

*Modelos, analogías y razón estadística, 1760-1900*¹

THEODORE M. PORTER

Departamento de Historia,
Universidad de California, Los Ángeles

En nuestro tiempo, el término estadística hace referencia a las herramientas matemáticas de inferencia y estimación por lo que es lógico que esta acepción ocupe un lugar central en los estudios de Historia. Stephen Stigler (1986) en su renombrado estudio sobre la historia de la estadística hasta 1900, emplea el subtítulo «la medida de la incertidumbre» pues su enfoque parte de considerar la teoría de los errores aparecida en la primera década del siglo XIX, como el centro de los tempranos desarrollos estadísticos. En relación con los conocimientos acerca de la probabilidad durante el siglo XVIII, los mínimos cuadrados se presentan como la síntesis de varias tradiciones que incluyen reglas y costumbres de la astronomía para combinar observaciones, el empleo de la probabilidad para estimar incertidumbres cuando se está frente a grandes cantidades de sucesos aleatorios, así como los métodos inversos de probabilidad basados en los estudios de Thomas Bayes y Pierre Simon de Laplace. El método de los mínimos cuadrados fue desarrollado matemáticamente alrededor de 1810 por Laplace y Karl Friedrich Gauss, y terminaría por convertirse en el principal modelo a la hora de analizar datos de diferentes campos utilizando la matemática de las probabilidades. No obstante, como ha observado Stigler, el método de los mínimos cuadrados se utilizó poco en el siglo XIX. Otras ciencias de la observación como la meteorología, llegaron a usarlo muy tempranamente y de hecho, las expediciones geodésicas dirigidas a la medición de las distancias entre los meridianos para poder fijar el metro, facilitaron el contexto para la formulación de los mínimos cuadrados hecha por Adrian María Legendre² en 1805 (Stigler 1999, cap. 20).

¹ Traducción y notas: José M. Arribas. Anexo: Antonio Vallejos.

² Ver anexo.

En las ciencias sociales, a pesar del entusiasmo de hombres como Adolfo Quetelet, las medidas probabilísticas de la variación y la incertidumbre nunca se desarrollaron con éxito hasta después de 1900. El campo experimental de la psicofísica liderada por Gustav Theodor Fechner³ fue tan solo una excepción parcial, pues la auténtica ruptura tuvo que esperar hasta la escuela biométrica de Francis Galton, Karl Pearson y G. Udny Yule.

Una visión alternativa de la historia de la probabilidad y la estadística puede hallarse en los dos volúmenes producidos por el grupo de investigación de Bielefeld en Alemania con el título «The Probabilistic Revolution» (Krüger, Daston y Heidelberger, 1987, Krüger, Gigerenzer and Morgan, 1987) La problemática de este trabajo fue definida, no tanto para cambiar los métodos del análisis cuantitativo (*data analysis*), como para mostrar una nueva vía de comprensión del mundo. Y aunque el informe comienza alrededor de 1800, los redactores de estos volúmenes escriben bastante poco sobre el método de los mínimos cuadrados. Con todo, lo abordan como parte del acervo filosófico contra el que se definieron las modernas concepciones probabilísticas. El método de los mínimos cuadrados fue uno de las más importantes logros de la probabilidad clásica, o al menos el que tuvo más éxito, pero fue parte del viejo mundo en el cual la probabilidad era ante todo, algo concerniente a los errores y la incertidumbre, pero no relativo a las variaciones y al mundo del azar. Tal vez, la aportación más importante de la revolución probabilística para este grupo es lo que Ian Hacking ha llamado la «erosión del determinismo» (Hacking, 1990). En este enfoque, los ingleses dedicados a la biometría ocuparon la parte central, aunque tal vez, más por su incorporación de la variabilidad en la estructura del mundo que por sus métodos de manipulación del error y la incertidumbre. Un papel similar es asignado a algunos desarrollos que son virtualmente omitidos en el trabajo de Stigler, como la emergencia en física de la mecánica estadística y la teoría cuántica. Aquí la visión determinista del mundo, se viene gradualmente abajo y es reemplazada por otra en la que el azar juega un papel fundamental.

Por supuesto, la historia de la probabilidad y la estadística puede ser abordada desde diferentes ángulos, pero aquí contrastaremos dos amplias perspectivas: una que examina la estadística como herramienta de razonamiento o de inferencia, y otra que desarrolla analogías estadísticas, modelos y teorías acerca del mundo natural y social. Sin que ninguna de estas dos vías devalúe la importancia de la historia de la estimación y la inferencia, me gustaría abordar aquí el significado de los modelos y las analogías para la historia de la razón estadística.

LA MODELIZACIÓN MATEMÁTICA: DANIEL BERNOULLI Y LA VIRUELA

El lenguaje de la modelización matemática parece haberse desarrollado tan solo en el siglo XX, aunque, siendo prudente con los anacronismos, podemos decir

³ Nacido en 1801 y fallecido en 1887 en Leipzig. Fundador de la psicofísica entendida como ciencia de las relaciones cuantitativas entre las sensaciones y los estímulos.

que en algún sentido, los modelos han sido usados en la ciencia desde tiempos muy antiguos. Alistair Crombie (1994) aborda la modelización como uno de los estilos básicos del razonamiento científico y proporciona abundantes ejemplos procedentes de los períodos medieval y moderno. Su principal objetivo era visualizar una imagen macroscópica de aquello que no puede ser observado. Desde esta perspectiva, la modelización alcanzó la cima de su influencia bajo la «filosofía mecánica» del siglo XVII en la que broches que se acoplan explican los enlaces químicos, los muelles dan saltos en el aire, y partículas puntiagudas sirven para experimentar la acidez. Semejantes ayudas para la imaginación, eran expuestas finalmente como teorías, pero con la idea de que incluso si la naturaleza contiene algo parecido a estos curiosos objetos, sus mecanismos no son exactamente los mismos de los muelles y palancas de la experiencia común. En otros casos, una imagen física precisa podía ser razonada, como en el análisis matemático que hace Descartes de la refracción de la luz en las gotas de agua para explicar el arco iris.

A finales del siglo XVIII, no era raro, en cambio, utilizar la probabilidad matemática para comprender y proponer reformas de las instituciones humanas. Uno de los casos más conocidos, sirvió de base para teorías sobre testimonios, elecciones y procesos judiciales. Trabajar de este modo era, como Lorraine Daston (1988) ha puesto de manifiesto, típico de la teoría de la probabilidad clásica, la cual fue concebida en los comienzos del siglo XIX como buen sentido matematizado. En algunos casos, la probabilidad iba más allá de la capacidad del sentido común facilitando consejo en cuestiones presentadas como puramente cuantitativas. De los meros juicios no podían deducirse las complejas consecuencias de disponer de un jurado más o menos grande, alterando así la mayoría necesaria para la condena (Baker, 1975; Brian 1994) y los cálculos probabilísticos podían facilitar estimaciones de cantidades que una inteligencia no adiestrada jamás podría reducir a números.

Un ejemplo particularmente notable de modelización matemática se encuentra en la memoria de Daniel Bernouilli cuando en 1760 explora los efectos de la mortalidad producida por la viruela, así como las potenciales ganancias de su eliminación. En términos más políticos, Bernouilli deseaba demostrar la racionalidad de la inoculación que desde Voltaire era una de las campañas favoritas de la Ilustración. En este empeño, Bernouilli usa la tabla de mortalidad de Edmond Halley, ligeramente modificada, como tasa de mortalidad en el primer año de vida. Dado que no había datos comparativos sobre la incidencia por edad de la viruela ni de la mortalidad, recurre a simplificar supuestos: los inequívocos e hipotéticos elementos a los que llamamos su modelo. Suponía que cada año, un octavo de los que no habían sufrido todavía viruela, la contraerían, y un octavo de ellos, morirían. Escribe y resuelve una ecuación diferencial para determinar cuantos en cada edad permanecen vulnerables, en el «estado natural» en el que la viruela está presente, y cuantos de ellos caerán enfermos y morirán. Esto le permite calcular una tabla que distingue anualmente las muertes debidas a la viruela de aquellas otras atribuidas a otras causas y finalmente calcula la tabla de mortalidad que podría preverse si se eliminase la viruela.

Bernouilli insistió en la imparcialidad de su posición. No hizo una defensa a ultranza de la inoculación, sino más bien quiso explorar sus consecuencias, en

una línea que muestra claramente su punto de vista acerca de la utilidad de la inoculación: no más de uno sobre 200 moriría como resultado de la inoculación quirúrgica, mientras que los niños o los jóvenes podrían prolongar sus expectativas de vida hasta más de tres años. En relación al Estado el argumento es más poderoso; muchas muertes de viruela, de acuerdo con sus cálculos, tenían lugar en la infancia, y aquellos que mueren antes de los 16 años consumen recursos sin contribuir significativamente a la riqueza de la nación. Si se eleva la proporción adultos/ niños, la inoculación contribuye al incremento de la prosperidad y el poder.

La memoria de Bernouilli fue una notable pieza de razonamiento matemático, pero no fue el inicio de la tradición; en realidad, la impugnación de este tipo de razonamiento aparece antes de 1766 cuando lo publica la Real Academia de Ciencias de París en las «Memoires de Mathématique et de Physique». La crítica vino de Jean d' Alembert, rival de Bernouilli y famoso escéptico de la teoría de la probabilidad, y objetó tales planteamientos en diferentes niveles, incluyendo el psicológico. El razonamiento de Bernouilli, aunque fuera encorsetado en términos de matemáticas y evidencia, pudo no parecer aún decisivo, especialmente a una madre que muy bien podía rehusar comprar una mayor esperanza de vida al precio de incrementar la posibilidad de pérdida de su niño en el siguiente mes. El Estado podía así confiar sin ambigüedad en la estadística y las medias, pero para un individuo siempre existía la posibilidad de que su vida particular se acortara (para ver las diferentes perspectivas del Estado y del individuo: Desrosiers, 1993). Para nuestro propósito, es particularmente importante destacar que d' Alembert rechazó los aspectos modelizadores de Bernouilli. D' Alembert quería datos, no suposiciones (d' Alembert, 1760):

«En primer lugar, la suposición que los ilustres matemáticos hacen sobre el número de personas que en cada edad cogen la viruela y sobre el número de los que mueren o no como consecuencia, parece estar absolutamente infundado. No es, de ningún modo cierto —o como mucho es bastante dudoso— que la viruela, a cualquier edad, ataque constantemente a una octava parte de los sanos; y es todavía más dudoso que, a cualquier edad, constantemente destruya un octavo de aquellos a los que ataca. Además, sería necesario conocer si la inoculación siempre supone, como es asumido, la misma fracción de individuos, independientemente de la edad a la que fueron inoculados».

Bernouilli introduce de hecho supuestos de modelización como substitutos de los datos que no pueden ser obtenidos, y donde es posible, compara las predicciones de la teoría con los datos disponibles. D' Alembert, por el contrario, rechazó el conjunto del planteamiento sin esfuerzo, alguno para explorar cómo las conclusiones podían cambiar si se alteraran las suposiciones simplificadoras. Y esto sucedía en el siglo XVIII, en pleno apogeo de la probabilidad clásica. El siglo XIX fue mucho menos tolerante con el uso de conjeturas probabilísticas como substituto de la información empírica y también menos próximo a las versiones matemáticas del razonamiento humano.

CIENCIA SOCIAL ESTADÍSTICA Y PENSAMIENTO ESTADÍSTICO

Un único modelo soporta el razonamiento estadístico del siglo XIX y no es, en principio, matemático, sino más bien social. El cambio se produjo como consecuencia de la dedicación a precisar medidas que después de 1750 progresivamente ha caracterizado a las ciencias (Wise, 1995). No fue, sin embargo, una mera consecuencia de la investigación científica, sino de la expansión burocrática y de la producción y circulación de información social. Fueron las publicaciones oficiales sobre estadística criminal en Francia las que primero provocaron perplejidad en Quetelet y en alguno de sus contemporáneos, pues crímenes de cualquier condición –asesinatos, suicidios, robos, etc– se sucedían con valores anuales constantes. Quetelet se preguntaba en 1829, en su primera publicación dedicada a este fenómeno, si los actos morales del hombre eran dirigidos por alguna extraña fuerza de la fatalidad que dejaba solo los pequeños actos de la vida cotidiana a la libertad individual (Porter, 1986) Las regularidades de la «estadística moral» dejaron un considerable fermento intelectual en las décadas siguientes. Fueron difundidas ampliamente en la «Historia de la civilización en Inglaterra» de Henry Thomas Bucle, publicada en 1857, y su influencia fue particularmente importante en Alemania, Rusia y Gran Bretaña.

A lo largo de este período, hubo más incertidumbre y desacuerdos sobre lo que significaban exactamente las regularidades estadísticas. El mismo Buckle las asociaba con necesidades científicas, cuando no con extrañas fatalidades. Quetelet tomó finalmente una línea menos rigurosa, insistiendo en que las regularidades en el nivel de las masas todavía dejan un modesto espacio para la libertad en cualquier individuo. Asumiendo, no obstante, que las estadísticas producían escasa o ninguna ayuda en la predicción de acciones individuales, estaba lejos de quedar claro lo que describen en el tiempo presente. Quetelet resolvió este problema a favor de la realidad de lo social. Las «leyes sociales» eran características no de cualquier individuo particular, sino de la especie humana en masa, de ese cuerpo especial que llamamos «sociedad», o más precisamente, caracterizó no al hombre individual sino al hombre medio.

La idea de sociedad, una entidad progresiva cuyo desarrollo podría no ser al final controlada por las autoridades políticas, fue realmente importante hacia mediados del siglo en Europa, y proporcionó fundamentos al discurso del liberalismo desde el momento en que reducía el papel potencial del Estado. Al mismo tiempo, el estudio de la enfermedad y el crimen en términos de grandes regularidades, más que una consecuencia de la debilidad de los individuos, fue visto por defensores y detractores como si debilitara o difuminara la responsabilidad moral. Mientras un criminal podía ser justamente condenado, la «física social» ponía el acento, por el contrario, en el fenómeno colectivo, en las tasas de criminalidad que pueden ser comparadas con otros tiempos y lugares para aprender algo sobre las causas sociales que hay detrás. Esto era la implicación más importante de lo que estoy denominando el modelo estadístico. Grandes números de individuos, actuando independientemente en un sistema, producen regularidades que no dependen de la coordinación o dirección de todas esas personas. Es posible razonar sobre colectividades sin ningún conocimiento detallado de los individuos.

Esta concepción liberal de la sociedad y de la producción espontánea del orden social, facilitó también un modelo para una forma de razonamiento matemático, cuyo desarrollo posterior fue enormemente productivo. Es así razonable que una palabra destinada a la investigación administrativa y social como «Estadística», haya llegado a convertirse en la denominación de un campo de la matemática aplicada. Después de todo, fue en el campo social donde el estudio cuantitativo de los fenómenos de masas se puso en práctica en primer lugar, y donde sus supuestos básicos comenzaron a utilizarse. El campo matemático de la estadística comienza alrededor de 1900, tomando su formalismo de la probabilidad matemática, y especialmente de los desarrollos hechos con el método de los mínimos cuadrados durante el siglo XIX. Bajo todo ello, subyace, no obstante, la concepción de cómo se produce el orden colectivo, y un esquema para detectar en el nivel superior las causas que permanecen menos claras en el nivel individual.

LA FÍSICA SOCIAL

Quetelet se dio cuenta de que las estadísticas podían facilitar el material necesario al Newton de las ciencias sociales que habrá de desarrollar la «mecánica social» o «física social» bajo la influencia de la mecánica celeste de Laplace. La cobertura de una física social era esperada desde hacía tiempo en los postreros años del siglo XIX aunque curiosamente su desarrollo se produjo más ampliamente en las ciencias físicas. Aquí es donde Ludwig Boltzmann introdujo su conocida memoria de 1872 sobre la convergencia de la distribución de velocidades de las moléculas del gas con la distribución exponencial, más tarde bautizada por él y por James Clerk Maxwell:

«Se atribuye exclusivamente a las circunstancias que incluso el proceso más irregular produzca los mismos valores medios cuando se le coloca bajo las mismas condiciones, es por ello que percibimos, por ejemplo, el comportamiento de los objetos calientes como completamente determinados por leyes. Las moléculas de un cuerpo son tantas y sus movimientos son tan rápidos que nada llega a ser perceptible para nosotros, excepto los valores medios.

La regularidad de estos valores puede ser comparada con la sorprendente constancia de los números medios facilitados por la estadística, que se derivan así mismo de procesos en los que cada suceso individual está condicionado por la participación de las más diversas e incalculables circunstancias externas. Las moléculas son como muchos individuos pues teniendo los más variados estados de movimiento y las propiedades de los gases, solo permanecen inalterados porque el número de estas moléculas, lo cual se manifiesta en la media, tienen un estado de movimiento constante. La determinación de las medias es la tarea del cálculo de probabilidades».

O como escribe en 1886 en una presentación divulgativa del mismo tema: «Como es bien conocido, Buckle demostró con la estadística que sólo si un

número suficiente de personas es tenido en cuenta en un recuento, no sólo tendremos el número de acontecimientos naturales como muertes, enfermedades, etc, perfectamente constante, sino también el número de lo que llamamos acciones voluntarias: matrimonios a determinada edad, crímenes, suicidios. Todo ello sucede en una forma similar al movimiento de las moléculas» (Porter 1986, 112-114) Maxwell, comienza a usar esta analogía en el mismo período. En una comunicación dirigida a la Asociación Británica para el Avance de la Ciencia en 1873, explica que las moléculas de un gas son muy numerosas y sus interacciones demasiado complejas para resolverlas utilizando las ecuaciones de la dinámica.

«Los modernos atomistas han adoptado, de todos modos, un método que es, según creo, nuevo en el departamento de física matemática, aunque ha estado mucho tiempo utilizándose en la sección de Estadística. Cuando los miembros de la sección F tienen un informe sobre el censo, comienzan distribuyendo el total de la población en grupos, de acuerdo con la edad, contribución fiscal, educación, pensamiento religioso o convicciones penales. El número de individuos esta lejos de permitir el seguimiento histórico de cada uno por separado por lo que en orden a reducir el trabajo dentro de unos límites humanos se concentra la atención en pequeños grupos artificiales: si varían el número de individuos en cada grupo y no varía el estado de cada individuo, es el primer dato a partir del cual trabajan...

Las ecuaciones de la dinámica expresan completamente las leyes del método histórico aplicadas a la materia, pero la aplicación de estas ecuaciones implica un perfecto conocimiento de los datos. Sin embargo, las porciones más pequeñas de materia que podemos someter a experimento están formadas por millones de moléculas y ninguna de ellas llega a ser individualmente perceptible a nuestros ojos. No podemos verificar, no obstante, el movimiento actual de alguna de esas moléculas por lo que estamos obligados a abandonar el método estrictamente histórico y adoptar el método estadístico del tratamiento de grandes grupos de moléculas» (Maxwell, 1873).

La combinación de complejidad y desorden en el nivel de individuos, con el orden estadístico en el nivel de las poblaciones, es análogo en los terrenos físico, y social. El modelo social proporciona una vía para visualizar un sistema de infinitas moléculas que colisionan como si estuvieran ordenadas y fueran capaces de soportar leyes como regularidades, que ya han sido formuladas en los términos de la termodinámica. La analogía proporciona un método atractivo de exposición para introducir el tema en audiencias de un bajo nivel educativo. Proporciona también una analogía para explorar las limitaciones del método estadístico en la teoría cinética de gases. Maxwell que estaba en desacuerdo con las visiones deterministas acerca del futuro humano, siguió la línea crítica que desarrollaron los estadísticos sociales, cuando puntualiza que las regularidades de grandes masas moleculares no están atadas a los individuos. Incluso las relaciones macroscópicas de temperatura y presión no tienen el mismo estatus que las leyes de la dinámica, pero reducen a fondo las regularidades estadísticas.

MODELOS SOCIALES Y PROCESOS BIOLÓGICOS

En física como en ciencia social, el pensamiento estadístico tuvo algunas consecuencias paradójicas. Surgió con el descubrimiento de una nueva forma de orden científico que a pesar de todo, llevaba aparejado el reconocimiento de la incertidumbre. En la biología evolucionista, la complejidad del modelo social tomó una forma diferente. El problema de la homogeneidad, -central para todo el siglo, como observa Stigler - fue mucho más un asunto de los estudios biométricos sobre la herencia biológica». La biometría, de todos modos, no lo resuelve, y el modelo social de Francis Galton, por ejemplo, permitió diferentes alternativas comprensivas de los mecanismos de la herencia biológica, así como los marcos alternativos a la estadística.

Charles Darwin, primo de Galton, había ofrecido su «hipótesis provisional de la Pangenesis» en 1868 al proporcionar una teoría de la herencia biológica que podía ser el soporte de su teoría de la evolución. Implicaba la existencia de pequeños elementos genéticos a los que llama «gémulas» procedentes de cada unidad -tal vez de cada célula- del cuerpo. La actividad sexual destinada a la reproducción concentra suficientes «gémulas» para permitir la formación de un embrión. Galton elaboró y modificó la hipótesis de Darwin a través de un símil que publicó en su «Genios hereditarios» de 1869. Su versión, en contraste con Darwin, excluye el uso de la herencia, más bien, permite la selección de «gémulas» por afinidades, y no solo a través de la selección aleatoria.

El modelo de Galton sobre la organización espontánea procede de un modelo que refleja la economía política liberal. Propuso que la formación de «gémulas» en un individuo, «puede ser comparada con el típico aspecto que siempre se encuentra en diferentes descripciones de conjuntos ... la reunión de los cuales es como si no estuviesen controlados por autoridad central alguna, pero han asumido su apariencia típica a través de la libre acción de los individuos que los componen, estando cada hombre inclinado hacia sus propios intereses inmediatos y encontrando su lugar bajo la única influencia o afinidad electiva de sus vecinos» (Galton, 1869, 364) La invocación de afinidades electivas es aquí crucial. La formación de una ciudad no se hace mediante una muestra aleatoria de los alrededores de la región sino mediante una selección muy especial. En algunos casos como el de una población destinada a la pesca o un centro turístico, los elementos combinan fácilmente, siempre que la costa tenga un paisaje suficientemente atractivo y los visitantes paguen contentos por el día elegido. Una ruidosa y contaminante factoría, por el contrario, puede dirigir el turismo hacia otro lado. En el mismo sentido, los elementos genéticos de padres heterogéneos pueden no combinarse en los herederos genéticos y sin embargo, un tipo particular puede prevalecer.

El modelo estadístico de la herencia biológica depende de un surtido de genes que es congruente, al menos, con lo que sucede bajo un sistema de selección aleatoria. Galton usó la analogía de los dibujos procedentes de una urna para caracterizar este proceso. Cuando comenzó sus estudios experimentales y estadísticos de la herencia, anticipó que cualquier rasgo de los vástagos (después de ser corregidas las diferencias de sexo) debería ser distribuido alrededor de la media de los padres. Descubrió que tendían más a «revertir» en proporciones

fijas hacia la media para el conjunto de la población. El término «reversión» tiene un significado específico en biología y su uso sugiere que esa tendencia a retornar sobre la media se debe a la presencia de genes procedentes de los ancestros más remotos que no han sido puestos de manifiesto en los padres. Un surtido de elementos genéticos procedentes de varias generaciones representa necesariamente una aproximación cercana a la media para la población que hizo que esto se expresara en forma individual.

A mediados de 1880, cambió el término «reversión» por el de «regresión». Mas tarde el término se referirá también al movimiento hacia atrás en el tiempo, pero Galton indicó un significado diferente. En 1885, durante su presidencia de la sección de antropología de la Asociación Británica, explicó que «El tipo es una forma ideal hacia la que los niños de quienes se desvían tienden a regresar... La estabilidad de un tipo puede, según creo, ser medida por la fuerza de su tendencia a regresar» (Galton, 1885,509-10) Es decir, Galton no consideraba la regresión como un movimiento hacia una forma ancestral que actúa como tipo ideal, más bien ponía el acento en la estabilidad pues piensa que la «herencia efectiva del niño» es «menor que las desviaciones acumuladas de sus antecesores» (Galton 1886,62). En estas cuestiones nunca dejó de tener en cuenta lo que Karl Pearson quiso poner en circulación como visión de la biometría: que el cambio evolutivo es siempre continuo. Galton y Pearson compartieron una inquebrantable dedicación a la estadística, pero la visión de Galton sobre la herencia biológica y la formación de los individuos requiere discontinuidades en el proceso de evolución. En la medida en que la evolución tenga lugar en los confines del genotipo, las desviaciones tenderán a retornar en las siguientes generaciones a un punto de estabilidad: el genotipo de la población. Los verdaderos cambios evolutivos requieren prácticas continuadas, cambios radicales que formen un nuevo punto de estabilidad, un nuevo tipo hereditario.

Mientras que, «reversión» y «regresión» se refieren a las relaciones entre generaciones, «correlación» significa para Galton interrelación entre rasgos de individuos particulares. En su papel de criminólogo buscaba una vía para identificar criminales –como el que más tarde se usaría utilizando las huellas de los dedos– y lo definió como un problema de correlación. En qué proporción una altura inusual implica una longitud particular del brazo o de los dedos. Siguiendo la hipótesis de la Pangenesis fue una cuestión de afinidades electivas, algo como que las «gémulas» están asociadas con la longitud del brazo, debe de estar también relacionado con las que determinen la longitud de las piernas y la medida del torso. Curiosamente, el hecho de pensar la relación como reversión matemática, de modo que una altura inusual implica un largo cúbito en cada individuo, implica, a su vez una longitud superior a la normal para él y su progeñie. La correlación de las partes no era, aún, fácilmente reconciliable con el modelo de los dibujos de bolas independientes que salen de una urna y que esta presente en las matemáticas de la «reversión» y la «regresión» La correlación estaba unida más a la lógica de afinidades electivas o inestabilidades y de diferentes tipos hereditarios. Si hubiera sido modelado en términos de dibujos procedentes de una urna, entonces habría sido prejuizado y no aleatorio.

La falta de coherencia en los modelos de Galton ha sido ligada, no obstante, a su escasa formación como matemático y a su rechazo a ser encorsetado por las

estrictas condiciones de la derivación matemática, lo cual contribuyo forzosamente a su orientación estadística. Como señala Stigler (1986,274), Galton se propone suavizar los restrictivos supuestos que ha habido en la derivación hecha por Laplace de la curva normal de errores. Correlación y regresión significa dividir la variación en una fracción que puede ser atribuida a variables que correlacionan y a otra parte aleatoria. Esta división no se deriva lógicamente de su modelo, sino que emerge de sus propias ambigüedades cuando se sitúa en el contexto de sus trabajos con semillas de guisantes, archivos criminales y medidas biológicas. Galton desplegó modelos de muestras aleatorias en algunas ocasiones e invocó la correlación de determinadas partes aunque en otras, prestó más atención al aprovechamiento práctico que a la consistencia lógica.

CONCLUSIÓN

Galton fue muy crítico con los estadísticos que veían la variación únicamente como error. Aunque el objeto de su desprecio no fueron los matemáticos sino los funcionarios de los censos y similares que raramente fueron más allá de presentar los valores medios, la tradición matemática de la teoría del error era todavía limitada en su objeto y perspectivas. En este sentido he señalado algunos momentos sobresalientes en un tradición alternativa que puede ser caracterizada en los términos de sus modelos. Con ocasión de la teoría cinética de gases, el modelo estadístico recurrió a matemáticas muy sofisticadas, pero la imagen de regularidad y estabilidad social que estaba en todo momento presente era elemental desde el punto de vista matemático. Las estadísticas del siglo XIX suponen en cierto sentido la retirada de los presupuestos utilizados por Bernouilli. Los estadísticos sociales prefirieron no ir más allá de los datos empíricos, la estadística aún no era un conjunto de herramientas para analizar datos sino que facilitaba los rudimentos de una teoría de la sociedad y luego a funcionar también como modelo en el sentido de que su estructura conceptual básica podía ser trasladada a otros campos como la física de gases y la evolución biológica donde, por cierto, desarrolló nuevas vías. Las teorías del error y la medida de la incertidumbre son solo parte de la historia. Los desplazamientos de unas disciplinas a otras y la elaboración de modelos estadísticos característicos forman un elemento clave en la historia de la estadística matemática durante el siglo XIX.

BIBLIOGRAFÍA

- ALEMBERT, JEAN LE ROND D' (1760): «On the Application of the Calculus of Probabilities to Inoculation against Smallpox», in T. Bradley, ed., *Smallpox Inoculation: An Eighteenth-Century Mathematical Controversy*, Adult Education Department of the University of Nottingham, Nottingham, 1971.
- BAKER, KEITH (1975): *Condorcet: From Natural Philosophy to Social Mathematics*, University of Chicago Press, Chicago.

- BERNOULLI, Daniel (1766): «An Attempt at a New Analysis of the Mortality Caused by Smallpox,» in T. Bradley, ed. *Smallpox Inoculation: An Eighteenth-Century Mathematical Controversy*, Adult Education Department of the University of Nottingham, Nottingham, 1971.
- BRIAN, Eric (1994): *La mesure de l'Etat: Administrateurs et géomètres au XVIIIe siècle*, Albin Michel, París.
- CROMBIE, Alistair (1994): *Styles of Scientific Thinking in the European Tradition*, 3 vols., Duckworth, London.
- DASTON, Lorraine (1988): *Classical Probability in the Enlightenment*, Princeton University Press, Princeton.
- DESROSIÈRES, Alain (1993): *La Politique des grands nombres: Histoire de la raison statistique*, Editions la Découverte, París.
- GALTON, Francis (1869): *Hereditary Genius*, Macmillan, London.
 — (1885): «Address to the Section of Anthropology of the British Association,» *Nature*, 32, 507-510.
 — (1886): «Family Likeness in Stature,» *Proceedings of the Royal Society of London*, 40, 42-73.
- HACKING, Ian (1990): *The Taming of Chance*, Cambridge University Press, Cambridge, U.K.
- KRÜGER, LORENZ, GERD GIGERENZER and MARY MORGAN, eds. (1987): *The Probabilistic Revolution, volume 2: Ideas in the Sciences*, MIT Press, Cambridge, Mass.
- KRÜGER, LORENZ, LORRAINE DASTON and MICHAEL HEIDELBERGER, eds. (1987): *The Probabilistic Revolution, volume 1: Ideas in History*, MIT Press, Cambridge, Mass.
- MAXWELL, James Clerk (1873): «Molecules,» in W. D. Niven, ed., *Scientific Papers of James Clerk Maxwell*, 2 vols., Cambridge University Press, Cambridge, U.K., 1890.
- PORTER, Theodore M. (1986): *The Rise of Statistical Thinking, 1820-1900*, Princeton University Press, Princeton.
- STIGLER, Stephen M. (1986): *The History of Statistics: The Measurement of Uncertainty before 1900*, Harvard University Press, Cambridge, Mass.
 — (1999). *Statistics on the Table: The History of Statistical Concepts and Methods*, Harvard University Press, Cambridge, Mass.
- WISE, M. Norton, ed. (1995): *The Values of Precision*, Princeton University Press, Princeton.

ANEXO

Legendre: *Nuevos métodos para la determinación de las órbitas de los cometas* París, 1805. (Jean-Luc Chabert y otros «Histoire d' algorithmes» Du caillou à la puce, BELIN)

«En la mayoría de las investigaciones que intentan de obtener las medidas dadas por la observación, los resultados más exactos que nos pueden ofrecer, se llega a un sistema de ecuaciones de la forma:

$$E = a + bx + cy + fz + \text{etc.}$$

en el cual a, b, c, f , etc. son coeficientes conocidos que varían de una ecuación a otra, y x, y, z , etc. son incógnitas que tenemos que determinar con la condición de que el valor de E se reduzca, en cada ecuación, a una cantidad nula o muy pequeña. Si hay tantas ecuaciones como incógnitas x, y, z , etc., no existe ninguna dificultad para la determinación de estas incógnitas, y se pueden encontrar errores E absolutamente nulos. Pero la mayoría de las veces el

número de ecuaciones es superior al de incógnitas y se hace imposible eliminar todos los errores.

En esta circunstancia, que es la de la mayor parte de los problemas físicos y astronómicos, cuando se busca determinar algunos elementos importantes, entra necesariamente lo arbitrario en la distribución de los errores, y no se puede decir que todas las hipótesis conducirán exactamente a los mismos resultados; pero hay que procurar que los errores extremos, sin tener que atender a sus signos, sean reducidos en los límites más estrechos posibles.

De todos los principios que se pueden proponer para este fin, yo pienso que no hay ninguno más general, de mayor exactitud, ni de una aplicación más fácil que el que hemos usado en las investigaciones precedentes, y que consiste en convertir en *mínima* la suma de los cuadrados de los errores. Por este medio se establece entre los errores una suerte de equilibrio que, haciendo desaparecer los valores extremos, es muy apropiado para conocer el estado del sistema más próximo a la verdad.

La suma de los cuadrados de los errores $E^2 + E'^2 + E''^2 + \text{etc.}$ eran

$$\begin{aligned} & (a + bx + cy + fz + \text{etc.})^2 \\ & + (a' + b'x + c'y + f'z + \text{etc.})^2 \\ & + (a'' + b''x + c''y + f''z + \text{etc.})^2 \\ & + \text{etc. ;} \end{aligned}$$

si lo que se busca son *mínimos* haciendo variar solamente la x , se tendrá la ecuación

$$0 = \int ab + x \int b^2 + y \int bc + z \int bf + \text{etc.}$$

en donde por $\int ab$ se entiende la suma de los productos semejantes $ab + a'b' + a''b'' + \text{etc.}$, por $\int b^2$ la suma de los cuadrados de los coeficientes de x , a saber, $b^2 + b'^2 + b''^2 + \text{etc.}$, y así todos los siguientes.

El *mínimo* por relación a y será, de manera similar:

$$0 = \int ac + x \int bc + y \int c^2 + z \int cf + \text{etc.}$$

y el *mínimo* por relación a z :

$$0 = \int af + x \int bf + y \int cf + z \int f^2 + \text{etc.},$$

donde se ve que los mismos coeficientes $\int bc$, $\int bf$, etc. son comunes a las dos ecuaciones, lo que contribuye a facilitar el cálculo.

En general, *para obtener la ecuación del mínimo por relación a una de las incógnitas, hay que multiplicar todos los términos de cada ecuación propuesta por el coeficiente de la incógnita en esta ecuación, tomado con su signo, y realizar la suma de todos estos productos.*

Se obtendrá de este modo tantas ecuaciones del *mínimo* como incógnitas haya, y se resolverán por los métodos ordinarios.

[...]

La regla por la cual se toma el mejor de entre los resultados de diversas observaciones (para un solo elemento), no es más que una simple consecuencia de nuestro método general, que llamaremos *método de los mínimos cuadrados*. De hecho, si en la experiencia hemos encontrado diversos valores a' , a'' , a''' , etc. para una cierta cantidad x , la suma de los cuadrados de los errores será $(a' - x)^2 + (a'' - x)^2 + (a''' - x)^2 + \text{etc.}$; e igualando esta suma a un mínimo, se tiene

$$0 = (a' - x) + (a'' - x) + (a''' - x) + \text{etc.},$$

de donde resulta que

$$x = \frac{a + a' + a'' + \dots + \text{etc.}}{n},$$

siendo n el número de observaciones.»

RESUMEN

En el presente artículo se aborda la noción de modelo en la estadística del siglo XIX. Mientras la estadística se convierte durante el siglo veinte en un conjunto de métodos de inferencia, durante el siglo XIX fue sobre todo, un conjunto de modelos que hundían sus raíces en una determinada comprensión de la sociedad y en las aplicaciones de otros campos como la física de gases o la evolución biológica.

ABSTRACT

In this article I try to show the idea of model in the statistics during the 19th century. Whereas statistics has become first of all a set of methods of inference in the 20th century, in the 19th it was first of all a set of models, rooted in an understanding of society and applied to other domains like gas physics and biological evolution.