EL RECONOCIMIENTO DEL MÉRITO CIENTÍFICO A TRAVÉS DE LA HISTORIA DE ALGUNOS DE LOS EPÓNIMOS ESTADÍSTICOS MÁS RELEVANTES. ¿SE VERIFICA LA LEY DE STIGLER?

José A. Roldán-Casas
Universidad de Córdoba

Resumen

La eponimia, proceso de generación de expresiones -epónimos- que tratan de asociar un hallazgo con su descubridor, es el sistema de reconocimiento del mérito científico más prestigioso y duradero en el tiempo. No obstante, en ocasiones, el epónimo establecido no hace honor al descubridor original del hallazgo en cuestión. Esto es lo que estipula la *Ley de Stigler*, que ningún descubrimiento científico recibe el nombre de su descubridor original, y cuyo cumplimiento se comprueba en algunos de los epónimos estadísticos más relevantes analizando la historia del proceso de formación de cada uno de ellos. El análisis revela que, de los seis casos considerados, tres evidencian la *Ley de Stigler*. En concreto, se constata que De Moivre fue el primero en descubrir tanto la *distribución normal* como la *distribución de Poisson*, siendo Bortkiewicz el primer autor en detectar la aplicación de la *Poisson* como distribución de los sucesos raros, mientras que la *distribución de Cauchy* fue descubierta por Poisson antes de que Cauchy hiciera referencia a ella. Por otro lado, hay dos casos (*distribución F de Snedecor* y *Teorema de Bayes*) en los que el análisis suscita dudas o controversias que no permiten pronunciarse sobre el cumplimiento o no de la *Ley*. Finalmente, en el caso de la *distribución t de Student*, desde un principio se supo que el pseudónimo *Student* hacía referencia a su descubridor, Gosset.

Abstract

Eponymy, the generating process of expressions -eponyms-, that try to associate a finding with its discoverer, is the most prestigious and lasting system for the recognition of scientific merit. However, sometimes the established eponymous does not honour the original discoverer. It is postulated by Stigler's Law, that is, "*no scientific discovery is named after its original discoverer*". Its fulfilment is studied in some of the most relevant statistical eponyms analysing the history of the generating process of each one of them. The study reveals that three out of the six cases considered, verify the Stigler´s Law.

Recibido el 7 de enero de 2021 — Aceptado el 1 de febrero de 2021
https://doi.org/10.47101/llull.2021.44.89.roldan

Specifically, it is confirmed that De Moivre was the first discoverer of the *normal distribution* and *Poisson distribution*, being Bortkiewicz the first author to associate the *Poisson* with rare events, whereas the *Cauchy distribution* was discovered by Poisson before Cauchy referred to it. On the other hand, in two cases (*Snedecor F distribution and Bayes´ Theorem*) the analysis raises doubts that do not allow to conclude about the fulfilment of the Law. Finally, in the case of the *Student's t distribution* it was well-known from the beginning that the pseudonym *Student* referred to the original discoverer, Gosset.

Palabras claves: Eponimia, Ley de Stigler, Estadística, Distribuciones de probabilidad.

Key words: Eponymy, Stigler's Law, Statistics, Probability distributions

1. Introducción

El logro del reconocimiento del mérito científico por el trabajo realizado siempre ha sido un asunto que ha estado muy presente en la comunidad científica. En ese sentido, si las aportaciones a la ciencia son relevantes, es comprensible que sus descubridores deseen pasar a la posteridad, ya sea por satisfacción personal o por una simple cuestión de justicia.

Entre las diversas formas que existen de reconocer a un autor el mérito por el trabajo realizado, las menciones públicas o la concesión de premios a sus logros suelen tener un efecto efímero en nuestra memoria. Es por ello, que sociólogos relevantes, como el estadounidense Robert K. Merton, coinciden en señalar a la eponimia, proceso de creación de expresiones -los epónimos- que asocian un hallazgo con su descubridor, como el mejor sistema de atribución del mérito científico, pues proporciona el reconocimiento más prestigioso y persistente en el tiempo con el que un autor consigue que su nombre forme parte del lenguaje científico de por vida. En definitiva, la eponimia es la mejor forma de inmortalizar a los artífices de los logros de la ciencia.

No obstante, la eponimia, como sistema de reconocimiento del mérito científico, presenta el inconveniente de que, en ocasiones, el epónimo establecido no hace honor al descubridor original del hallazgo en cuestión. Estas situaciones de injustica, provocadas porque el mérito se concede a un autor de reconocido prestigio en detrimento del descubridor original menos conocido o porque el descubrimiento del autor original no fue tenido en cuenta en su época y acaba llevando el nombre de otro autor en un momento posterior, quedan reflejadas a través de tres conocidas leyes relacionadas con la eponimia: *Efecto Mateo*, *Ley de Stigler* y *Ley de Boyer*. Precisamente, la *Ley de Stigler* postula que los descubrimientos científicos no llevan el nombre de su descubridor original.

Puesto que nuestro interés se centra en el ámbito de la Estadística, el objetivo de este trabajo es comprobar si algunos de los epónimos estadísticos más conocidos (distribución normal, distribución de Poisson, distribución t de Student, distribución F de Snedecor, distribución de Cauchy y teorema de Bayes) son casos o no de la *Ley de Stigler*. Para ello, se analiza la historia del proceso de formación de cada uno de los epónimos para conocer cómo se ha llegado a la expresión que les da nombre y poder así determinar si evidencian o no la citada Ley.

Por tanto, el trabajo comienza (sección 2) describiendo el papel de la eponimia como sistema de reconocimiento del mérito científico. En la sección 3 se desarrollan las tres leyes relacionadas con la eponimia que tratan de reflejar las situaciones de injusticia que suelen producirse al establecer los epónimos en el ámbito científico. El análisis de la historia de algunos de los epónimos más conocidos en Estadística se presenta en la sección 4 con el objetivo de determinar si son o no casos de la *Ley de Stigler*. Finalmente, en la sección 5 se recogen las principales conclusiones del estudio.

2. La eponimia y el reconocimiento del mérito científico

La eponimia, como proceso de creación de expresiones –epónimos-, que se asignan al autor de un descubrimiento o hallazgo, tiene como principal objetivo atribuir el merecido reconocimiento a dicho autor, y desde hace varios siglos es un recurso que se viene empleando en numerosos ámbitos de la ciencia, principalmente, en medicina, química, física, matemáticas y economía.

Precisamente, el estadounidense Robert K. Merton, uno de los sociólogos más relevantes del siglo XX, precursor de la Sociología moderna y fundador de la sociología de la ciencia[1], siempre destacaba la importancia del papel de la eponimia en sus trabajos relacionados con el sistema de reconocimiento en la ciencia. De hecho, en su discurso de 1957, *Priorities in Scientific Discovery…*[2], defiende que la eponimia, que define como:

> …la práctica de poner el nombre del científico a todo o parte de lo que ha descubierto, como la Ley de Hooke, la constante de Planck o el Cometa Halley [Merton, 1957, p. 643],

es la mejor forma de reconocer el mérito científico, pues considera que es la más duradera y prestigiosa. En efecto, para Merton la eponomia es la manera en la que los científicos dejan su huella en la historia, de que sus nombres formen parte de todos los lenguajes científicos del mundo.

Merton indica que los epónimos que se establecen en la práctica se pueden ordenar jerárquicamente en tres niveles. En el nivel superior se encuentran los más escasos, los que nombran una época entera en honor de hombres y mujeres que han puesto su sello en la ciencia, como, por ejemplo, época Newtoniana, era Darwiniana o era Freudiana.

El siguiente nivel aglutina un número mayor de epónimos mediante los cuales, a científicos brillantes se les reconoce "padres" de una ciencia particular, es decir, se les atribuye haber engendrado una nueva ciencia o una nueva rama de la ciencia. Entre los reconocidos padres

1. La sociología de la ciencia considera los aspectos sociales al analizar a una comunidad científica, es decir, investiga el comportamiento y organización social de los científicos y todo aquello que los estimula, recompensa o intimida.
2. Discurso presidencial leído en la reunión anual de la *American Sociological Society* en agosto de 1957.

ilustres de una ciencia o especialidad podríamos citar a Faraday, padre de la Electrotecnia; Aristóteles, padre de la Lógica; Daniel Bernoulli, padre de la Física Matemática; Euclides, padre de la Geometría; Pearson, padre de la Biometría; Hipócrates, padre de la Medicina; Keynes, padre de la macroeconomía;… En otros casos, los epónimos dan testimonio del primer científico que desarrolló una vertiente de una disciplina determinada, como ocurre con la lógica aristotélica, la geometría euclidiana, la medicina hipocrática, la economía keynesiana, por citar algunos ejemplos.

Finalmente, el tercer nivel está formado por miles de epónimos referidos a leyes, teorías, teoremas, hipótesis, instrumentos, constantes y distribuciones, con las que gran cantidad de hombres y mujeres han contribuido a la ciencia, por lo que cualquier lista de epónimos de este nivel que se confeccione no sería representativa del total. Sea como fuere, cada ciencia desarrolla sus propios patrones distintivos de eponimia para honrar a quienes hacen de ella lo que es. Así, por ejemplo, en Medicina, la posteridad está asegurada para el primer descubridor o descriptor de una parte del cuerpo humano (trompa de Eustaquio, conducto de Stenon, nódulo de Aschoff, trompas de Falopio, esfínter de Oddi) o para la primera persona que diagnostica una enfermedad (Parkinson, Alzheimer, malformación de Chiari, síndrome de Asperger, hernia de Bochdalek, tiroiditis de Hashimoto); o en Física, donde se rinde homenaje a los grandes físicos asociando sus nombres a unidades eléctricas y magnéticas (ohmios, amperios, culombios, faradios, vatios). En el ámbito de la Estadística, objeto de este trabajo, son habituales, entre otros, los epónimos que tratan de reconocer el mérito al descubridor de una distribución de probabilidad, como sucede en casos tales como distribución de Gauss, distribución de Poisson, distribución de Cauchy, distribución F de Snedecor o distribución t de Student, que son precisamente los que se analizan en este artículo.

3. Leyes relacionadas con la eponimia

En el establecimiento de epónimos en el ámbito científico suelen producirse situaciones de injusticia, confusión o controversia que, a lo largo de la historia, han provocado que determinadas figuras científicas no hayan visto reconocida su labor. Estas situaciones quedan reflejadas por tres conocidas leyes que giran en torno a la eponimia: *Efecto Mateo*, *Ley de Stigler* y *Ley de Boyer*.

3.1. Efecto Mateo

Robert K. Merton [1968] denominó *Efecto Mateo* a la injusta asignación del mérito científico consistente en favorecer a los autores de reconocido prestigio frente a aquellos que aún no han dejado huella en la ciencia, que está inspirado en una cita bíblica que aparece recogida en el capítulo 13, versículo 12 del Evangelio de San Mateo:

> Qui enim habet, dabitur ei, et abundabit; Qui antem non habet, et quodhabet, auferetur ab eo

y que se traduce como:

> Porque al que tiene se le dará y tendrá en abundancia; pero al que no tiene incluso lo que tiene se le quitará

En esencia, Merton establece que el simple hecho de que un autor sea conocido y goce de una reputación en el ámbito científico, favorece el grado de reconocimiento de su trabajo, mientras que ser un autor poco conocido o que está iniciando su labor, conlleva muchas más dificultades para lograr dicho reconocimiento, incluso en los casos en los que se efectúa un hallazgo importante. En consecuencia, como señala Merton, el *Efecto Mateo* encierra una doble injusticia no intencionada ya que, de forma injustificada, por un lado, los científicos famosos se benefician, mientras que, por otro lado, los menos conocidos quedan en el anonimato. El propio Merton indica que incluso científicos galardonados con el Premio Nobel aportan evidencias del *Efecto Mateo*[3] como beneficiarios involuntarios.

3.2. Ley de Stigler

Stephen Stigler, profesor de estadística de la Universidad de Chicago nacido en 1941, en su trabajo de 1980, *Stigler's law of eponymy*, pone de manifiesto las dificultades que se pueden producir en el reconocimiento del mérito científico basado en el tercer nivel de eponimia que establece Merton, al afirmar que:

> No scientific discovery is named after its original discoverer" [Stigler, 1980, p. 147 y 1999, p. 277]

que traducido a nuestro idioma significa:

> Ningún descubrimiento científico recibe el nombre de quien lo descubrió en primer lugar.

Esta afirmación representa la forma más simple de lo que se conoce como *Ley de Stigler* y, según su autor, la generalidad con la que se manifiesta en las diferentes áreas científicas rivaliza con la de cualquier otra ley de las ciencias sociales. Precisamente, la *Ley de Stigler* está relacionada con una famosa hipótesis que Merton enuncia en su obra de 1973, *The Sociology of Science*,

> Todos los descubrimientos científicos son en principio múltiples. [Merton, 1973, p. 356]

siendo evidente que si todos los descubrimientos científicos pueden ser múltiples es razonable pensar que puedan acabar no llevando el nombre de su primer descubridor. Por tanto, en cierta forma, la *Ley de Stigler* fue anticipada por Merton, y esta es la razón por la que se dice que la *Ley de Stigler* es un ejemplo de ella misma.

3. Para definir el Efecto Mateo, Merton se basó en la investigación que, durante la década de los años 60, una joven colaboradora de su grupo de trabajo, Harriet Zuckerman, realizaba para su tesis doctoral en relación con las características de la élite científica y en la que entrevistó a científicos estadounidenses galardonados con el Premio Nobel. Paradójicamente, Zuckerman resultó ser "víctima" del Efecto Mateo, pues Merton no le reconoció públicamente su trabajo ya que en su artículo de 1968 las referencias a Harriet se limitaron a dos frases cortas en las que mencionaba sus entrevistas a premios nobel estadounidenses. Precisamente, el caso de Zuckerman y el de la activista norteamericana por los derechos de la mujer, Matilda Joslyn Gage, inspiraron a la historiadora de la ciencia Margaret W. Rossiter [1993] para enunciar el Efecto Harriet-Matilda (más conocido como Efecto Matilda) que trata de reflejar la discriminación sufrida por las mujeres en el ámbito científico respecto a sus hallazgos pues, mayormente, son sus compañeros de trabajo de sexo masculino quienes se llevan el mérito de los mismos.

Aunque el propio Stigler reconoce la relación existente entre su ley y la hipótesis de Merton (llega a señalar incluso que cualquier idea que desarrolla en su artículo, si no está implícita en el trabajo de Merton de 1973, es fruto de la casualidad), indica que son distintas y que la ley no contiene necesariamente a la hipótesis, pues un hallazgo puede llevar el nombre de alguien que no sea su descubridor original ni de ninguno de los que se consideran múltiples descubridores.

En realidad, un estudio exhaustivo de la historia de cualquier disciplina permite encontrar evidencias a favor de la *Ley de Stigler*. Así, en el campo de la Estadística matemática (ámbito objeto de análisis por parte de Stigler) podría citarse a este respecto, por ejemplo, que Laplace empleó la Transformada de Fourier con anterioridad a que Fourier publicara sobre ese tema; que Lagrange presentó la Transformada de Laplace antes de que Laplace comenzara su carrera científica; o que Poisson publicó la Distribución de Cauchy en 1824, 29 años antes de que Cauchy hiciera referencia a ella de forma accidental. Esta es solo una muestra de los muchos casos que podrían mencionarse en este ámbito, algunos de los cuales serán objeto de análisis en la sección 4 de este artículo.

Principalmente, hay tres razones por las que encontramos evidencias del cumplimiento de la *Ley de Stigler* en los diferentes ámbitos científicos. La primera de ellas hace referencia a que, en ocasiones, se produce un descubrimiento múltiple de forma simultánea, pero por algún motivo sólo uno de los descubridores se lleva el mérito del hallazgo. Si en este mismo contexto uno de los múltiples descubridores goza de prestigio en el ámbito correspondiente, probablemente el descubrimiento acabe llevando su nombre, circunstancia que también se produce cuando la figura relevante simplemente tiene acceso al descubrimiento y no ha participado en el mismo. Por último, hay casos en los que una aportación no es tomada en consideración en un momento determinado bien por contradecir algún dogma de la época o, simplemente, por ser demasiado avanzada a su tiempo. Este hecho provoca que en un momento posterior, la aportación vuelva a ser postulada con el reconocimiento merecido, pero portando el nombre del nuevo descubridor y no el del original.

3.3. Ley de Boyer

La esencia de la *Ley de Stigler* ya fue planteada en el ámbito exclusivo de las Matemáticas en 1972 por un autor americano profesor de matemáticas, Hubert Kennedy, como se infiere inequívocamente de su afirmación:

> Por lo general, las fórmulas y teoremas matemáticos no reciben el nombre de sus descubridores originales. [Kennedy, 1972, p. 67]

la cual es fruto del análisis de la obra de Carl Benjamin Boyer [1968], *A History of Mathematics,* que contiene numerosos ejemplos de dicha afirmación[4]. Kennedy concluyó que debería de existir una ley que agrupara esos hechos y le asignó el nombre del autor del libro que le condujo a formular su reflexión, *Ley de Boyer*.

4. Análisis de algunos epónimos estadísticos

En esta sección se analizan en profundidad algunos de los epónimos más conocidos surgidos en el ámbito de la Estadística con el objetivo de conocer la historia de cada uno de ellos y poder así determinar si son o no casos de la *Ley de Stigler*.

4.1. Distribución normal

La distribución de probabilidad normal de parámetros μ y σ, cuya función de densidad es

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

con forma acampanada y simétrica respecto a μ, tradicionalmente se asocia con el matemático Carl Friedrich Gauss. De hecho, es habitual que se la denomine distribución de Gauss y a su densidad campana de Gauss.

El origen de esta atribución se encuentra en un trabajo de 1809 en el que el autor relacionó la distribución normal con el método de mínimos cuadrados[5] en el contexto de la determinación de órbitas planetarias. En concreto, el objetivo de Gauss era la estimación de las k constantes desconocidas que determinan la órbita elíptica en base a $n > k$ observaciones, teniendo en cuenta la incertidumbre introducida por los errores de medición. Así, Gauss utilizó el criterio de mínimos cuadrados para realizar la citada estimación, es decir, para localizar la órbita que mejor se ajustaba a los datos que se tenían, y justificó su aplicación considerando una teoría de los errores de medición que estaba basada en los tres supuestos siguientes:

1. Los errores pequeños son más probables que los errores grandes.
2. Para cualquier número real ε la probabilidad de que se cometa en la medición un error de magnitudes $+\varepsilon$ y $-\varepsilon$ es la misma.

---

4. Alguno de los ejemplos citados por Boyer son las *Series de Taylor* que ya fueron trabajadas por el matemático y astrónomo escocés James Gregory cuando Brook Taylor era un niño; las Leyes de De Morgan, formuladas formalmente por Augustus De Morgan inspirándose en las afirmaciones del matemático británico George Boole; o el Triángulo de Pascal que no se debe al matemático francés Blaise Pascal pues únicamente se limitó a reunir resultados ya conocidos sobre el triángulo y a utilizarlos para demostrar diversas propiedades de la teoría de la probabilidad.
5. Parece ser que en 1808, esto es, un año antes de que apareciera el libro de Gauss, el matemático Robert Adrain ya había publicado algo relacionando la distribución normal con los mínimos cuadrados, en una revista poco conocida [Stigler, 1999, p. 284].

3. Dadas varias mediciones de la misma cantidad, el valor más probable de la cantidad que se mide es el promedio de dichas mediciones.

Basándose en estas suposiciones, Gauss llegó a la conclusión de que la densidad de probabilidad del error de medición era,

$$\phi(x) = \frac{h}{\sqrt{\pi}} e^{-h^2 x^2}$$

donde *h* es una constante positiva que el autor consideró como la "precisión del proceso de medición", y que en la notación actual sería la densidad de una normal con $\mu = 0$ y $\sigma = 1/\sqrt{2}h$

No obstante, Pierre Simon Laplace ya había aludido a la distribución normal en 1774, en su trabajo sobre el error en la medida de las observaciones. En dicho trabajo Laplace propuso la (primera) curva de error, que denotó $\phi(x)$, estableciendo que debía ser simétrica en *x* y monótona decreciente para $x > 0$. Como no se sabía nada acerca de la relación entre $\phi(x)$ y $\phi'(x)$, Laplace asumió que eran proporcionales[6], lo cual le llevó a la ecuación diferencial

$$\frac{d\phi(x)}{dx} = -m\phi(x)$$

cuya solución es precisamente la mencionada primera curva de error de Laplace

$$\phi(x) = \frac{m}{2} e^{-m|x|}$$

siendo su representación la que aparece en la figura 1 y en la que destaca el hecho de que no es diferenciable en $x = 0$.

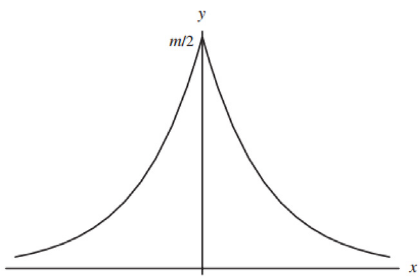

Figura 1. Primera curva de error de Laplace. Fuente: *The Evolution of the Normal Distribution* [Stahl, 2006, p. 103]

6. Laplace argumentó [Stigler, 1986b, 1, p. 372]: "…*como no tenemos una razón para suponer una ley diferente para las ordenadas que para sus diferencias, se asume, en base a las reglas de la probabilidad, que la proporción de dos diferencias consecutivas infinitesimales es igual a la de las correspondientes ordenadas. Así, se tiene*

$$\frac{d\phi(x+dx)}{d\phi(x)} = \frac{\phi(x+dx)}{\phi(x)}"$$

Aunque la densidad $\phi(x)$ de Laplace es distinta de la de la normal -la primera no es derivable en 0 y la segunda lo es en todo punto-, es evidente que la expresión y forma de $\phi(x)$ constituyen una aproximación a la densidad de la distribución normal[7] (hecho que el propio Gauss reconoció en su obra de 1809). Por esta razón, en ocasiones, a la distribución normal se la denomina distribución de Laplace-Gauss.

El caso de Laplace hizo pensar que el origen de la distribución normal podría remontarse a fechas más tempranas, y propició investigaciones que han descubierto nombres que se pueden vincular con dicha distribución pero que no están reconocidos mediante un epónimo. En este sentido, destaca, por encima de todas las demás, la figura de Abraham De Moivre, autor que se le relaciona con el descubrimiento de la distribución normal en el contexto de su trabajo con la distribución binomial[8] que, como es sabido, su función de probabilidad permite determinar la probabilidad de obtener x éxitos en las n repeticiones de un experimento de Bernoulli con probabilidad de éxito p y fracaso $q = 1 - p$

$$p(x) = \binom{n}{x} p^x q^{n-x} \tag{1}$$

En concreto, De Moivre se centró en el problema de cálculo que se presentaba en (1) cuando n era un valor grande o simplemente cuando se planteaban sumas del tipo

$$\sum_{x=i}^{j} \binom{n}{x} p^x q^{n-x}$$

y propuso dos aproximaciones [1733][9] para el caso particular $p = 1/2$

$$P\left(X = \frac{n}{2} + l\right) = \binom{n}{\frac{n}{2}+l} \left(\frac{1}{2}\right)^n \approx \frac{2}{\sqrt{2\pi n}} e^{-2l^2/n} \tag{2}$$

y

$$\sum_{l=0}^{s\sqrt{n}} P\left(X = \frac{n}{2} + l\right) = \sum_{l=0}^{s\sqrt{n}} \binom{n}{\frac{n}{2}+l} \left(\frac{1}{2}\right)^n \approx \frac{2}{\sqrt{2\pi n}} \int_0^{s\sqrt{n}} e^{-2l^2/n} dl \tag{3}$$

7. Laplace propuso en 1777 otra curva de error alternativa a $\phi(x)$ pero en la que no estaba implicada la función exponencial e^x.

8. La distribución binomial surgió de entre los problemas que consideraron Pierre de Fermat y Blaise Pascal como resultado de la correspondencia que ambos mantuvieron en 1654, a raíz de las consultas que les realizó el jugador Chevalier de Méré [David, 1962]. A esta famosa correspondencia se le atribuyen los orígenes de la teoría matemática de la probabilidad.

9. La investigación que realizó De Moivre le llevó, además, a ser el primero -en 1730- en obtener una aproximación para el factorial de un número n elevado

$$n! \approx (constante)\sqrt{n}n^n e^{-n}$$

Poco después, James Stirling determinó que $constante = \sqrt{2\pi}$, y el reconocimiento que de esta aportación efectuó De Moivre en su libro de 1738, hizo que la expresión pasara a la historia como fórmula de Stirling, que es como se la conoce actualmente.

que son válidas siempre que n fuese una cantidad grande en relación a l. Como fácilmente se puede deducir, en ambas expresiones está implicada la densidad de una normal con $\mu = 0$ y $\sigma = \sqrt{n}/2$.

A partir de estos resultados De Moivre pudo aproximar la probabilidad de que una binomial simétrica (p=1/2) tomara valores en un intervalo de la forma $\left[\frac{1}{2}n - l, \frac{1}{2}n + l\right]$ donde[10] $l = \frac{1}{2}a\sqrt{n}$. Así, teniendo en cuenta que $s = \frac{1}{2}a$ en (3), tras el correspondiente cambio de variable, resulta

$$P\left(\frac{n}{2} - a\frac{\sqrt{n}}{2} \leq X \leq \frac{n}{2} + a\frac{\sqrt{n}}{2}\right) \approx \frac{1}{\sqrt{2\pi}}\int_{-a}^{a} e^{-t^2/2}dt$$

siendo ahora la densidad de la $N(0,1)$ la que aparece en la expresión de la aproximación. Las aproximaciones de las probabilidades de la binomial que obtuvo De Moivre para diferentes valores de a [1733, pp. 4-6; 1738, pp. 238-241] fueron bastante precisas cuando se comparan con los cálculos exactos que proporcionan los métodos modernos [Stigler, 1986a, p. 82].

De Moivre generalizó las aproximaciones propuestas a un contexto asimétrico y estableció la media y desviación típica de la binomial cuando n tiende a infinito

$$\binom{n}{x} p^x q^{n-x} \xrightarrow{n \to \infty} N\left(np, \sqrt{npq}\right)$$

quedando así formalizado el conocido Teorema de De Moivre que, en la práctica, permite realizar la denominada aproximación de la binomial a la normal[11].

Aunque De Moivre no desarrollara una función de densidad de probabilidad, hay autores, como Karl Pearson [1926], que consideran la aproximación (2) como el origen de la curva normal. Stigler [1986a, p. 76] señala que, de hecho, el propio De Moivre, a pesar de no conferirle más importancia que la que atribuía a cualquier otra aproximación a la distribución binomial, siempre la concibió como una curva [De Moivre, 1730, pp. 109-110].

Pero, volviendo al tradicional vínculo que ha existido entre la distribución normal y los errores de medición, podemos identificar el primer autor, que en cierta forma, hizo referencia a características propias de dicha distribución. Se trata de Galileo Galilei, el primer científico en señalar por escrito, en su obra de 1632 (*Dialogue Concerning the Two Chief Systems of the World-Ptolemaic and Copernican*), que los errores de medición merecían un tratamiento sistemático y científico. Su análisis informal de las propiedades de los errores aleatorios inherentes a las observaciones de los fenómenos celestes, es resumido por Hald [2003, p. 150] en cinco puntos:

10. Puesto que una binomial de parámetro p = 1/2 tiene esperanza $n/2$ y varianza $n/4$, lo que expresa el intervalo planteado es que la variable en cuestión no se separe de su media más de a veces su desviación típica ($\sqrt{n}/2$).

11. Aunque la segunda edición de *The Doctrine of Chances* (1738) incluye el teorema límite, De Moivre ya lo había publicado por separado en 1733.

1. Sólo hay un número que da la distancia de la estrella al centro de la Tierra, la distancia verdadera.

2. Todas las observaciones que se toman están cargadas de errores, debido al observador, a los instrumentos y a las demás condiciones observacionales.

3. Las observaciones se distribuyen simétricamente alrededor del valor verdadero, es decir, los errores se distribuyen simétricamente alrededor de cero.

4. Los errores pequeños ocurren con mayor frecuencia que los errores grandes.

5. La distancia calculada es una función de las observaciones angulares directas, de modo que pequeños ajustes en las observaciones pueden provocar un gran ajuste de la distancia.

Si bien los puntos 3 y 4 son características de la distribución normal, Galileo no abordó la cuestión de cómo debía estimarse la distancia verdadera [Stahl, 2006].

4.2. Distribución de Poisson

La distribución de Poisson se utiliza para caracterizar los denominados eventos o sucesos raros, sucesos con frecuencias de ocurrencia pequeñas dentro de una población amplia. Se trata pues, de sucesos que ocurren con una probabilidad baja, que viene dada por la función

$$p(x) = e^{-\lambda}\frac{\lambda^x}{x!} \qquad x = 0,\ 1,\ 2, \ldots \qquad (4)$$

donde λ es el parámetro que la caracteriza.

Esta distribución debe su nombre al matemático francés Siméon Denis Poisson con motivo de su trabajo de 1837[12] (*Recherches sur la Probabilité des Jugements en Matière Criminelle et en Matière Civile Précédées des Règles Générales du Calcul des Probabilités*) relacionado con la distribución binomial negativa. En concreto, Poisson, para un experimento Bernoulli en el que el suceso fracaso (F) lleva asociada probabilidad q, determinó la probabilidad de que F no ocurriera más de m veces en n repeticiones de dicho experimento sin ser F el último resultado que sucede. Pues bien, el límite de esta probabilidad acumulada o función de distribución (de una binomial negativa) cuando q tiende a cero y n tiende a infinito, manteniéndose constante el producto $nq = \lambda$, llevó a Poisson a la expresión

$$\sum_{x=0}^{m} e^{-\lambda}\frac{\lambda^x}{x!}$$

que es la de la función de distribución que lleva su nombre.

12. Aunque algunos autores fechan el trabajo de Poisson en 1832, ni el *Catalogue Generale des Livres Imprimes de la Biblioteque Nationale* (vol. 139, p. 982) ni la propia obra publicada hacen referencia a ninguna edición anterior a 1837 [Haight, 1967, p. 113].

No obstante, para algunos autores[13], fue Abraham De Moivre quien, aproximadamente un siglo antes -en 1718-, descubrió la distribución de Poisson, también como límite de la distribución binomial negativa, pero, a diferencia de Poisson, De Moivre no proporcionó de forma explícita la fórmula (4)[14]. En realidad, Poisson, siendo probablemente consciente del trabajo de De Moivre, añadió poco a lo que este autor ya había aportado al respecto [Good, 1986, 1, p. 166]. De hecho, el propio Poisson no se percató, en su momento, de la importancia que la distribución acabaría teniendo y la consideró como una más de las aproximaciones que estudió. Prueba de todo ello es que Poisson menciona la distribución una sola vez en toda su obra, y lo hace en una única página, la 206, de su libro de 1837 [Stigler, 1982, 1, p. 33].

Sin embargo, el primer autor en observar que las ocurrencias de acontecimientos con pequeñas frecuencias en una población amplia (los denominados sucesos raros) se ajustaban a una distribución de Poisson fue Ladislaus von Bortkiewicz, economista y estadístico ruso de ascendencia polaca, en 1898. El hallazgo se produce en el contexto de un estudio en el que Bortkiewicz quería conocer la posibilidad de que un soldado prusiano muriera por la coz de un caballo. Para ello, tomó el número de soldados muertos por coz de caballo en 14 cuerpos del ejército prusiano en un período de 20 años, de 1875 a 1894 (figura 2). Pero, Bortkiewicz, buscando que la muestra estuviera constituida por cuerpos similares en cuanto al número de escuadrones de caballería se refiere[15], prescindió de 4 de ellos (G, I, VI y XI). Por tanto, la muestra final quedó constituida por los datos de 10 cuerpos a lo largo de 20 años, es decir, por un total de 200 observaciones, entre las que se contabilizaron un total de 122 muertes por coz de caballo, lo cual dio lugar a un promedio de 122/200 = 0,61 muertes por cuerpo y año. El valor de la media (inferior a la unidad) inducía a pensar que en un año dado se esperaba observar cero muertes, a veces una, ocasionalmente dos, quizá de vez en cuando tres, y muy raramente más. En consecuencia, se estaba ante un suceso raro y que probabilísticamente se describía mediante la ya conocida distribución de Poisson, tal y como Bortkiewicz comprobó al tomar $\lambda = 0{,}61$ en (4) para determinar el número de años en el que se esperaba que ocurriera un número concreto de muertes por coz de caballo en el periodo de 200 años, pues coincidía prácticamente con el número real de años en el que esa cantidad de muertes fueron observadas (tabla 1).

13. Haight [1967, p. 113] cita, en concreto, a tres: Newbold [1927], Jensen [1954] y David [1962].

14. Se podría afirmar, incluso, que, en cierta forma, Pierre Rémond de Montmort pudo anticiparse a De Moivre -en 1708- con su discusión del problema de la concordancia (*the matching problem*). Este problema plantea lo siguiente: dadas dos barajas de n cartas, numeradas cada una de 1 a n, que se barajan por separado y se extienden en dos hileras, se trata de determinar la probabilidad de que se produzcan r coincidencias, es decir, de encontrar r parejas de cartas iguales situadas cada una en la misma posición en su respectiva hilera. Pues bien, De Moivre demostró que cuando $n \to \infty$ dicha probabilidad venia dada por $e^{-1}/r!$, lo cual se corresponde con una Poisson de parámetro 1, pero antes, en 1708, Montmort ya había deducido el resultado para $r = 0$, pues, había concluido que era e^{-1}. Incluso en 1713 publicó una demostración de este caso que recibió de Nicolas Bernoulli.

15. El motivo de elegir cuerpos con similar número de escuadrones de caballería es que podrían compartir una tasa común de mortalidad por coz de caballo.

Por lo expuesto en el párrafo anterior, hay autores [Good, 1986] que consideran que la distribución de Poisson debió acabar denominándose *distribución de Bortkiewicz*.

	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94
G	—	2	2	1	—	—	1	1	—	3	—	2	1	—	—	1	—	1	—	1
I	—	—	—	2	—	3	—	2	—	—	—	1	1	1	—	2	—	3	1	—
II	—	—	—	2	—	2	—	—	1	1	—	—	2	1	1	—	—	2	—	—
III	—	—	—	1	1	1	2	—	2	—	—	—	1	—	1	2	1	—	—	—
IV	—	1	—	1	1	1	1	—	—	—	—	1	—	—	—	—	1	1	—	—
V	—	—	—	—	2	1	—	—	1	—	—	1	—	1	1	1	1	1	1	—
VI	—	—	1	—	2	—	—	1	2	—	1	1	3	1	1	1	—	3	—	—
VII	1	—	1	—	—	—	1	—	1	1	—	—	2	—	—	2	1	—	2	—
VIII	1	—	—	—	1	—	—	1	—	—	—	—	1	—	—	—	1	1	—	1
IX	—	—	—	—	—	2	1	1	1	—	2	1	1	—	1	2	—	1	—	—
X	—	—	1	1	—	1	—	2	—	2	—	—	—	—	2	1	3	—	1	1
XI	—	—	—	—	2	4	—	1	3	—	1	1	1	1	2	1	3	1	3	1
XIV	1	1	2	1	1	3	—	4	—	1	—	3	2	1	—	2	1	1	—	—
XV	—	1	—	—	—	—	—	1	—	1	1	—	—	—	2	2	—	—	—	—

Figura 2. Datos prusianos de Bortkiewicz* [1898]. Fuente. *Das Gesetz der Kleinen Zahlen* [Bortkiewicz, 1898, p. 25].

*Datos tomados de 20 volúmenes de la revista *Preussischen Statistik*, publicada en Berlín en los siglos XIX y XX con estadísticas oficiales relativas a Prusia. En las columnas se indican los años (1875 a 1894) y en las filas se identifica el cuerpo.

Tabla 1. Los datos prusianos de Bortkiewicz [1898] se ajustan a una Poisson. Fuente. Elaboración propia a partir de los datos de Bortkiewicz [1898].

Nº de muertes anuales en el cuerpo		Nº de años con x_i muertes en el cuerpo (en 200 años)	
Valores (x_i)	Probabilidad ocurrencia (p_{x_i}) ↓ Poisson (λ = 0,61)	Esperado ↓ $200 * p_{x_i}$	Observado
0	0,54335	108,67	109
1	0,33145	66,29	65
2	0,10109	20,22	22
3	0,02056	4,11	3
4	0,00313	0,63	1
≥5	0,00042	0,08	0

4.3. Distribución *t* de Student

La distribución *t* de Student surge del estudio publicado en 1908 por William Sealy Gosset, con el título *The Probable Error of a Mean*[16], en relación con la distribución del estadístico

$$z = \frac{\bar{X} - \mu}{S} \qquad (5)$$

donde

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad\qquad S = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}$$

para pequeñas muestras en un contexto de normalidad, esto es, asumiendo que $X_i \sim N(\mu,\sigma)$ $i = 1,\ldots, n$. Bajo este supuesto, lo que se sabía hasta ese momento era que

$$\frac{\bar{X}}{\sigma} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

y que esta normalidad se mantenía para tamaños $\bar{S}$ muestrales grandes cuando el valor de σ no estaba disponible y se sustituía por S (o por como sugería Gauss)[17]. Pues bien, Gosset cuestionó la normalidad de (5) cuando n era pequeña, y mediante un análisis profundo basado en algunas suposiciones certeras que no estaban respaldadas por demostraciones rigurosas, obtuvo la distribución exacta de (5)[18]. Ronald Fisher [1925a] la acabaría denominando distribución *t* de Student con $n - 1$ grados de libertad tras verificar que, la transformación del estadístico z que sugirió

$$t = \frac{\bar{X} - \mu}{S}\sqrt{n-1}$$

para contrastar la media poblacional en el contexto de normalidad (lo que se conoce como estadístico *t* para una muestra), tenía la misma distribución que Gosset había obtenido para z.

16. Gosset escribió su artículo durante la estancia que la fábrica de cerveza *Guinness* (en Dublín), para la que trabajaba como químico desde 1899, le permitió realizar en el laboratorio de Karl Pearson de la Universidad de Londres, en dos períodos distintos en el transcurso de los años 1906 y 1907. La política de empresa de *Guinness* prohibía que sus químicos publicaran sus hallazgos, pero Gosset, tras argumentar que las conclusiones de su trabajo no tenían utilidad práctica para cerveceros competidores, consiguió autorización para publicar sus resultados, pero hubo de hacerlo bajo el pseudónimo "Student". Según Hotelling [1930, p. 189] el anonimato de Gosset fue diseñado por *Guinness* para ocultar su identidad a sus propios cerveceros y evitar así conflictos con la plantilla. Fue, finalmente, el director general de *Guinness*, Christopher Digges La Touche, quien sugirió el pseudónimo "Student" [Box, 1987, 2, p. 49].

17. $\bar{S} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}$

18. Para este resultado, Gosset asumió que la incorrelación entre $\bar{X}$ y S^2 -circunstancia que llegó a demostrar- implicaba su independencia, lo cual fue una suerte, pues esto sólo es cierto para el caso normal.

En su mismo trabajo de 1908, Gosset quiso proporcionar una evidencia visual de la no normalidad de z para muestras pequeñas y facilitó la representación gráfica de la densidad de su distribución para el caso n = 10 (la de una t de Student con 9 grados de libertad) en comparación con la densidad de una normal con la misma media (0) y la misma desviación típica[19] ($1/\sqrt{7} = 0{,}378$).

En la figura 3 se aprecian las diferencias entre ambas funciones de densidad (la línea sólida se corresponde con la t (9) y la línea discontinua con la normal) en la que Gosset indicó que, aunque la concordancia no era mala, para grandes desviaciones (de la media muestral en relación a la media poblacional) la normal da "una falsa sensación de seguridad". Así, tras observar que cuanto mayor era n, su curva se aproximaba más a la normal, decidió tabular el valor de la función de distribución (área acumulada) de z para n = 4, 5,…, 10 [Gosset, 1908, 6, p. 19].

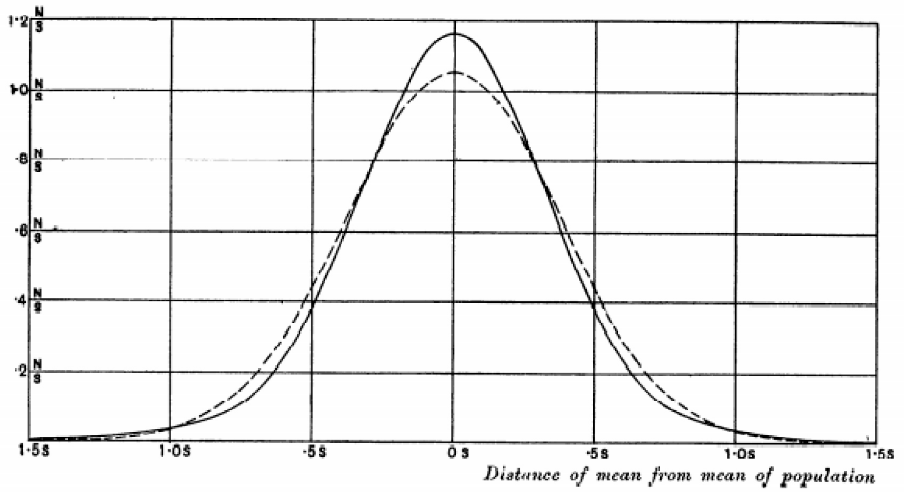

Figura 3. Densidad de la distribución t de Student y de la Normal*. Fuente. *The probable Error of a Mean"* [Gosset, 1908, 6, p. 12]
*línea sólida → t(9) línea discontinua → normal

El artículo de Gosset fue prácticamente ignorado tras su publicación, pues inicialmente no contó con el atractivo de su potencial aplicabilidad al control de calidad de la cerveza ni

19. A partir de la distribución exacta del estadístico (5), Gosset [1908, 6, p. 11] demostró que su varianza era $1/(n - 3)$.

fue considerado como un resultado destacable del grupo de trabajo de Pearson. Su impacto en la práctica estadística no se produjo hasta que Ronald Fisher lo incluyó en su manual *Statistical Methods for Research Workers* de 1925b [Stigler, 2016, p. 96].

4.4. Distribución *F* de Snedecor

El origen de la distribución *F* de Snedecor está vinculado a la distribución del estadístico que propuso Ronald Fisher [1924][20] para comparar las varianzas, σ_1^2 y σ_2^2, de dos poblaciones en un contexto de normalidad e independencia.

En concreto, estimando dichas varianzas a partir de muestras aleatorias de tamaños *n* y *m* extraídas de cada población, mediante las correspondientes varianzas muestrales, s_1^2 y s_2^2, respectivamente, Fisher sugirió el estadístico

$$z = \frac{1}{2}\log_e\left(\frac{s_1^2}{s_2^2}\right)$$

basado en la razón de varianzas s_1^2/s_2^2, para contrastar la igualdad[21] entre σ_1^2 y σ_2^2.

Fisher [1924] dedujo la distribución de probabilidad de *z*, la cual dependía de los valores $n_1 = n - 1$ y $n_2 = m - 1$ (grados de libertad en los que estaban basadas las estimaciones s_1^2 y s_2^2, respectivamente), y obtuvo algunos de sus cuantiles, los de orden 0,05, para determinados valores de n_1 y n_2, que publicó en la primera edición de su obra *Statistical Methods for Research Workers* [1925b, Tabla VI, p. 210][22].

En 1934, con el objetivo de eliminar el logaritmo natural de la expresión del estadístico *z*, George W. Snedecor, en su monografía *Calculation and Interpretation of Analysis of Variance and Covariance*, propuso un estadístico alternativo, que denotó *F*, y que no era más que la razón de varianzas

$$F = \frac{s_1^2}{s_2^2} \tag{6}$$

dando origen así a lo que hoy conocemos como prueba *F*.

Es evidente que los estadísticos *z* y *F* se relacionan en la forma

$$z = \frac{1}{2}\log_e\left(\frac{s_1^2}{s_2^2}\right) = \frac{1}{2}\log_e F \tag{7}$$

20. Fisher presentó su trabajo en The International Congress of Mathematics, celebrado en Toronto en 1924, cuyas actas no fueron publicadas hasta 1928.

21. El conocido análisis de la varianza de Fisher no es más que la aplicación de su prueba *z* a un caso particular [Fisher, 1925b, p. 192].

22. A partir de la tercera edición de la obra (1930), Fisher, además de ampliar el rango de valores de n_2, incluyó la tabla con los cuantiles de orden 0,01.

la cual permitió a Snedecor [1934] derivar la distribución de probabilidad de su estadístico[23], además de tabular algunos de sus cuantiles [1934, tabla XXXV y 1938, pp. 184-187] a partir de los que Fisher había obtenido para *z*.

A pesar de que el fin perseguido por Snedecor no era introducir una nueva distribución sino facilitar los cálculos del análisis de la varianza de Fisher, lo cierto es que la distribución de probabilidad asociada al estadístico (6) acabaría denominándose distribución *F* de Snedecor[24].

Se cree que la intención de Snedecor al elegir la letra *F* para designar su estadístico fue su manera de admitir la relación (7) y, por extensión, de realizar un reconocimiento a Fisher. No obstante, a menudo se ha comentado que Fisher no lo entendió así, pues se dice que no le pareció bien que Snedecor renombrara su estadístico *z* como *F*, y que ni siquiera se mostró satisfecho con la elección de la letra. Esta reacción de Fisher se desprende de la correspondencia que mantuvo con otro estadístico, H.W. Heckstall-Smith, y el propio Snedecor. Así, en una carta dirigida a Heckstall-Smith el 9 de febrero de 1956, señalaba que la ventaja que pudiera tener la razón de varianzas frente a su logaritmo natural requería de una máquina de calcular de la que muy pocos disponían:

> Gracias por su carta[25]. Siempre he considerado que la prueba *F* y la prueba *z* son lo mismo y, en efecto, en 1924 yo evalué la razón de varianzas a través de la obtención de su logaritmo natural. Por supuesto, si todos tuvieran una máquina de calcular a mano en su escritorio, los ratios-varianza son los más rápidos de obtener, pero aquellos que no tienen computadoras y que pueden usar tablas de cuatro cifras son cinco veces más numerosos… [Bennett, 1990, p. 319]

En la misma carta, Fisher ponía en duda que Snedecor hubiese elegido la *F* en su honor:

> …Creo que fue sólo una reflexión tardía lo que llevó a Snedecor a decir que la letra mayúscula *F* que había usado tenía el propósito de ser un cumplido para mí… [Bennett, 1990, p. 319]

Asimismo, indicó que, en la 12ª edición de *Statistical Methods for Research* Workers, publicada en 1954, había añadido una breve nota histórica advirtiendo que el gran uso que se hacía en Estados Unidos del símbolo *F* había llevado a que la distribución de *z* fuera referenciada a menudo, erróneamente, con la letra *F*. Con esta nota, Fisher, albergaba la esperanza de que los profesionales de la enseñanza evitaran identificar la prueba *F* con la prueba *z*, y que de esta forma dejaran de producirse confusiones:

> … Debido a los malentendidos que se han extendido en el extranjero, especialmente en los Estados Unidos, en la 12ª edición de *Statistical Methods* (página 226) he añadido una breve nota histórica que espero que evite que los profesores identifiquen la prueba *F*, o 'test de Snedecor' como dicen los franceses, con la prueba *z* o 'test de Fisher'… [Bennett, 1990, p. 319]

---

23. La función característica de la distribución *F* se estuvo utilizando de manera incorrecta durante 48 años, hasta que en 1982 el economista Peter Phillips encontró la expresión correcta de dicha función.

24. En Aroian [1941] ya aparece por primera vez el término distribución *F*.

25. Heckstall-Smith había escrito previamente a Fisher indicando que, en un artículo que estaba elaborando para una revista médica, necesitaría usar la prueba *F* en lugar de la *z* ya que los logaritmos naturales no serían tolerados.

No obstante, las reticencias de Fisher a que Snedecor le hubiese asignado el símbolo *F* a la razón de varianzas provenían de mucho antes, como se pone de manifiesto en una carta que dirigió al propio Snedecor, el 16 de febrero de 1938, en respuesta a una sugerencia[26] que éste le había realizado:

> … Me temo que tengo una dificultad, que puede que usted no conozca, al asignar el símbolo particular *F* a la razón de varianzas, y es que, una tabla de este ratio derivada de mi tabla de *z*, se publicó en la India, creo que en Sankhya, antes que la suya. El autor, si recuerdo bien, usó un símbolo diferente - probablemente *x*. Creo que sería demasiado complicado hacer un subtítulo utilizando las diferentes denominaciones… [Bennett, 1990, p. 323]

En efecto, en 1932, el científico y estadístico indio P. C. Mahalanobis, en respuesta a las demandas de los investigadores de campo, que no estaban familiarizados con el uso de logaritmos naturales y tenían dificultades con el estadístico *z* de Fisher, elaboró seis tablas de cuantiles, alternativas a las de *z*, que publicó en el *Indian Journal of Agricultural Science*. Dos tablas fueron diseñadas para trabajar con logaritmos ordinarios (con base 10 en lugar de base *e*); otras dos tablas fueron elaboradas para trabajar directamente con el ratio de las desviaciones típicas en lugar de con el de las varianzas, y dos tablas más fueron diseñadas para la razón de varianzas directamente, sin recurrir a logaritmos naturales. En cada pareja de tablas, una correspondía al nivel de significación del 5% y la otra al del 1%. Aun admitiendo que Snedecor hubiera obtenido sus tablas sin tener conocimiento de las de Mahalanobis (que se cree que fue lo que sucedió a la vista del inicio del fragmento arriba trascrito), Fisher consideraba que la tabulación de Mahalanobis, que había utilizado el símbolo *x* para representar la razón de varianzas, tenía prioridad y, es por esta razón, que se mostraba contrario al uso del símbolo *F*.

4.5 Distribución de Cauchy

La distribución de Cauchy de parámetros x_0 y $\gamma > 0$, cuya función de densidad es

$$f(x) = \frac{1}{\pi\gamma\left[1+\left(\frac{x-x_0}{\gamma}\right)^2\right]}$$

se asocia con el matemático francés Augustin Louis Cauchy como resultado de una disputa mantenida con otro matemático galo, Irénée Jules Bienaymé, que prácticamente monopolizó las páginas de *Comptes Rendus* en el verano de 1853[27]. La disputa se inició con una crítica de Bienaymé [1853] al método de interpolación propuesto por Cauchy en 1835,

26. Snedecor había sugerido a Fisher que "sería clarificador asociar de alguna manera su tabla de la razón de varianzas con el símbolo F" [Bennett, 1990, p. 323].

27. En el verano de 1853, Cauchy publicó en *Comptes Rendus* siete memorias relacionadas con el tema, a lo que Bienaymé contestó con comentarios en 3 ocasiones y con un artículo.

y acabaría focalizada en una cuestión que Laplace mantenía en su memoria de 1810 en relación con el método de mínimos cuadrados.

Laplace defendía en dicho trabajo que, como mínimos cuadrados proporcionaba el mejor estimador lineal asumiendo errores de medida distribuidos normalmente, sucedía lo mismo para cualquier distribución del error, al menos, para muestras grandes. Pero Laplace basó su afirmación únicamente en el supuesto de que la distribución del error fuera simétrica. Pues bien, Cauchy, en defensa de su método de interpolación, demostró que, para ciertas distribuciones simétricas no normales de los errores de medida, mínimos cuadrados no proporcionaba el mejor estimador lineal. La densidad de una de estas distribuciones era[28]

$$f(\varepsilon) = \frac{k}{\pi}\frac{1}{1+k^2\varepsilon^2}$$

que se corresponde con lo que hoy conocemos como una Cauchy de parámetros $x_0 = 0$ y $\gamma = 1/k$, y es por ello que recibe el nombre del matemático francés.

No obstante, Cauchy le prestó poca atención a este caso pues, como ya se ha comentado, estaba más centrado en defender su método de interpolación frente a Bienaymé, que en cuestionar la demostración de Laplace.

Prueba de ello, es que no aludió a las dificultades que $f(\varepsilon)$ podría presentar para muestras grandes, y lo consideró como un caso más, entre los que se incluía la doble exponencial, de los que conducían a estimadores lineales diferentes a los mínimos cuadrados.

En esta misma línea, años antes, en 1824, Simeon Denis Poisson, tratando de clarificar y extender el trabajo de Laplace, demostró que la argumentación de Laplace fallaba cuando consideró una distribución de los errores de medida con densidad

$$f(x) = \frac{1}{\pi(1+x^2)}$$

es decir, la de una Cauchy de parámetros $x_0 = 0$ y $\gamma = 1$.

No obstante, Poisson no le dio importancia a este caso por considerar que no se daba en la práctica:

> No debemos tener este caso en consideración. Será suficiente señalar su particularidad y que verdaderamente no lo encontraremos en la práctica [Poisson, 1824]

Es posible que Cauchy no estuviera familiarizado con el trabajo de Poisson de 1824, aunque es bastante probable, como rival de Poisson, que leyera el artículo en el momento de

28. De las siete memorias que Cauchy publicó en 1853 en *Comptes Rendus*, con motivo de su disputa con Bienaymé, la densidad $f(\varepsilon)$ aparece en la cuarta [1853a] y en su continuación, la quinta [1853b].

su publicación. En cualquier caso, lo que sí es cierto es que Cauchy, no siendo estadístico[29], se había aproximado a la misma distribución que Poisson desde una perspectiva distinta: el primero cuestionando a un defensor de Laplace, y el segundo extendiendo y clarificando el trabajo de Laplace [Stigler, 1999, p. 336].

En realidad, se puede decir que la densidad de Cauchy había sido referenciada en trabajos anteriores como curva proporcional a un caso especial de la curva denominada Bruja de Agnesi, que aparece descrita al final del primer libro de los cuatro que componen el tratado de cálculo y geometría analítica, *Instituzioni Analitiche*, que en 1748 publicó la matemática italiana Maria Agnesi[30]. La curva, que fue estudiada previamente por Pierre de Fermat en 1659 y Guido Grandi en 1703, tiene por ecuación

$$f(x) = \frac{8a^3}{x^2 + 4a^2}$$

donde el parámetro *a* es el radio de la circunferencia con la que se construye la curva. En consecuencia, la densidad de una Cauchy de parámetros $x_0 = 0$ y $\gamma = 1$ es proporcional al caso particular de la Bruja de Agnesi para $a = 1/2$

$$f(x) = \frac{1}{1 + x^2}$$

4.6. Teorema de Bayes

El Teorema de Bayes fue atribuido a Thomas Bayes[31] tras la publicación póstuma[32] -en 1763- que de dicho teorema organizó Richard Price, clérigo y escritor británico amigo de Bayes, a partir de diversos escritos que la familia de Bayes, a la muerte de éste, le suministraron a Price para que decidiera si eran o no interesantes. La publicación se concretó en un ensayo

29. Las siete memorias en *Comptes Rendus* de 1853 son prácticamente los únicos artículos de Cauchy sobre estadística.

30. Maria Agnesi denominó la curva que describió en su obra de 1748 como "la versiera", pero fue Guido Grandi el primero que, en 1718, la llamó "versiera", del latín "versoria" (término de naturaleza naval), haciendo referencia a que la curva había surgido del seno verso. Aunque el único significado relevante que tenía la palabra "versiera" para los diccionarios de la época era "virar o girar", también era una abreviatura de "avversiera", que en italiano significa "demonio" [Stigler, 1999, pp. 333-334]. John Colson, profesor de Cambridge que se encargó de traducir al inglés la obra de María sin tener un gran dominio del idioma italiano, lo tradujo como "Witch" [1801, p. 222], que significa bruja, y la curva pasó a llamarse en todas las versiones inglesas del manual como "La curva de la Bruja de Agnesi".

31. Thomas Bayes (1702(?)-1761) fue un ministro disidente inglés que vivió en Tunbridge Wells desde 1731.

32. Aunque Bayes no publicó en vida ningún trabajo sobre Matemáticas, sus contemporáneos debieron reconocerle como una autoridad en la materia, pues fue elegido como miembro de la *Royal Society of London* en 1742 [Stigler, 1986a, p. 99]

titulado[33], *An essay towards solving a problem in the doctrine of chances*, en el que el propio Price contribuyó con una introducción, un apéndice y un escolio, al menos, pero incluidos de tal manera que no hay forma de determinar qué parte se debe a Price, pues el propio autor señala que Bayes ya tenía preparada una introducción.

A pesar de que durante muchos años no apareció un autor anterior reclamando la autoría del teorema, por la forma en la que se le atribuyó el mérito a Bayes y, a tenor de la Ley de Stigler, era previsible que esta situación no durara indefinidamente. Y así fue, pues en 1979 el psicólogo inglés Bernard Singer señaló por escrito -en particular, en la página 6 de su trabajo- que en un pasaje de la obra del filósofo británico David Hartley, *Observations on Man, His Frame, His Duty, and His Expectations,* publicada en 1749, podría encontrarse la evidencia de que Bayes no fuese el primer autor del teorema[34]. El pasaje en cuestión se ubica en la sección "propositions and the nature of assent", concretamente, en un párrafo que se podría traducir como sigue:

> Un ingenioso amigo me ha comunicado una solución para el problema inverso, en la que ha demostrado cuál es la esperanza, cuando un suceso ha ocurrido *p* veces y ha fallado *q* veces, de que la proporción original de las causas de ocurrencia o fallo del suceso se desvíe en cualquier grado de la de *p* respecto de *q*. Y de esta solución se desprende que, cuando el número de pruebas es muy grande, la desviación debe ser insignificante: lo cual demuestra que podemos esperar determinar las proposiciones y, a través de los grados, mediante una observación suficiente de sus efectos, toda la naturaleza de causas desconocidas. [Hartley, 1749, p. 339]

La descripción que se recoge en el párrafo transcrito en relación a la solución que *un ingenioso amigo* propone a Hartley para el denominado problema inverso[35], induce a pensar, de manera razonable, que se trata de la solución al problema de Bayes[36] [Stigler, 1983, 1999]. En efecto, se trata de una solución que permite inferir las proporciones originales de las causas del citado problema inverso, y con el número suficiente de efectos observados se pueden determinar la íntegra naturaleza de las causas desconocidas.

A la vista del contenido del citado párrafo y de ciertas fechas clave -el libro de Hartley se publicó 12 años antes de la muerte de Bayes y 15 antes de que Price publicara los documentos de Bayes-, es evidente que el *ingenioso amigo* al que alude Hartley es el primer autor que desarrolló el teorema. Por tanto, el interrogante que a continuación se suscita es cuál es la identidad del ingenioso amigo al que se refiere Hartley en el extracto.

---

33. Stigler [2013] ha descubierto que, inicialmente, el título del ensayo era "A Method of Calculating the Exact Probability of All Conclusions founded on Induction", tal y como se puede comprobar en la separata que la revista enviaba a los autores de un artículo antes de proceder a la publicación del mismo. Este título lleva a Stigler a especular con la idea de que Price y, posiblemente, Bayes buscaran con el ensayo una herramienta para combatir a David Hume en relación con una cuestión de Teología.
34. Al parecer, Bernard Singer fue el primero que destacó por escrito la relevancia del descubrimiento efectuado en la obra de Hartley [Stigler, 1999, p. 291].
35. El problema inverso es como era denominada la determinación de las causas a partir de las observaciones.
36. Edwards [1986] y Dale [1988] defienden que Hartley no aludió al problema de Bayes como sostiene Stigler, sino al uso inverso del teorema de Bernoulli. El propio Stigler [1999, p. 301] considera que la argumentación de estos autores es inconsistente con la redacción de Hartley.

En un principio, se podría pensar que el ingenioso amigo fuera Abraham De Moivre, porque justo en el párrafo anterior, Hartley alude al matemático francés, pudiéndose entender que siguiese hablando de él en el siguiente párrafo, pero refiriéndose a este como su ingenioso amigo. No obstante, Stigler [1999, p. 292] descarta esta posibilidad pues considera que Hartley contrasta ambos párrafos de forma deliberada adjudicando, en el segundo párrafo, a un ingenioso amigo la solución al inverso del problema que atribuye explícitamente a De Moivre en el párrafo precedente.

Descartado De Moivre como descubridor del teorema, Stigler [1983, 1999] llevó a cabo una investigación con el objetivo de obtener nombres de posibles autores que podrían haber sido ese ingenioso amigo de Hartley. Comenzó por la lista de suscripción del libro *Miscellanea Analytica…* de 1730 escrito por Abraham De Moivre, ya que dedujo que el autor del teorema tendría que haberse inspirado en este libro. La lista contenía 62 nombres entre los cuáles pueden citarse a John Craig, James Stirling, Colin Maclaurin o Brook Taylor, y en la que no aparece el nombre de Bayes.

En este punto de la investigación, Stigler decidió abrir otra línea centrada en los amigos de Hartley. En esta lista tampoco figuraba Bayes, pero destacó el nombre de Nicholas Saunderson, científico y matemático inglés que llegaría a ser el cuarto profesor lucasiano de Matemáticas en Cambridge. La particularidad de este nombre es que, además de ser amigo de Hartley, figuraba en la lista de suscripciones del libro de De Moivre. Por tanto, Stigler concluye que Nicholas Saunderson es el autor con más opciones de ser el "ingenioso amigo", pero no encontró pruebas fehacientes de ello.

En consecuencia, Stigler propone que puede escogerse como autor entre Saunderson y Thomas Bayes y, aunque su investigación conduce a Saunderson como posible descubridor del teorema, reconoce que la falta de evidencias concluyentes impide decantarse entre él o Thomas Bayes.

Finalmente, en un contexto en el que se cuestiona a Thomas Bayes como autor del teorema que lleva su nombre, es justo considerar que dicha autoría pudiera recaer en la figura de Richard Price, académico de la *Royal Society* que, como se ha indicado, publicó el teorema a través de un ensayo en el que incluyó aportaciones propias.

5. Conclusiones

El análisis de la historia de los epónimos estadísticos considerados en este trabajo ha puesto de manifiesto que la distribución normal, la distribución de Poisson y la distribución de Cauchy son tres casos claros de la Ley de Stigler. En efecto, la distribución normal tradicionalmente se ha asociado a Gauss, y el análisis de la historia de esta distribución demuestra que fue De Moivre el primero que la descubrió como límite de la distribución binomial. En lo que se refiere a la distribución de Poisson, se ha podido constatar que De Moivre, muchos años antes que Poisson, también obtuvo la distribución como límite de la distribución binomial negativa, mientras que Bortkiewicz fue el primero en detectar su aplicación como distribución de los sucesos raros. La distribución de Cauchy fue descubierta por Poisson antes de que Cauchy hiciera referencia a ella en su disputa matemática con Bienaymé.

En el caso de la distribución *t* de Student es evidente que en la expresión no aparece el nombre del autor que la descubre, Gosset, pero también es cierto que, desde un principio, siempre se supo que el pseudónimo Student hacía referencia a su persona. En consecuencia, en sentido estricto, este epónimo no debe considerarse como una evidencia de la Ley de Stigler.

En principio, la distribución *F* de Snedecor plantea dudas en relación a si es o no un caso de la Ley de Stigler. Es evidente, y así lo muestra el análisis, que la distribución de probabilidad del estadístico *F* es una aportación de Snedecor, pero también es cierto que la distribución de probabilidad del estadístico *z* de Fisher es función de la de *F*, por lo que la aportación de Snedecor es, digamos, de escasa originalidad. La introducción de la letra *F* en el epónimo, es un reconocimiento expreso de la relación entre ambas distribuciones, con independencia de las intenciones de Snedecor al incluirla. En consecuencia, en este caso sería más justo utilizar la expresión que han acuñado algunas obras en la que se hace referencia de manera explícita a Fisher: distribución *F* de Fisher-Snedecor.

Finalmente, en relación con el Teorema de Bayes, las investigaciones realizadas hasta ahora no permiten negar lo que históricamente se ha venido afirmando, es decir, que Thomas Bayes es su autor. Tan solo hay indicios, que no pruebas, de que el verdadero autor pudiera ser otra persona: Stigler [1980] apunta, en su investigación, a Nicholas Saunderson como la opción más probable.

Bibliografía

Adrain, Robert (1808) "Research concerning the probabilities of the errors which happen in making observations". *Analyst*, *1*, 93-109.

Agnesi, Maria Gaetana (1748) *Instituzioni analitiche ad uso della gioventú italiana.* Milano, Nella Regia Ducal Corte.

Agnesi, Maria Gaetana (1801) *Analytical Institutions.* London, Taylor & Wilks. Traducción de John Colson.

Aroian, Leo A. (1941) "A Study of R. A. Fisher´s *z* Distribution and the Related *F* Distribution", *The Annals of Mathematical Statistics*, *12(4)*, 429-448.

Bayes, Thomas (1763) "An essay towards solving a problem in the doctrine of chances". *Phylosophical Transactions of the Royal Society of London*, *53*, 370-418.

Bennett, John Herny (ed.) (1990) *Statistical Inference and Analysis: Selected Correspondence of R. A. Fisher*. Oxford, Oxford University Press.

Bienaymé, Irénée Jules (1853) "Remarques sur les différences qui distinguent l'interpolation de M. Cauchy de la methode des moindres carrés, et qui assurent la supériorité de cette méthode". *Comptes Rendus de l'académie des sciencies de Paris*, *37*, 5-13.

Bortkiewicz, Ladislaus von (1898) *Das Gesetz der Kleinen Zahlen.* Leipzig, Teubner.

Box, Joan Fisher (1987) "Guinness, Gosset, Fisher and Small Samples". *Statistical Science*, *2(1)*, 45-52.

Boyer, Carl B. (1968) *A history of mathematics.* New York, John Wiley & Sons, Inc.

Cauchy, Augustin-Louis (1835) "Mémoir sur l'interpolation". Litografía, Praga. Reimpreso en 1837 en *Journal de mathématiques pures et appliquées*, *1re série, 2*, 193-205.

Cauchy, Augustin-Louis (1853a) "Sur les résultats moyens d'observations de même nature, et sur les résultats les plus probables". *Comptes Rendus de l'académie des sciencies de Paris*, *37*, 198-206.

Cauchy, Augustin-Louis (1853b) "Sur la probabilité des erreurs qui affectent des résultats moyens d'observations de même nature". *Comptes Rendus de l'académie des sciencies de Paris*, *37*, 264-272.

Dale, Andrew I. (1988) "On Bayes' Theorem and the Inverse Bernoulli Theorem". *Historia Mathematica*, *15*, 348-360.

David, Florence Nightingale (1962) *Games, Gods and Gambling.* New York, Hafner Publishing Company.

De Moivre, Abraham (1718) *The Doctrine of Chances.* London, W. Pearson.

De Moivre, Abraham (1730) *Miscellanea analytica de seriebus et quadraturis.* London, J. Tonson and J. Watts.

De Moivre, Abraham(1733) *Approximatio ad Summam Terninorum Binomii $(a + b)^n$ in Seriem Expansi*. Suplemento en latín añadido a la obra *Miscellanea analytica…* de 1730.

De Moivre, Abraham (1738) *The Doctrine of Chances,* 2nd edition. London, H. Woodfall.

De Montmort, Pierre Remond (1708) *Essay d'Analyse sur les Jeux de Hazard.* Paris, Jacque Quillau.

Edwards, Anthony William Fairbank (1986) "Is the reference in Hartley (1749) to Bayesian Inference?". *The American Statistician*, *40(2)*, 109-110.

Fermat, Pierre (1659) "De aequationum localium transmutatione et emendatione ad multimodaum curvilineorum inter se vel cum rectilineis comparationem, cui annectitur proportionis geometricae in quadrandis infinitis parabolis et hyperbolis usus", publicado en *Varia Opera Mathematica* en 1679, *8*, 44-57.

Fisher, Ronald Aylmer (1924) "On a distribution yielding the error functions of several well known statistics". *Proceedings of the International Congress of Mathematics*, Toronto, *2*, 805-813.

Fisher, Ronald Aylmer (1925a) "Applications of Student's Distribution". *Metron*, *5*, 90-104.

Fisher, Ronald Aylmer (1925b) *Statistical Methods for Research Workers.* Edinburgh and London, Oliver & Boyd.

Fisher, Ronald Aylmer (1930) *Statistical Methods for Research Workers*. 3rd edition, Edinburgh and London, Oliver & Boyd.

Fisher, Ronald Aylmer (1954) *Statistical Methods for Research* Workers. 12th edition, New York, Hafner Publishing Company Inc..

Galilei, Galileo (1967) *Dialogue Concerning the Two chief World Systems-Ptolemaic & Copernican*. 2ª ed., Berkeley, University of California Press. S. Drake traductor.

Gauss, Carl Friedrich (1809) *Theoria Motus Corporum Celestian.* Hamburg, Friedrich Perthes and I. H. Besser.

Good, Irving John (1986) "Some Statistical Applications of Poisson's Work". *Statistical Science*, *1(2)*, 157-170.

Gosset, William Sealy (1908) "The probable Error of a Mean". *Biometrika, 6*, 1-25.

Grandi, Guidonis (1703) *Quadratura circuli, et hyperbolae per infinitas hyperbolas, et parabolas geometricè exhibita*. Pisa, Ex Typographia Francisci Bindi.

Grandi, Guidonis (1718) "Note al Trattato del Galileo del moto naturalmente accellerato". En: *Opere di Galileo Galilei*, *3*, 385-423, Firenze.

Haight, Frank A. (1967) *Handbook of the Poisson Distribution.* New York, John Wiley & Sons.

Hald, Anders (2003) *A History of Probability and Statistics and their Applications before 1750.* New Jersey, John Wiley & Sons.

Hartley, David (1749) *Observations on Man, His Frame, His Duty, and His Expectations.* London, S. Richardson.

Hotelling, Harold (1930) "British Statistics and Statisticians Today". *Journal of the American Statistical Association*, *25*, 186-190.

Jensen, Arne (1954). *A Distribution Model Applicable to Economics.* Copenhagen, Munksgaard.

Kennedy, Hubert C. (1972) "Who discovered Boyer's Law?". *The American Mathematical Monthly*, *79(1)*, 66-67.

Laplace, Pierre-Simon (1774) "Mémoire sur la probabilité des causes par les évènements". *Mémoires de l'Academie Royale des Sciences Presentés par Divers Savans, 6*, 621-656.

Laplace, Pierre-Simon (1777) "Recherches sur le milieu qu'il faut choisir entre les resultat de plusieurs observations". In: Charles Coulston Gilliespie. "Memoris inédits ou anonyous de Laplace on the théorie das des erreurs, les polynômes de Legendre et la philosophie des probabilités". Revue d'Hisotir des Sciences, 32 [1979], 228-256.

Laplace, Pierre-Simon (1810). "Mémoire sur les approximations des formules quit sont fonctions de très-grands nombres, et sur leur application aux applications probabilités". *Mémoires de la classe des sciences mathématiques et physiques de l'Institute de France, anné 1809,* 353-415, Supplément, 559-565.

Mahalanobis, Prasanta Chandra (1932) "Auxiliary tables for Fisher's *Z*-test in analysis of variance. (Statistical notes for agricultural workers, nº 3)". *Indian Journal of Agricultural Science*, *2*, 679-693.

Merton, Robert K. (1957) "Priorities in Scientific Discovery: A Chapter in the Sociology of Science". *American Sociological Review, 22(6)*, 635-659.

Merton, Robert K. (1968) "The Matthew Effect in Science". *Science, 159*, 56-63.

Merton, Robert K. (1973) *The Sociology of Science: Theoretical and Empirical Investigations.* Chicago and London, The University of Chicago Press.

Newbold, Ethel M. (1927) "Practical applications of the statistics of repeated events, particularly to industrial accidents". *Journal of the Royal Statistical Society, 90(3)*, 487-547.

Pearson, Karl (1926) "Abraham De Moivre". *Nature*, *117*, 551-552.

Phillips, Peter C.B. (1982) "The true characteristic function of the *F* distribution". *Biometrika,* 69(1), 261-264.

Poisson, Siméon Denis (1824) "Sur la probabilité des résultats moyens des observations". *Connaissance des Tems pour l'an 1827*, 273-302.

Poisson, Siméon Denis (1837) *Recherches sur la Probabilité des Jugements en Matière Criminelle et en Matière Civile Précédées des Règles Générales du Calcul des Probabilités.* Paris, Bachelier.

Rossiter, Margaret W. (1993) "The Matthew Matilda Effect in Science". *Social Studies of Science*, 23(2), 325-341.

Singer, Bernard (1979) *Distribution-Free Methods for Non-parametric Problems: A Classified and Selected Bibliography*. Leicester, British Psychological Society.

Snedecor, George Waddel (1934) *Calculation and Interpretation of Analysis of Variance and Covariance.* Ames, Iowa, Collegiate Press.

Snedecor, George Waddel (1938) *Statistical Methods.* Ames, Iowa, Collegiate Press.

Stahl, Saul (2006) "The Evolution of the Normal Distribution". *Mathematics magazine, 79(2)*, 96-113.

Stigler, Stephen M. (1980) "Stigler's Law of Eponymy". *Transactions of The New York Academy of Sciences*, *39(1)*, 147-157.

Stigler, Stephen M. (1982) "Poisson on the Poisson distribution". *Statistics and Probability Letters, 1*, 33-35.

Stigler, Stephen M. (1983) "Who Discovered Bayes's Theorem?". *The American Statistician*, *37(4)*, 290-296.

Stigler, Stephen M. (1986a) *The History of Statistics. The Measurement of Uncertainty before 1900.* Cambridge, The Belknap Press of Harvard University.

Stigler, Stephen M. (1986b) "Laplace's 1774 memoir on inverse probability". *Statistical Science*, *1(3)*, 359-378.

Stigler, Stephen M. (1999) *Statistics on the Table: The History of the Statistical Concepts and Methods.* London, Harvard University Press.

Stigler, Stephen M. (2013) "The True Title of Bayes´s Essay 1774". *Statistical Science*, *28(3)*, 283-288.

Stigler, Stephen M. (2016) *The Seven Pillars of Statistical Wisdom.* London, Harvard University Press.