sentido de las manecillas del reloj, y se colocan los mejores rangos al centro (rangos bajos - tamaños grandes); si los rangos crecen, los tamaños decrecen, y eso se interpreta como que el objeto representado se ha movido hacia la circunferencia, alejándose del centro. Así, se puede observar gráficamente si un elemento va mejorando su posición, al acercarse al centro, o si va empeorando, al irse saliendo.

Batty toma los datos de tres casos, relativos a las distribuciones de las  $N$  ciudades más pobladas; lo hace para tres conjuntos: los Estados Unidos de Norteamérica, el Reino Unido y a nivel mundial. El número de ciudades consideradas depende de las bases de datos y de los archivos estadísticos, pero en todos los casos encuentra que se cumple la ley de Zipf. Las diferencias radican en  $N$ , el número de ciudades, pues en los dos primeros casos los datos están basados en los censos de población y usualmente no contienen las poblaciones pequeñas, es decir se habla de un Top 100 o Top 50. Así, no se observa la cola derecha. Pero como se mencionó en el capítulo anterior, existe una cola para las poblaciones pequeñas, y como se ha mostrado en éste, tiene un grado importante de ubicuidad. En la figura (3.6) se tiene un rank clock para los datos del Reino Unido.

Para explorar la volatilidad de los datos, supongamos que una ciudad pujante hoy

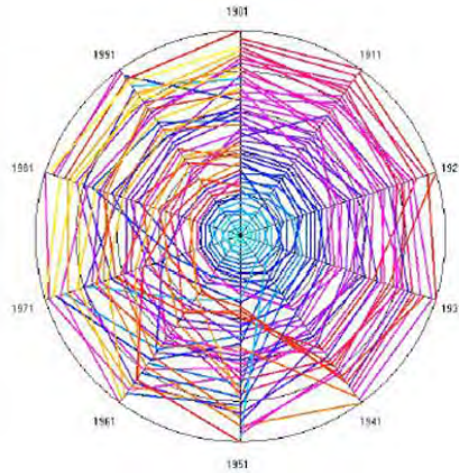


Figura 3.6: *Rank clock*, según Batty, para el Reino Unido.

está en el rango 34, en el pasado estuvo en otro rango menor, y en el futuro estará en otro rango aún mayor (si es que su pujanza no decae), lo que se verá como un giro hacia el centro; de tal suerte que se puede observar para cada caso cómo va cambiando de sitio desde que entra al Top 100, si logra entrar al Top 50 y si, en su caso, accede al Top 5, o incluso si logra ocupar el primer sitio. Desde luego, ello podría ir en sentido inverso, pues como se sabe, las ciudades declinan, lo que implica que se van vaciando de población.

En la figura 3.7 se muestra un rank clock para algún indicador y los nombres de algunas universidades, y en la gráfica 3.8 se muestra otro rank clock para el mismo indicador, pero ahora con 50 universidades (de un total de cerca de 500); en ambos casos el indicador es irrelevante, pues *el comportamiento cualitativo observado es el mismo*, independientemente del indicador y del año.

Por supuesto que lo descrito aquí no es lo único que hace Batty; por el contrario, define cierto tipo de cuentas para intentar dilucidar la estructura subyacente y propone a los interesados recurrir a información complementaria, así como a un programa de cómputo que contiene la base de datos que él empleó. Las referencias aparecen en [9].

En el siguiente capítulo se abordarán un par de modelos estocásticos que intentan dar cuenta de la microdinámica volátil, fluctuante, y que derivan a la ecuación 3.2, partiendo de consideraciones muy generales.



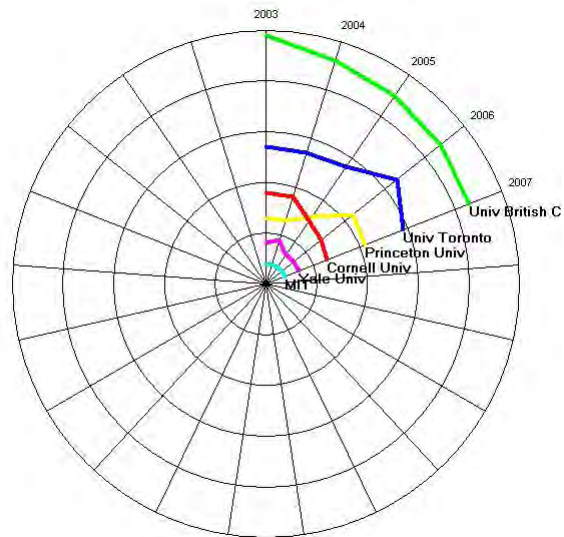


Figura 3.7: *Rank clock* años 2003-2007 para casos específicos.

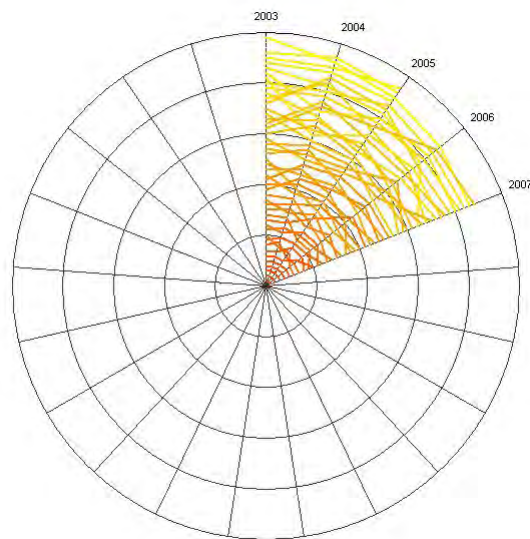


Figura 3.8: *Rank clock* años 2003-2007.

---

## Capítulo 4

# Modelaje

---

Como hemos visto en el capítulo anterior, existe una variedad grande de sistemas que dan lugar a las leyes de potencias de dos colas con una descripción específica: la distribución beta de dos parámetros. En este capítulo se presentan un par de modelos generativos que dan lugar a dicha distribución.

### 4.1 Ecuación maestra para comunidades ecológicas

Uno de los principales problemas en la teoría de la diversidad es poder explicar la abundancia de especies en comunidades ecológicas. Expuesta en su libro *The Unified Neutral Theory of Biodiversity and Biogeography*,<sup>1</sup> Stephen P. Hubbell elaboró una teoría neutral de la biodiversidad, la cual describe la dinámica de taxones con especies tróficamente similares en bosques o comunidades marinas que compiten por recursos alimenticios. En su enfoque, la dinámica poblacional obedece las reglas de un *juego de suma cero*: la comunidad estaría saturada biológicamente con individuos y a cada muerte seguiría rápidamente el nacimiento de un individuo de la misma especie o de otra. La dependencia temporal en una comunidad dada sería el resultado de la competición interna y de la inmigración del exterior, de lo que denomina *metacomunidad*. Hubbell supone, además, que en la metacomunidad se tiene *neutralidad*, o sea que todas las especies tienen la misma capacidad de competencia, con la misma probabilidad de morir y de reproducirse.

Así, los modelos que discute Hubbell obedecen la dinámica de un proceso de nacimiento y muerte. En particular, para la dinámica de la metacomunidad, en una aproximación de campo medio, se tiene que la probabilidad  $P_k(t)$ , donde  $k$  individuos pertenecen a la misma especie al tiempo  $t$ , obedece la ecuación maestra:

$$dP_k(t)/dt = A_{k-1}P_{k-1} + C_{k+1}P_{k+1} - (A_k + C_k)P_k$$

---

<sup>1</sup>Para una exploración electrónica se sugiere ver [31].

donde los coeficientes son las tasas de incremento y decremento de la abundancia de especies  $k$ , respectivamente, expresadas en unidades del tiempo de vida de un individuo. Estos coeficientes están dados por:

$$A_k = \frac{(J_m - k)k}{J_m(J_m - 1)}(1 - \nu),$$

$$C_k = \frac{k[J_m - k + (k - 1)\nu]}{J_m(J_m - 1)}$$

La tasa de incremento  $A_k$ , por ejemplo, es proporcional a la probabilidad de muerte de un individuo que no pertenece a la especie considerada  $\frac{(J_m - k)}{J_m}$ , multiplicada por la probabilidad de nacimiento de un individuo que pertenece a la especie considerada  $\frac{k}{(J_m - 1)}$ , multiplicada por la probabilidad de no mutación  $(1 - \nu)$ .

Cuando se considera la emigración, o sea cuando se busca la distribución de especies en una comunidad local en contacto con la metacomunidad, en uno de los modelos se tiene la siguiente representación integral:<sup>2</sup>

$$g_s(\omega) = \mu\theta \int_0^1 \binom{\mu}{\mu u} (1 - \omega)^{\mu u - 1} \omega^{\mu(1-u) - 1} u^\theta du$$

y utilizando el teorema del valor medio, podemos aproximar dicha integral por:

$$g_s(\omega) \approx \mu\theta \binom{\mu}{\mu \bar{u}} \bar{u}^\theta (1 - \omega)^b \omega^{-a} \quad (4.1)$$

donde  $b = \mu\bar{u} - 1$ ,  $a = 1 - \mu(1 - \bar{u})$ ,  $u < \bar{u} < 1$ . Si  $\theta$  no es pequeña,  $\bar{u} = 1 - \epsilon$ , con  $\epsilon$  pequeña. Tenemos así la representación  $\{a, b\}$  para la abundancia de especies.

## 4.2 Generalización de la ecuación maestra

Con base en lo expuesto anteriormente se ha propuesto una generalización de las ecuaciones maestras de nacimiento y muerte asociadas a los modelos de la dinámica neutral de las comunidades ecológicas.

Sea la ecuación maestra:

$$dP(k)/dt = F(k)P(k, t) + r_{k+1}P(k + 1, t) + g_{k-1}P(k - 1, t) - (g_k + r_k)P(k, t)$$

con  $r_k = W(k-1 | k)$  (tasa de *pérdida*),  $g_k = W(k+1 | k)$  (tasa de *ganancia*). Tomemos  $F(k)$  igual a cero, para encontrar la solución estacionaria, que está dada por:

<sup>2</sup>En lo que sigue, nos basamos en los resultados de [80].

$$P_s(k) = \frac{g_{k-1}g_{k-2}\dots g_0}{r_k r_{k-1} \dots r_1} P_s(0) \quad (4.2)$$

y se puede obtener  $P_s(0)$  a partir de la condición de normalización, es decir, de que

$$1 = \sum_{k=0}^N P_s(k) = P_s(0) + \sum_{k=1}^N P_s(k)$$

Para la probabilidad de crecimiento o de ganancia, podemos considerar el “efecto San Mateo” como un primer aporte a esta tasa de transición, es decir, algo proporcional a  $k$ , y para evitar problemas con los ceros, algo proporcional a  $(k + \text{cte.})$ . Además, podemos suponer que, como en el caso de la logística, hay un factor de saturación, así  $g_k \propto (k + c_1)f(k)$ , con  $f(k) = (N_1 - k)$ , y se cumple que  $k \leq N$  y  $N_1 \geq N + 1$

Tenemos pues,  $g_k$ , de la forma

$$g_k = \lambda_1(k + C_1)(N_1 - k)$$

Para que haya solución estacionaria, sin explosión o implosión factorial que haga que el proceso se anule o diverja, necesitamos que  $r_k$ , la tasa de pérdida, sea también un polinomio de segundo orden, para que así cada factor, de los  $k$  que hay en (4.2), esté equilibrado:

$$r_k = \lambda_2(k + C_2)(N_2 - k)$$

Entonces, para el caso estacionario, tenemos:

$$\begin{aligned} P_s(k) &\approx \frac{(k-1+C_2)\dots C_2(N_2+1-k)\dots N_2}{(k+C_1)\dots(C_1+1)(N_1-k)\dots(N_1-1)} \left(\frac{\lambda_1}{\lambda_2}\right)^k \\ &\longrightarrow \frac{(k+C_2-1)^{k+C_2-1}}{(k+C_1)^{k+C_1}} \frac{(N_1-1-k)^{N_1-1-k}}{(N_2-k)^{N_2-k}} \left(\frac{\lambda_1}{\lambda_2}\right)^k \end{aligned}$$

Para  $k$  grande, está última expresión va a

$$\frac{(N_2-k)^{N_1-1-N_2}}{(k+C_1)^{C_1-C_2+1}} \left(\frac{\lambda_1}{\lambda_2}\right)^k$$

Sea  $b = N_1 - N_2 - 1 > 0$ , i.e.  $N_1 - N_2 > 1$ ;  $0 < a < 1$ ,  $0 < C_2 - C_1 < 1$ ,  $N_1 > N_2 + 1$ ,  $C_1 < C_2 < C_1 + 1$ ,  $C_2 > C_1 > C_2 - 1$ , tendremos que (para  $k$  grande) la podemos escribir como:

$$P_s(k) \approx \frac{(N_2-k)^b}{(k+C)^a} \exp\left(-\ln\left(\frac{\lambda_2}{\lambda_1}\right)^k\right)$$

donde  $k = 1, 2, \dots, N$ ,  $N_2 = N_1 + 1$ , y  $N_2 \geq N - k$ , entonces:

$$P_s(k) \approx \frac{(N + 1 - k)^b}{(k + C)^a} e^{-\lambda k}$$

y nuevamente, para que no diverja, necesitamos que  $\lambda \geq 0$ . Si  $\lambda = 0$  implica que  $\lambda_1 = \lambda_2$ , en cuyo caso tendremos una ley de potencias de dos colas asociada a una distribución beta de dos parámetros, que era lo que queríamos encontrar.

### 4.3 Procesos estocásticos

Tanto la ecuación de Langevin como la ecuación de Fokker-Planck, constituyen los tipos de ecuaciones diferenciales estocásticas (EDE) más usados en física. Una tercera forma de escribir una EDE es mediante su forma diferencial, que es la que se usa en matemáticas y finanzas, y es la que se usó en el capítulo 2. Dado que las ecuaciones diferenciales estocásticas tienen que manejar de alguna manera los procesos de Wiener (para describir al movimiento browniano), es natural esperar que el cálculo diferencial e integral deba ser modificado para proceder formalmente, pues el proceso de Wiener no es diferenciable. Hay dos formas de hacerlo, el cálculo estocástico de Itô y el cálculo estocástico de Stratonovich. La formulación de Itô tiene ciertas ventajas formales, pues con él se construye de manera natural la notación diferencial, lo que le confiere una gran elegancia, y por ello es el más usado por los matemáticos; sin embargo, tiene la desventaja de que es necesario aprender nuevas reglas de cálculo. Por el contrario, la ventaja principal del cálculo de Stratonovich radica en que no hay que aprender nuevas reglas de cálculo, pero para beneficiarse del formalismo matemático, hay que encontrar ciertas reglas de transformación, y resulta un poco engorroso tener que hacer esas transformaciones cada vez que se quieren comparar resultados expresados en la formulación de Itô, por ello, en lo que sigue, optamos por adoptar esta última.

Debe quedar claro que lo que queremos hacer en esta sección es encontrar condiciones generales para que obtengamos una función beta de dos parámetros.

#### 4.3.1 Ecuación de Langevin

La ecuación de Langevin surge en el contexto del movimiento browniano, donde se considera la ecuación de movimiento de una partícula de masa  $m$ , sujeta al bombardeo constante de moléculas de un medio en equilibrio termalizado (de masa mucho menor a  $m$ ):

$$m \frac{dv}{dt} = m\dot{v} = -\gamma v + F(t) \quad (4.3)$$

El término  $-\gamma v$  constituye la fuerza de fricción, opuesta a la velocidad; y  $F(t)$  es la fuerza que toma en cuenta los bombardeos moleculares aleatorios. Haciendo  $\beta = -\gamma$ , y  $\Gamma(t) = F(t)/m$  (conocida como fuerza de Langevin), la ecuación 4.3 queda de la siguiente manera:

$$\dot{v} = \beta v + \Gamma(t) \quad (4.4)$$

A la función aleatoria  $\Gamma(t)$  se le piden ciertas condiciones, las llamadas condiciones de normalización: 1) que su promedio sea nulo, para que no haya efectos netos en una dirección específica; y 2) que si se multiplican dos fuerzas de Langevin a tiempos distintos,  $t, t'$ , su promedio sea cero. Es decir, que se cumplan:

$$\langle \Gamma(t) \rangle = 0$$

y que siempre que  $\Delta = |t - t'| \geq \tau_0$ , donde  $\tau_0$  es el tiempo que dura la colisión:

$$\langle \Gamma(t)\Gamma(t') \rangle = 0$$

Esta segunda condición es razonable, pues las colisiones de diferentes moléculas del fluido con la partícula que observamos son aproximadamente independientes. Usualmente, el tiempo de colisión,  $\tau_0$ , es mucho menor al tiempo de relajación de la velocidad de la partícula,  $\tau = 1/\gamma$ . Por tanto, es razonable tomar el límite  $\tau_0 \rightarrow 0$ , como una buena aproximación, teniendo la condición:

$$\langle \Gamma(t)\Gamma(t') \rangle = q\delta(t - t') \quad (4.5)$$

y se dice que  $\Gamma(t)$  está  $\delta$ -correlacionada, con intensidad  $q$ . También es común decir que la fuerza de Langevin,  $\Gamma(t)$ , es *ruido blanco*, y eso se debe a que su densidad espectral es independiente de la frecuencia  $\omega$ .

Una generalización de la ecuación de Langevin, en la forma de la ecuación 4.4, con la variable estocástica  $\xi$ , perturbada por ruido blanco, se vería:

$$\dot{\xi} = A(\xi, t) + B(\xi, t)\Gamma(t) \quad (4.6)$$

Y sujeta a las condiciones de normalización, (la constante  $q$  se absorbe en la función  $B$ ):

$$\langle \Gamma(t) \rangle = 0$$

$$\langle \Gamma(t)\Gamma(t') \rangle = Q\delta(t - t')$$

Si  $B = cte.$ , se dice que se tiene ruido aditivo. Si  $B = B(\xi)$ , se dice que se tiene ruido multiplicativo, pero eso es un tanto arbitrario, pues se pueden hacer ciertos cambios para que siempre se elimine el factor que multiplica a  $\Gamma(t)$ .

### 4.3.2 Construcción de la ecuación de Fokker-Planck

Hay varias maneras de obtener la ecuación de Fokker-Planck:

- Partiendo de la ecuación de Chapman-Kolmogorov, que hace uso de las tasas de transición entre estados, para un proceso markoviano, ya sea usando un generador infinitesimal o mediante un desarrollo de Taylor a segundo orden.
- Construyendo, a partir de la ecuación maestra generalizada (anteriormente expuesta), la análoga para el caso continuo mediante el cálculo en diferencias.
- Desarrollando la ecuación diferencial estocástica del valor promedio de la función  $P(X)$ , a partir del cálculo de Itô.

Las últimas dos construcciones se encuentran en el apéndice. Por ahora sólo se hará la siguiente observación: en las condiciones de normalización se introdujo el factor  $A$ ; resulta que en la literatura aparecen varias normalizaciones; las más frecuentes son: i)  $Q = 2$ , ii)  $Q = 1$ , lo cual tiene implicaciones en la forma de la ecuación de Fokker-Planck resultante. En el primer caso, el término difusivo requiere un factor 1; y en el segundo caso, un factor igual a  $1/2$ . Optaremos por este caso, en el entendido de que bien se podría haber usado el otro.

### 4.3.3 De la ecuación de Langevin a la de Fokker-Planck

En el caso de una dimensión, donde  $A$  y  $B$  no dependen de  $t$ , la ecuación de Langevin en su forma 4.6, pero en la variable  $k$ , queda:

$$\dot{k} = A(k) + B(k)\Gamma(t) \quad (4.7)$$

y la ecuación de Fokker-Planck, con el factor  $Q = 1$  en la normalización, queda de la forma:

$$\frac{\partial P(k, t)}{\partial t} = -\frac{\partial(A(k)P)}{\partial k} + \frac{1}{2} \frac{\partial^2(B(k)^2 P)}{\partial^2 k} \quad (4.8)$$

Es decir:

$$\frac{\partial P(k, t)}{\partial t} = -\frac{\partial}{\partial k} \left[ A(k)P - \frac{1}{2} \frac{\partial(B(k)^2 P)}{\partial k} \right] \quad (4.9)$$

Los detalles de la construcción de la ecuación de Fokker-Planck, partiendo de la ecuación de Langevin, se pueden consultar en el apéndice.

#### 4.3.4 Caso estacionario

Para encontrar la solución del caso estacionario de la ecuación 4.9, en cuyo caso el lado izquierdo  $\dot{P}(k, t) = 0$ , tenemos que resolver la siguiente ecuación:

$$0 = -\frac{\partial}{\partial k} \left[ A(k)P_s - \frac{1}{2} \frac{\partial(B^2 P_s)}{\partial k} \right]$$

Integrando, y tomando la constante de integración nula, es equivalente a resolver:

$$0 = \left[ AP_s - \frac{1}{2} \frac{\partial(B^2 P_s)}{\partial k} \right]$$

$$0 = 2P_s \frac{[AP_s - BB']}{B^2} - \frac{dP_s(k)}{dk}$$

Así, si separamos variables, tendremos:

$$\frac{dP_s}{P_s} = \frac{2[A - BB']dk}{B^2} \quad (4.10)$$

Considerando los resultados de las comunidades ecológicas, podemos tomar:

$$A(k) = \alpha(N_1 - k) + \beta(k + C_1) \quad (4.11)$$

$$B(k) = [(N_2 - k)(k + C - 2)]^{\frac{1}{2}} \quad (4.12)$$

Haciendo  $\Omega$  el factor que multiplica a  $dk$  en la ecuación (4.10), y calculando  $BB'$ , tenemos que:

$$BB' = \frac{[N_2 - 2k - C + 2]}{2}$$

así, separando por fracciones parciales:

$$\Omega = \frac{2A - (N_2 - 2k - C + 2)}{(N_2 - k)(k + C - 2)} = \frac{A_0}{N_2 - k} + \frac{B_0}{k + C - 2}$$

integrando, con  $u = N_2 - k$  y  $v = k + C - 2$  y  $C_0 = cte.$  tenemos:

$$\ln P_s = -A_0 \ln u + B_0 \ln v + C_0$$

Por lo tanto, la solución estacionaria es:

$$P_s(k) = C_0(N_2 - k)^{-A_0}(k + C - 2)^{B_0}$$

Y se encuentra que los coeficientes de las fracciones parciales son:



$$A_0 = \frac{2[\alpha(N_1 - N_2) + \beta(N_2 + C_1)]}{N_2 + C - 2} + 1$$

$$B_0 = \frac{2[\alpha(N_1 + C - 2) + \beta(C_1 - C + 2)]}{N_2 + C - 2} - 1$$

Si hacemos  $N_1 = N_2 = N$ , y  $C_1 = C - 2 = \bar{C}$ , tendremos que:

$$A_0 = \frac{2\beta(N + \bar{C})}{N + \bar{C}} + 1 = 2\beta + 1 = b$$

$$B_0 = \frac{2\alpha(N + \bar{C})}{N + \bar{C}} - 1 = 2\alpha - 1 = a$$

Y finalmente tenemos:

$$P_s(k) = C' \frac{(N - k)^a}{(k + \bar{C})^b} \quad (4.13)$$

Que es nuevamente una ley de potencias de dos colas asociada a una distribución beta de dos parámetros. En otras palabras, la solución del caso estacionario de la ecuación de Fokker-Planck (4.8), con las condiciones (4.11) y (4.12), es una función beta de dos parámetros. Vemos también que, de acuerdo al ajuste que hemos realizado en las gráficas de los datos del ARWU, se tiene que  $\bar{C} = 0$ , lo que es análogo a lo que pasa en el caso de Zipf, sin la corrección de Mandelbrot.

#### 4.3.5 Observaciones

Antes de continuar, debemos hacer una observación importante. Vimos que partiendo de la ecuación de Langevin se construye la ecuación de Fokker-Planck, pero lo inverso no es unívoco, y ello acota las posibilidades que pudieramos tener para interpretar nuestros resultados.

# Discusión y conclusiones

---

Por lo expuesto en el capítulo anterior, tenemos al menos dos modelos generativos de la distribución beta de dos parámetros.

Sugerido por la ubicuidad, tales modelos no pueden dar cuenta de todos los sitios donde aparece la función beta de dos parámetros, pues la naturaleza tan diversa donde aparece constituye una limitante importante. El hecho de mostrar dos modelos sugiere también que habría posiblemente toda una pléyade de mecanismos que den lugar a la relación funcional. De hecho, los mapeos unimodales en secuencias genéticas [11] se basan en el trabajo de W. Li sobre expansión-modificación y sugieren analogías del tipo de las mutaciones. Así pues, los valores de los exponentes se pueden asociar a la lucha entre lo que permanece (se expande) y lo que cambia (se modifica). De hecho, en más de un sentido.

Cuando observamos los rank-clocks para alguno de los indicadores de las universidades, se tiene la sensación de observar el cambio en el tiempo, la transitoriedad de una posición, y si eso fuera lo único que se observara, difícilmente se pensaría que hay algo que se preserve: la forma funcional de la distribución. De hecho, el modelo convencional para los asentamientos humanos, que hace uso de la ley de Gibrat (un proceso de Yule, o el *preferential attachment*), y que se emplea en el capítulo 2, no refleja la característica de volatilidad, pero tendría que incluirse para considerar la evolución y quizás poder predecir.

Lo que se observa es que las universidades que se encuentran entre los primeros sitios, tienden a quedarse alrededor de ellos, es decir, es difícil decir cuándo decaerán para ceder su sitio a otra universidad.

Por ahora, las reflexiones en torno a las fluctuaciones y el trabajo de Batty indican que hay factores que permanecen (en el caso de las universidades, las que se encuentran en los primeros lugares -y quizás también las que están en los últimos, y habría que construir rank-clocks inversos-), y hay factores que cambian mucho (en el caso de las universidades, las que están por el medio); quizás en la representación  $\{a, b\}$  se encuentre sintetizada alguna información al respecto.

Por lo pronto, pensando en el primer modelo, y tratando de asociarlo a las universidades, no queda claro cómo hacer la identificación, pues la neutralidad parece no cumplirse para las universidades; más bien parece que el número de universidades crece, y habrá que esperar a recabar más datos para los próximos años, o de plano, mapear el ranking hacia años pasados. La idea es tener más mediciones, para poder observar, si fuera el caso, algo de estructura, o agrupamientos; una visita reciente al sitio de la Universidad de Shanghai, muestra que ahora intentan agrupar la información también por área, y una tarea pendiente sería indagar si en este nuevo caso se hallan funciones Beta.

Para el caso de la ecuación de Langevin, podemos pensar que el término de arrastre condensa información sobre la resistencia a moverse de una universidad, digamos que tendería a permanecer en su sitio; el otro sumando indicaría el resultado neto de los factores que la empujan a cambiar de sitio... así, viendo la forma de las condiciones auxiliares para encontrar la solución estacionaria, uno puede asociar a los exponentes el cambio y la permanencia.

## 5.1 Conclusiones

Podemos decir que la distribución beta de dos parámetros presenta un muy buen ajuste en todos los casos citados, y para el intervalo completo de definición de los datos; tal característica la coloca en una posición de ventaja con respecto a las otras distribuciones; además de ser ubicua y corresponder a una cantidad de fenómenos de naturaleza diversa, se destacan los de carácter social, como lo son los indicadores del ranking mundial de las universidades y los factores de impacto, y sugieren que quizás haya muchos más lugares donde aparezca.

Basados en los resultados de [49] podemos tener la esperanza de seguir trabajando para encontrar una distribución límite bajo ciertas condiciones, que por ahora se presentan teóricamente complicadas, pero que en caso de hallarse harían frente a la diversidad de naturalezas donde ha aparecido.

---

Capítulo 6

# Apéndice

---

Hay varias maneras de construir la ecuación de Fokker-Planck. Se presentan dos. Pero para un tratamiento riguroso y formal, se sugiere consultar las referencias [65] y [67].

## 6.1 Desde una ecuación maestra generalizada

Se puede también construir una ecuación de Fokker-Planck a partir de la ecuación maestra, pues será el equivalente de ésta en caso continuo. Tomemos la ecuación maestra que hemos trabajado:

$$dP(k)/dt = r_{k+1}P(k+1, t) + g_{k-1}P(k-1, t) - (g_k + r_k)P(k, t)$$

Consideremos el cálculo en diferencias, donde, para una función  $g$ :

$$\frac{\partial h}{\partial x} = \lim_{\epsilon \rightarrow 0} \frac{h(x + \epsilon) - h(x)}{\epsilon}$$

$$\frac{\partial^2 h}{\partial x^2} = \lim_{\epsilon \rightarrow 0} \frac{h(x + \epsilon) + h(x - \epsilon) - 2h(x)}{\epsilon^2}$$

Usando tales expresiones, podemos manipular la ecuación maestra, agregando nuevos términos, pero para hacerlo, introduzcamos una notación apropiada. Sean:  $P_+ = P(k+1)$ ,  $P_- = P(k-1)$  y  $P = P(k)$

Nuestra ecuación maestra se transforma en:

$$\Phi = \frac{dP}{dt} = r_+P_+ + g_-P_- - (g + r)P = (r_+P_+ - rP) + (g_-P_- - gP)$$

Primero sumamos a cada paréntesis un sumando adecuado, para tener:

$$\Phi = (r_+P_+ + r_-P_- - rP) - r_-P_- + (g_-P_- + g_+P_+ - gP) - g_+P_+$$

expresión a la que se le puede ordenar como:

$$\Phi = (r_+P_+ + r_-P_- - 2rP) - r_-P_- + rP + (g_-P_- + g_+P_+ - 2gP) - g_+P_+ + gP$$

De manera que se puede ya ver una forma reconocible, en términos de la segunda derivada. La otra parte del cálculo se basa en que, si  $\Delta(f) = f(x+1) - f(x) = f_+ - f$ , entonces  $\Delta(fg) = fg - f_-g_- = f_+g_+ - fg$ , y por tanto,  $\Delta(fg) = \frac{1}{2}(f_+g_+ - f_-g_-)$ , que se usa como equivalente de la derivada del producto.

Por otro lado, en la notación de las deltas, la segunda derivada se ve como:  $\Delta^2 f = f_+ + f_- - 2f$ .

Así, tenemos que:

$$\Phi = \frac{1}{2}\Delta^2(rP) + \frac{1}{2}\Delta^2(gP) + \frac{1}{2}(r_+P_+ - r_-P_-) - \frac{1}{2}(g_+P_+ - g_-P_-)$$

Entonces,

$$\Phi = \frac{1}{2}\Delta^2(rP) + \frac{1}{2}\Delta^2(gP) + \Delta(rP) - \Delta(gP)$$

Finalmente, identificando términos y reagrupando sumandos:

$$\frac{\partial P(x, t)}{\partial t} = -\frac{\partial((g-r)P)}{\partial x} + \frac{1}{2}\frac{\partial^2((r+g)P)}{\partial^2 x}$$

## 6.2 Dentro del cálculo de Itô

Consideremos la ecuación de Langevin generalizada (ecuación 4.6), con las condiciones de normalización correspondientes, pero cambiemos un poco la notación, empleando las usuales  $f$  y  $g$  para las funciones  $A$  y  $B$ , esto es:

$$\dot{\xi} = h(\xi, t) + g(\xi, t)\Gamma(t)$$

Queremos hallar una ecuación para la distribución<sup>1</sup>  $P$ , de forma tal que  $P = P(x; t)$ . La ecuación de Fokker-Planck es una ecuación de movimiento de la distribución de probabilidad, y la describe en términos de las variables macroscópicas fluctuantes, por lo que se presta a las consideraciones que se hacen desde el punto de vista “físico” y que corresponden al desarrollo de la ecuación diferencial estocástica del valor promedio de la función  $P(X)$ .

<sup>1</sup>En física es muy común hablar de *distribución de probabilidad*, donde lo correcto sería emplear *densidad de probabilidad*. Para no cambiar lo que ya se ha convertido en costumbre, nos apegamos a ella.

Sea  $\tau > 0$ , y  $\xi(t) = x$ , entonces, escribiendo la ecuación de Langevin en forma integral, tenemos:

$$\xi(t + \tau) - x = \int_t^{t+\tau} [h(\xi(t'), t') + g(\xi(t'), t')\Gamma(t')] dt' \quad (6.1)$$

Suponiendo que  $h$  y  $g$  se pueden expandir en serie de Taylor, y denotando la derivada con respecto a  $x$  con la notación primada usual,<sup>2</sup> entonces:

$$h(\xi(t'), t') = h(x, t') + h'(x, t')(\xi(t + \tau) - x) + \dots$$

$$g(\xi(t'), t') = g(x, t') + g'(x, t')(\xi(t + \tau) - x) + \dots$$

Sustituyendo esta expansión en la ecuación integral 6.1, se tiene, para los términos en  $h$ , y los términos en  $g$ :

$$\begin{aligned} \xi(t + \tau) - x &= \int_t^{t+\tau} h(x, t') dt' + \int_t^{t+\tau} h'(x, t')(\xi(t') - x) dt' + \dots + \\ &+ \int_t^{t+\tau} g(x, t')\Gamma(t') dt' + \int_t^{t+\tau} g'(x, t')(\xi(t') - x)\Gamma(t') dt' + \dots \end{aligned}$$

Como en muchos términos a integrar aparece el factor  $(\xi(t') - x)$ , podemos introducir nuevamente la expansión de Taylor para cada término; haciéndolo reiteradamente tenemos como integrandos solamente productos de las funciones conocidas  $h(x, t)$ ,  $g(x, t)$ , sus derivadas, y  $\Gamma(t)$ . Una vez hecho eso, tomamos el promedio y usamos las condiciones de normalización. Así, todos los términos donde aparezca multiplicando la fuerza de Langevin, se anulan; y donde aparezcan productos de la fuerza de Langevin (a tiempos distintos), se pueden sustituir por sus correlaciones, lo que deja:

$$\begin{aligned} \langle \xi(t + \tau) - x \rangle &= \int_t^{t+\tau} h(x, t') dt' + \int_t^{t+\tau} \int_t^{t'} h'(x, t') h(x, t'') dt'' dt' + \dots + \\ &+ \int_t^{t+\tau} g'(x, t') \int_t^{t'} g(x, t'') 2\delta(t'' - t') dt'' dt' + \frac{1}{2} \int_t^{t+\tau} \int_t^{t+\tau} g(x, t') g(x, t'') 2\delta(t' - t'') dt' dt'' + \dots \end{aligned}$$

Si se hacen las cuentas, se observa que los términos que sobreviven son los términos correspondientes a los primeros dos momentos de una cierta distribución de probabilidad.

<sup>2</sup>Es decir:  $\frac{\partial}{\partial \xi(t')} h(\xi(t'), t')|_{\xi(t')=x} \equiv \frac{\partial}{\partial x} h(x, t') \equiv h'(x, t')$

Quedando así, la ecuación de Fokker-Planck, ec. (4.8):

$$\frac{\partial P(x,t)}{\partial t} = -\frac{\partial(h(x)P)}{\partial x} + \frac{1}{2} \frac{\partial^2(g(x)^2 P)}{\partial^2 x} \quad (6.2)$$

# Bibliografía

---

- [1] Abramson, Norman (1974). *Teoría de la información y codificación*, tercera edición, Paraninfo, España.
- [2] Aitchison, J., Brown, J. A. C. (1957). *The Lognormal Distribution*, Cambridge University Press.
- [3] Amit, Daniel J. (1978). *Field Theory, the Renormalization Group and Critical Phenomena*, McGraw Hill.
- [4] Araújo Ruiz, J. y Arencibia Jorge, R. (2002). “Informetría, bibliometría y ciencias: aspectos teórico-prácticos”, *Acimed, Revista Cubana de los Profesionales de la Información y la Comunicación en Salud*. Vol. 10, no. 4, julio-agosto. <http://bvs.sld.cu/revistas/aci/indice.html>
- [5] Arita, H., Rodríguez, P. y Solís, L. (2000). “De ratones, hombres y elefantes: el tamaño sí importa”, *Ciencias* No. 57. enero-abril, pp. 28-38.
- [6] Bak, Per (1996). *How Nature Works: The Science of Self-Organized Criticality*, Copernicus Springer-Verlag.
- [7] Barabási, A. L., Stanley, H. E. (1995). *Fractals Concepts in Surface Growth*, Cambridge University Press.
- [8] Barabási, Albert-László (2002). *Linked: The New Science of Networks*, Perseus. Cambridge Massachusetts.
- [9] Batty, Michael (2006). “Rank clocks”, *Nature* 444, pp. 592-596.
- [10] Batty, M., Longley, P. (1996). *Fractal Cities*, Academic Press.
- [11] Beltrán del Río García, Manuel (2005). “Leyes de potencias en secuencias genéticas y en la dinámica simbólica de mapeos unimodales”. Tesis de licenciatura (física), FC-UNAM.



- [12] Beltrán del Río, M., Cocho, G., Naumis, G. G. (2008). “Universality in the tail of musical note rank distribution”. *Physica A: Statistical Mechanics and its Applications*, Vol 387, Issue 22, pp. 5552-5560.
- [13] Benítez Keinrad, Mariana (2004). “Criticalidad autoorganizada en sistemas ecológicos”. Tesis de licenciatura (biología), FC-UNAM.
- [14] Braun, Eliezer (1999). *Caos, fractales y cosas raras*, FCE, México. Disponible en: <http://omega.ilce.edu.mx:3000/sites/ciencia/volumen3/ciencia3/150/htm/caos.htm>
- [15] Callen, Herbert B. (1985). *Thermodynamics and an Introduction to Thermostatistics*, John Wiley & Sons.
- [16] Carmona, Gerardo (2000). *Fenómenos críticos y el grupo de renormalización*, FC-UNAM.
- [17] Çinlar, Erhan (1975). *Introduction to Stochastic Processes*, Prentice Hall.
- [18] Cocho, G., Miramontes, P. (2000). “Patrones y procesos en la naturaleza. La importancia de los protectorados”, *Ciencias* No. 59. julio-septiembre, pp. 12-20.
- [19] Crow, E. L., Shimizu, K. (editores) (1988). *Lognormal Distribution: Theory and Applications*, Markel Dekker, Inc., New York.
- [20] Dubrelle, B., Graner, F., Sornette, D., editores (1997). *Scale Invariance and Beyond*, Les Houches Workshop, March 10-14, 1997, Series Les Houches, EDP Sciences and Springer, Berlín.
- [21] Dresden, Arnold (1922). “A report on the scientific work of the Chicago section, 1897-1922”, *Bulletin of the American Mathematical Society* 28, pp. 303-307.
- [22] Feder, Jens (1988). *Fractals*, Plenum Press.
- [23] Frame, M., Mandelbrot, B., Neger, N. (2004). *Fractal Geometry*, Yale University. Disponible en: <http://classes.yale.edu/fractals/index.html>
- [24] Gabaix, Xavier (1999). “Zipf’s Law for Cities: An Explanation”, *The Quarterly Journal of Economics*, 114, pp. 739-967.
- [25] García Naumis, Gerardo (2002). “Los fractales: una nueva geometría para describir el espacio geográfico”. Conferencia presentada en el simposio: “La reurbanización de la Ciudad de México”. Disponible en: [http://www.fisica.unam.mx/~naumis/index\\_archivos/fractales.htm](http://www.fisica.unam.mx/~naumis/index_archivos/fractales.htm)

- [26] Garfield, Eugene (1955). "Citation Indexes for Science. A New Dimension in Documentation through Association of Ideas", *Science* 122 (3159), pp. 108-111. Disponible en:  
<http://www.garfield.library.upenn.edu/pub.html>
- [27] Gouyet, Jean-Francois (1996). *Physics and Fractal Structures*, Springer-Verlag.
- [28] Gnedenko, Boris V. (1978). *The Theory of Probability*, English translation, Mir Publishers.
- [29] Hacking, Ian (1991). *La domesticación del azar*, Gedisa.
- [30] Huang, Kerson (1987). *Statistical Mechanics*, 2nd. Edition, John Wiley & Sons.
- [31] Hubbell, Stephen P. (2001). *The Unified Neutral Theory of Biodiversity and Biogeography*. El capítulo 1 está disponible en:  
<http://www.pupress.princeton.edu/chapters/s7105.html>
- [32] Jiménez Contreras, Evaristo (2000). "Los métodos bibliométricos: estado de la cuestión y aplicaciones", ponencia presentada en el I Congreso Universitario de Ciencias de la Documentación. Teoría, historia y metodología de la documentación en España (1975-2000), celebrado del 14 al 17 de noviembre de 2000. Disponible en: <http://www.ucm.es/info/multidoc/multidoc/revista/num10/paginas/pdfs/ejcontreras.pdf>
- [33] Johnson, N. L., Kotz, S. (1970). *Continuos Univariate Distributions in Statistics*. Vol. I. John Wiley & Sons, USA.
- [34] Kadanoff, Leo P. (1993). *From Order to Chaos. Essays: Critical, Chaotic and Otherwise*, World Scientific Series on Nonlinear Science, Series A, Vol. 1, Singapore.
- [35] Kao, Edward P. C. (1997). *An Introduction to Stochastic Processes*, Duxbury Press, USA.
- [36] Kim, Myungshik. Chapter 4: *Brownian motion*. Disponible en su *Homepage*:  
<http://web.am.qub.ac.uk/users/m.s.kim/chap4.pdf>
- [37] Koski, Timo (2005). *Power Law Distributions in Bioinformatics*.
- [38] Li, Wentian (2003). "References on Zipf's Law".  
Del sitio: <http://linkage.rockerfeller.edu/wli/zipf/>

- [39] Li, W. (1991). "Expansion-modification systems: a model for spatial  $1/f$  spectra", *Physical Review A*, 43(10), pp. 5240-5260.  
Disponible en: <http://www.nslj-genetics.org/wli/pub/pr91.pdf>
- [40] Lotka, Alfred J. (1926). "The frequency distribution of scientific productivity", *Journal of the Washington Academy of Science* 16, pp. 317-323. Disponible en el sitio: <http://www.bibliometria.com/ley-de-lotka>
- [41] Lugo Velez, Carlos Antonio (2003). "Procesos multiplicativos y la distribución de los tamaños de las ciudades". Tesis de licenciatura (física), FC-UNAM.
- [42] Mandelbrot, Benoit (1971). *Fractals: Form, Change and Dimension*, Freeman and Company, USA.
- [43] Mansilla, R., Köppen, E., Cocho, G. y Miramontes, P. (2007). "On the Behavior of Journal Impact Factor Rank-Order Distribution", *Journal of Informetrics*, Vol. 1, Issue 2, abril, pp. 155-160. También disponible en [arXiv:cs.IR/0610091](http://arxiv.org/abs/cs/0610091).
- [44] Mansilla, R., Cocho, G. (2000). "Multiscaling in expansion-modification systems: an explanation for the long-range correlation in DNA", *Complex Systems*, 12, pp. 207.
- [45] Miramontes, Pedro (1999). "El color del ruido", *Ciencias* No. 54. abril-junio, pp. 4-11.
- [46] Mitzenmacher, Michael. "A Brief History of Generative Models for Power Law and Lognormal Distributions". Una primera versión está disponible en: <http://www.eecs.harvard.edu/~michaelm/NETWORK/papers.html>  
Existe una versión publicada en la revista *Internet Mathematics*, vol 1, no. 2, pp. 226-251, 2004, y disponible en: <http://www.internetmathematics.org/>
- [47] Montroll, E. W., Shlesinger, M. F. (1983). "Maximum Entropy Formalism, Fractals, Scaling Phenomena, and  $1/f$  Noise: A Tale of Tails", *Journal of Statistical Physics*, Vol. 32, N. 2, pp. 209-230.
- [48] Naumis, G. G., Cocho, G. (2007). "Tail universalities in rank distributions as an algebraic problem: the beta-like function", [arXiv:0705.0551](http://arxiv.org/abs/0705.0551)
- [49] Naumis, G. G., Cocho, G. (2007). "The tails of rank-size distributions due to multiplicative processes: from power laws to stretched exponentials and beta-like functions", *New Journal of Physics* 9 (286), pp. 286-292.
- [50] Nelson, Edward. Sección Books, del sitio:  
<http://www.math.princeton.edu/~nelson/books/>  
El fichero es: <http://www.math.princeton.edu/~nelson/books/bmotion.pdf>

- [51] Newman, M. E. J. (2005). "Power laws, Pareto distributions and Zipf's law", University of Michigan, Ann Arbor, USA. [arXiv:cond-mat/0412004](https://arxiv.org/abs/cond-mat/0412004).
- [52] Noyola Iglesias, Arturo (1971). *Antología de física*, Lecturas Universitarias No. 9, UNAM.
- [53] OERI - Observatorio de Estudios Relacionados con la Información, de la Escuela Graduada de Ciencias y Tecnologías de la Información (EGCTI) del Recinto de Río Piedras de la Universidad de Puerto Rico; entrada: "Bibliometría, Cienciometría". Disponible en: [http://oeri.uprrp.edu/index.php?option=com\\_content&task=view&id=65&Itemid=68](http://oeri.uprrp.edu/index.php?option=com_content&task=view&id=65&Itemid=68)
- [54] Parzen, Emanuel (1987). *Teoría moderna de probabilidad y sus aplicaciones*, Limusa, México.
- [55] Persky, Joseph (1992). "Retrospectives. Pareto's Law", *Journal of Economic Perspectives*, Vol. 6, No. 2, pp. 181-192.
- [56] Popescu, Ioan-Iovitz (2003). "On a Zipf's Law extension to impact factors", *Glottometrics*, Vol. 6, pp. 83-93. Disponible en: [http://www.geocities.com/iipopescu/Zipf\\_Law.html](http://www.geocities.com/iipopescu/Zipf_Law.html)
- [57] Ravichandra Rao, I. K. (1998). "Classical Laws of Informetrics: An Overview", DRTC Workshop on Informetrics and Scientometrics, 16-19 March 1998, Bangalore.
- [58] Redner, Sidney (1990). "Random Multiplicative Processes: An Elementary Tutorial", *Am. J. Phys.* Vol. 58, p. 267. Disponible en: <http://physics.bu.edu/redner/pubs.html>  
<http://physics.bu.edu/redner/pdf/ajp58p267.pdf>
- [59] Reed, William J. (1999). "The Pareto Law of Incomes - an Explanation and an Extension", *Journal of Bussiness and Economic Statistic (Physica A)*, 319, pp. 469-486. Disponible en: <http://www.math.uvic.ca/faculty/reed/income.2.pdf>
- [60] Reed, William J. (2000). "On the Rank-size Distribution for Human Settlements". Disponible en: <http://www.math.uvic.ca/faculty/reed/Rank-size.pdf>
- [61] Reed, William J. (2000). "The Pareto, Zipf and Other Power Laws". Disponible en: [http://www.math.uvic.ca/faculty/reed/Econ\\_letters.pdf](http://www.math.uvic.ca/faculty/reed/Econ_letters.pdf)

- [62] Reed, William J. (2002). “On Pareto’s Law and the determinants of Pareto exponents”. Disponible en: <http://www.math.uvic.ca/faculty/reed/JEI.pdf>
- [63] Reed, William J. “Power-law Distributions from Exponential Processes: An Explanation for the Occurrence of Long-tailed Distributions in Biology and Elsewhere”. Disponible en: [http://www.math.uvic.ca/faculty/reed/JAMS\\_sub](http://www.math.uvic.ca/faculty/reed/JAMS_sub)
- [64] Reichl, Linda E. (1980). *A Modern Course in Statistical Physics*, University of Texas Press.
- [65] Risken, Hannes (1989). *The Fokker-Planck Equation: Methods of Solution and Applications*, Springer, Berlín.
- [66] Ross, Sheldon (1998). *A first Course in Probability*, Prentice Hall, USA.
- [67] Ross, Sheldon (1983). *Stochastic Processes*, John Wiley & Sons, Inc., USA.
- [68] Ruelle, David (1993). *Azar y caos*, Madrid, Alianza Universidad.
- [69] Ruiz Baños, Rosario (2005). “Indicadores Bibliométricos”, Proyecto Cognosfera del Departamento de Biblioteconomía y Documentación, Facultad de Documentación, Universidad de Granada, España. También se consultaron los materiales (en diversos formatos): “Productividad de los autores científicos. Ley de Lotka”, “Bases para un concepto de las métricas”, “El crecimiento de la Ciencia. Las Leyes de Price”, todos disponibles en la *Sala de Estudio* del sitio: <http://www.ugr.es/rruizb/cognosfera/>
- [70] Sancho, J. M., San Miguel, M. (1980). *Ecuaciones diferenciales estocásticas: ruido blanco y de color*. Actas del II Congreso de Ecuaciones Diferenciales y Aplicaciones, Barcelona, 28-31 de mayo, 1979. Disponible en el sitio de la revista *Publicacions de la Secció de Matemàtiques*, Vol. 19: <http://www.raco.cat/index.php/PublicacionsSeccioMatematiques/>
- [71] Schroeder, Manfred (1991). *Fractals, Chaos, Power Laws: Minutes from an Infinite Paradise*, Freeman & Co.
- [72] Solé, Ricard V., Manrubia, S. C. (1996). *Orden y caos en sistemas complejos*, Universitat Politècnica de Catalunya, España.
- [73] Sornette, Didier (1996). “Convergent Multiplicative Process Repelled from Zero”. Disponible en: [arXiv:cond-mat/9609074](http://arxiv.org/abs/cond-mat/9609074).
- [74] Sornette, Didier (2006). *Critical Phenomena in Natural Sciences. Chaos, Fractals, Selforganization and Disorder: Concepts and Tools*. Second Edition. Springer Verlag, Germany.

- 
- [75] Stanley, H. Eugene (1971). *Introduction to Phase Transitions and Critical Phenomena*, Oxford University Press.
- [76] Stauffer, D., Stanley, H. E. (1996). *From Newton to Mandelbrot. A Primer in Theoretical Physics with Fractals for the Macintosh*, 2nd. Edition, Springer-Verlag.
- [77] Tague-Sutcliffe, Jean (1994). “Introducción a la informetría”, *Acimed, Revista Cubana de los Profesionales de la Información y la Comunicación en Salud*. Vol. 2, no. 3, septiembre-diciembre. <http://bvs.sld.cu/revistas/aci/indice.html>
- [78] Thompson, D’Arcy W. “Sobre la magnitud”. Reproducido en SIGMA. *El mundo de las matemáticas*, Vol 2., compilado por James R. Newman (1979), Grijalbo, España.
- [79] Torres Agudo, Joaquín J. “Ecuación maestra”. Disponible en: <http://www.ugr.es/~jtorres/leccion4.pdf>  
En su página personal del Instituto Carlos I de física teórica, España: <http://www.ugr.es/~jtorres/docencia.html>
- [80] Vallade, M., Houchmandzadeh, B. (2003). “Analytical solution of neutral model of biodiversity”, *Physical Review E*, 68, pp. 061902-1-061902-5.
- [81] Vogel, Eugenio E. (2000). “Usos correctos e incorrectos der los factores de impacto”, *Boletín de la Sociedad Mexicana de Física*, vol. 14, no. 1, enero-marzo, pp. 11-18.
- [82] Wilson, Kennet G. (1979). “Problems in Physics with many Scales of Length”, *Scientific American*, Vol. 241, pp. 140-55.
- [83] Zipf, George K. (1935). *The Psycho-biology of Language: An Introduction to Dynamic Philology*, MIT Press, USA.