II CUADERNOS DE TEORÍA DE LA PROBABILIDAD PARA MODELOS ECONOMÉTRICOS Y ACTUARIALES

# EL CONCEPTO DE ALEATORIEDAD EN LAS REPRESENTACIONES ECONOMÉTRICAS: DE BERNOULLI A WOLD

ALBERTO H. LANDRO

MIRTA L. GONZALEZ

Universidad de Buenos Aires
Facultad de Ciencias Económicas

ISBN: 978-950-29-1765-8

Universidad de Buenos Aires
Facultad de Ciencias Económicas

# CUADERNOS DE

# TEORIA DE LA PROBABILIDAD

# PARA MODELOS

# ECONOMÉTRICOS Y ACTUARIALES

**Alberto H. LANDRO** **Mirta L. GONZALEZ**

Centro de Investigaciones en Econometría
Instituto de Investigaciones en Administración, Contabilidad
y Métodos Cuantitativos para la Gestión – IADCOM
Facultad de Ciencias Económicas
Universidad de Buenos Aires

González, Mirta L.
El concepto de aleatoriedad en las representaciones econométricas : de Bernoulli a Wold / Mirta L. González ; Alberto Landro. - 1a ed . - Ciudad Autónoma de Buenos Aires : Universidad de Buenos Aires. Facultad de Ciencias Económicas, 2019.
Libro digital, PDF - (Cuadernos de Teoría de la Probabilidad para modelos econométricos y actuariales / González, Mirta L.; Landro, Alberto; 2)

Archivo Digital: descarga y online
ISBN 978-950-29-1765-8

1. Teoría de las Probabilidades. 2. Econometría. 3. Análisis de Riesgo. I. Landro, Alberto II. Título
CDD 332.04

ISBN 978-950-29-1765-8
9 789502 917658

Imagen de tapa: *"El Nombre de la Rosa"*, xilografía de Carmen Gilardi.

Hecho el depósito legal que establece la Ley 11.723 – Registro de la Propiedad Intelectual.

## INTRODUCCION A LA SERIE

El objetivo último de esta serie de Cuadernos sobre *Teoría de la Probabilidad para Modelos Econométricos y Actuariales* es el tratamiento de aquellos aspectos de la teoría de la probabilidad necesarios para "explicar" el comportamiento de fenómenos dinámicos de las ciencias económicas y actuariales. Fenómenos cuyo estado natural es de no-equilibrio. Un no-equilibrio constructivo que, como consecuencia de su propiedad fundamental de auto-organización, genera nuevos estados y nuevas estructuras infinitamente complejas en el ámbito de la irreversibilidad temporal y que, por tanto, permite asimilarlos a procesos estocásticos que evolucionan en el dominio del tiempo y cuya configuración varía en el dominio de los estados, definidos éstos por la realización simultánea de infinitas variables aleatorias. Esta interpretación de un proceso como una sucesión de variables aleatorias ordenadas en el tiempo implica, en consecuencia, la necesidad de analizar la naturaleza estocástica de sus características individuales, intentando una aproximación a las distribuciones de probabilidades que, sin resentir la rigurosidad del tratamiento de los complejos formalismos de su fundamento matemático, evite la distorsión de la esencia de los fenómenos analizados.

Mediante un riguroso tratamiento conceptual y aplicado se pretende no sólo ofrecer una fuente acabada de conocimiento para la enseñanza y aprendizaje sino brindar una herramienta integral para las aplicaciones profesionales en economía y finanzas, el análisis de riesgo y los problemas actuariales. Cada Cuaderno presentará un contenido preciso y autónomo de un capítulo de la teoría de la probabilidad, adecuado para su uso por profesores y alumnos de grado y posgrado, investigadores y profesionales de las ciencias económicas y actuariales. A todos les damos la bienvenida.

*Alberto Landro y Mirta González*

Universidad de Buenos Aires
Facultad de Ciencias Económicas

CUADERNO NÚMERO DOS

# EL CONCEPTO DE ALEATORIEDAD EN LAS REPRESENTACIONES ECONOMETRICAS: DE BERNOULLI A WOLD

**Alberto H. LANDRO** **Mirta L. GONZALEZ**

Centro de Investigaciones en Econometría
Instituto de Investigaciones en Administración, Contabilidad
y Métodos Cuantitativos para la Gestión – IADCOM
Facultad de Ciencias Económicas
Universidad de Buenos Aires

## INTRODUCCION AL SEGUNDO CUADERNO

El tratamiento del concepto de aleatoriedad en las representaciones econométricas resulta de particular interés para esta serie cuyo objetivo es “explicar” el comportamiento de fenómenos dinámicos en su asimilación a procesos estocásticos.

En una escueta reseña sobre el desarrollo de la teoría de la probabilidad y su influencia en la explicación del comportamiento de los fenómenos económicos, este Cuaderno comienza demostrando que el teorema de Bernoulli constituye el verdadero origen de la teoría de los modelos estocásticos dinámicos y avanza en el estudio de la evolución de los métodos econométricos, que evolucionaron desde una posición inicial determinística hacia el teorema de descomposición predictiva de Wold, fundamento de todo el tratamiento causal moderno de los fenómenos dinámicos.

Queda el lector invitado a compartir este apasionante viaje por los fundamentos metodológicos de las aplicaciones econométricas modernas.

**Agradecimientos**

Queremos agradecer a los grandes maestros de la historia por sus aportes a la solución del difícil intento de lograr la “explicación” de los fenómenos de las ciencias económicas y por habernos permitido tomar prestada su pluma para justificar las técnicas desarrolladas en este trabajo, cuyo objetivo es la transmisión de un sólido conocimiento sobre el cual puedan avanzar las nuevas generaciones.

*A.L y M.G*

*Para mi valiosa y querida mamá*
*M.G.*

**CONTENIDO**

## 1.- El problema de la inversión de la probabilidad

El "*Ars conjectandi*" de Jakob Bernoulli fue publicado en 1713[1] -en forma póstuma-, como recopilación de sus trabajos inéditos sobre teoría de la probabilidad de los 20 años anteriores -y por obra de su sobrino Nickolaus. Una obra que, por su tratamiento absolutamente novedoso de la evidencia empírica, modificó los fundamentos de la probabilidad en tal forma que debería ser considerada como el verdadero punto de partida de la teoría de la probabilidad y la culminación del proceso de formación del concepto de probabilidad.

Un proceso que tuvo su origen en el siglo XVII, en la obra de un grupo de pensadores (entre los que cabe mencionar a Joseph Glanvill, John Wilkins, Marin Mersenne, Pierre Gassendi, Hugo Grotius, John Tillotson, Robert Boyle y John Locke) quienes, a partir del deterioro del criterio de *"creencia"* en la religión, la filosofía y la ciencia, provocado fundamentalmente por las polémicas de la Reforma y la Contra-Reforma, generaron el movimiento filosófico conocido como "*escepticismo constructivo"*. Esta corriente coincidió con los escépticos en que la "*certeza absoluta"* se encontraba fuera del alcance de cualquier "*observador"*, pero admitió, sin embargo, que este observador poseía la habilidad de diseñar su comportamiento a un nivel inferior de "*certeza parcial"*. Este planteo se convirtió en el argumento habitual de la apologética de la teología natural, especialmente en Gran Bretaña[2] y tuvo su culminación en "*Analogy of religion, natural and revealed, to the constitution of nature"* (1736) de Butler[3].

---

[1] Si bien muchos de los desarrollos ya se encontraban en sus "*Meditationes"* (1684-1690) -según se puede concluir de su correspondencia con Leibniz-, J. Bernoulli dedicó a esta obra los dos últimos años de su vida, es decir que estaba fundamentalmente concluida en 1705. Después de su fallecimiento (a la edad de 50 años) sólo fue sometida a ligeras modificaciones pero, debido a disputas familiares, no fue publicada hasta 1713.

[2] Boyle ("*Some considerations about the reconcileableness of reason and religión"* (1675)) reconoció tres grados de certeza: "*metafísica*" (absoluta), "*física*" y "*moral*" (parciales). Consideró a la primera como derivada de axiomas que "*...nunca pueden ser sino ciertos"*, a la "*certeza física"* como "*...deducida de principios físicos"* que podían ser modificados o eliminados por Dios, y a la "*certidumbre moral"* (estado intermedio entre la certeza y la duda) como aquélla en la que "*...la conclusión se funda en alguna concurrencia de probabilidades"* (p. 182). Entendiendo por "*concurrencia de probabilidades"* no probabilidades matemáticas, sino la "*...convicción adicional que proporciona la evidencia convergente"* (Daston (1980, p. 63)).

[3] En esta obra, en la que Butler incorporó además un tipo de razonamiento inductivo y analógico concebido en términos frecuencistas, aparece por primera vez el "ejemplo del amanecer". En el que la asignación de una probabilidad a la ocurrencia del fenómeno "que mañana amanezca" sólo es posible a partir de un gran número de observaciones. Este ejemplo, utilizado luego por Richard Price en sus comentarios sobre el teorema de Bayes, se convirtió en un lugar común en la literatura sobre la inversión de la probabilidad (ver Sec. 10).

La racionalidad pragmática de esta "*certeza parcial*" -ubicada entre la "*certeza dogmática*" de los escolásticos y los cartesianos, y la "*duda absoluta*" de los escépticos- le permitía al observador definir "*grados de certeza*" cuya cuantificación dio origen a las probabilidades.

Este intento de cuantificación, que estuvo a cargo de matemáticos como Pascal y Fermat, se estructuró a partir de: **i)** la posibilidad de la justificación de analogías entre el comportamiento de los fenómenos y los modelos matemáticos disponibles, y **ii)** la hipótesis metafísica de que dicho comportamiento era constante y ordenado. Los filósofos racionalistas del siglo XVII[4] procuraron, de la misma forma que los teoremas eran deducidos de los axiomas y definiciones de la geometría, deducir el comportamiento de los fenómenos naturales -mediante un método de síntesis- a partir de "*leyes*" que se suponía constituían la estructura de la materia. Pero, en muchos casos la limitación de los sentidos, la falibilidad de la razón y la infinita complejidad de la naturaleza los obligaron a emplear un esquema inverso (analítico) de razonamiento empirista: desde los "*efectos evidentes*" hacia las "*causas ocultas*". Ahora bien, dado que este método empírico, basado en un conjunto limitado de observaciones, no permitía generalizaciones a cualquier otro caso, las previsiones que se obtuvieron resultaron inevitablemente inciertas. Fueron en principio, los matemáticos del siglo XVIII quienes, mediante la cuantificación de las probabilidades de las hipótesis, intentaron dar mayor precisión a estos métodos de predicción y ampliar los alcances de la teoría de la probabilidad a la inferencia empírica[5].

A comienzos de este siglo también los filósofos empiristas trasladaron su interés de los aspectos cualitativos a los aspectos cuantitativos de la experiencia. Todo este proceso produjo como resultado una "*teoría de la inferencia inductiva*" aplicable a series de eventos que pudieran ser considerados naturalmente idénticos, para los cuales la agregación de la experiencia adquirida a partir de su observación repetida, se transformaba en formas de "*expectativa*" acerca de su comportamiento futuro (expectativa entendida como la expresión matemática del concepto de racionalidad)[6].

[4] Por ejemplo, Descartes ("*Principia philosophia*" (1644)).

[5] Para una cuidadosa reseña sobre la probabilidad previa a Pascal y un análisis detallado de su obra probabilística y su correspondencia con Fermat, vert Franklin (2001), Shea (2003), Hammond (2003).

[6] La primera versión de este tratamiento cuantitativo de la "expectativa" se halla en "*La logique, ou l'art de penser*" (1695) de Arnauld y Nicole, obra que constituye el nexo entre la tradición cualitativa de los retóricos del excepticismo constructivo y J. Bernoulli (es más, el "*Ars conjectandi*" puede ser considerado como un complemento de "*La logique*", conocida en latín como el "*Ars cogitandi*"). Posteriormente Pascal en sus "*Pensées*" (1669), utilizando el argumento hoy conocido como la "*apuesta de Pascal*", interpretó a la racionalidad como la maximización de

En este sentido, la contribución fundamental de J. Bernoulli -incluida en la "*Pars quarta, Fradens usum et applicationem procedentis doctrinæ in civilibus, moralibus et œconomicis"* del "*Ars conjectandi"*- consistió en: **i)** proporcionar la primera definición estricta de probabilidad ("*probabilitas*") en un contexto matemático como "*...grado de certeza (gradus certitudinis), que difiere de ésta de la misma forma que una parte con respecto al todo"*[7] y, en consecuencia, a partir del tratamiento cualitativo formal de la idea planteada por los mencionados apologistas ingleses de la teología natural del siglo XVII y por Arnauld y Nicole, los lógicos de Port Royal, postular que se podía "*aprender"* a partir de la experiencia, que cuanto mayor fuera la cantidad de información con que contara el observador más aproximada sería la estimación de su verdadero valor; **ii)** demostrar que este aprendizaje era cuantificable a través de un proceso de inversión de la probabilidad[8] y, en consecuencia, asumiendo ciertas hipótesis de simplicidad y regularidad, **iii)** establecer el nexo entre las probabilidades "a priori" (definidas a partir de un razonamiento que va de las causas a los efectos, de la simetría de los resultados posibles al concepto de equiprobabilidad) y las probabilidades "a posteriori" (definidas a partir de un razonamiento que va de los efectos a las causas, de las estadísticas sobre la mortalidad introducidas por Graunt a la probabilidad de fallecimiento)[9].

Esta demostración del principio intuitivo que la incertidumbre disminuía en la medida que se incrementaba el número de observaciones, y la cuantificación de

---

la expectativa en condiciones de incertidumbre.

[7] En *"Die Werke von Jakob Bernoulli"*, vol. 3 (p. 239).

[8] Si bien figura en algunos cursos dictados en la École Polytechnique a fines del siglo XVIII y principios del siglo XIX (presuntamente por Jean-Baptiste J. Fourier o Jean G. Garnier), la expresión *"probabilidad inversa"* es atribuida (de acuerdo con Arne Fisher (1926)) a Augustus de Morgan (*"An essay on probabilities and their application to life contingencies and insurance offices"* (1838)).

[9] Como se verá en la sección siguiente, para la elaboración de este nexo entre probabilidades "a priori" y "a posteriori" (según surge de la correspondencia intercambiada con Gottfried W. Leibniz entre octubre de 1703 y abril de 1704), Bernoulli utilizó un modelo elaborado a partir de una probabilidad "a posteriori" calculada en base a extracciones "con reposición" de una urna con bolillas de diferentes colores que representaban enfermedades terminales. Las hipótesis sobre las cuales basó su razonamiento fueron: **i)** que, de la misma forma que la urna contenía un gran número de bolillas en una proporción estable, "*...las personas estaban afectadas por un gran número de enfermedades*" y **ii)** que la "*naturaleza observa los comportamientos más simples"* (en Gerhardt (ed.)(1962), vol. 3, p. 77)).

dicho proceso de inferencia -conocida luego como la "*ley (débil) de los grandes números"*-, constituyó el primer teorema límite de la teoría de la probabilidad[10].

Hasta la aparición del "*Ars conjectandi"* los avances producidos en la teoría de la probabilidad no habían conseguido proporcionar una respuesta eficaz a la formalización de este proceso de inferencia. Como se mencionó en la sección precedente, los principales tratados de los autores clásicos -empleando el método de razonamiento desde las "*causas"* a los "*efectos"*- se referían exclusivamente a la resolución de problemas del tipo: dada una urna que se sabe que contiene $a$ bolillas rojas y $c$ bolillas azules, la probabilidad de obtener una bolilla roja al realizar una extracción al azar es $\theta = \frac{a}{a+c}$. J. Bernoulli fue el primero en tratar el esquema empírico inverso: la estimación de los valores $a$ y $c$, basándose en la evidencia que proporcionaban los resultados de las sucesivas extracciones. Fue el primero en advertir a los probabilistas sobre "*...la relación entre la 'conjectandum' probabilística y la inferencia inductiva"* (Daston (1988, p. 228)), en sustituir el concepto clásico (deductivo) de probabilidad "a priori", basado en la simetría de los resultados posibles y en el concepto Humeano de equiprobabilidad[11] apropiado solamente para resolver problemas relacionados

[10] La denominación de "*ley de los grandes números*" referida a una generalización de este teorema referida al comportamiento de los promedios es muy posterior y se debe a Poisson (1837). La forma más rudimentaria de la ley de los grandes números es atribuible a Girolamo Cardano ("*Practica aritmeticæ et mensurandi singularis"* (1539)), quien demostró que el número de ocurrencias de un resultado, en una "*larga serie"* de repeticiones independientes de un fenómeno de comportamiento eventual ($X$), es aproximadamente igual a $n\theta$ (donde $\theta$ denota la probabilidad "a priori" de ocurrencia de dicho resultado en una repetición determinada). Cardano –quien puede ser considerado como el primer gran aleatorista- consideró que la aproximación obtenida al cabo de infinitas repeticiones estaba sujeta a las "*perturbaciones de la fortuna"* (la "*suerte"* en su nomenclatura). Su metafísica podría ser resumida en la siguiente expresión, debida a Daston (1988): "*...el cálculo no puede gobernar lo contingente"* (p. 36). Una idea similar sobre el tratamiento cualitativo del aprendizaje a partir de la experiencia puede hallarse en Halley ("*An estimate the degrees of mortality of mankind drawn from the curious tables of the births and funerals at the city of Breslau; with an attempt to ascertain the price of annuities upon lives"* (1693)), en "*De incerti æstimatione"* de Leibniz (1703-1705) y, comose viomásarriba, en "*La logique, ou l'art de penser"* de Arnauld-Nicole.

[11] Hume (1739): "*Dado que una indiferencia total es esencial a la aleatoriedad, no es posible suponer que unaaleatoriedad pueda ser superior a otra, en su defecto puede estar compuesta por un número de aleatoriedades iguales. Afirmar que, de alguna forma, una aleatoriedad puede ser superior a otra implica reconocer que existe algo que origina esa superioridad. Es decir, implica suponer la existencia de una causa y, por lo tanto, la negación de la hipótesis de aleatoriedad. De modo que una indiferencia perfecta y total es esencial a la aleatoriedad y una indiferencia total nunca puede ser intrínsecamente superior o inferior a otra. Esta verdad no es inherente exclusivamente a mi sistema, sino que es reconocida por todos los probabilistas"* (p. 125). Los números de página que figuran en las referencias a "*A treatise of human nature*" de Hume

con juegos de azar, por la idea de probabilidad "a posteriori", definida como una medida del conocimiento (la expectativa) que posee el observador acerca de la veracidad de una proposición[12].

Utilizando notación moderna, el resultado obtenido por Bernoulli puede ser expresado de la siguiente forma: Sea $Y_n = \frac{X}{n}$ la frecuencia relativa correspondiente al resultado "bolilla roja", obtenida al cabo de una sucesión de $n$ extracciones al azar "con reposición" de una urna cuya composición -desconocida para el observador- es de $a$ bolillas rojas y $c$ bolillas azules[13]. Entonces, dados un valor $\varepsilon$positivo y arbitrariamente pequeño, y un valor $t$ positivo y arbitrariamente grande, se demuestra que es posible hallar un $n > n(\theta, \varepsilon, t)$ tal que se puede asegurar que, con una probabilidad mayor a $\frac{t^2-1}{t^2}$, la frecuencia relativa del resultado "bolilla roja" se encontrará a una distancia menor o igual que ε del verdadero valor de la probabilidad binomial $\theta = \frac{a}{a+c}$:

---

corresponden a la versión inglesa editada por Clarendon Press (1978).

[12] J. Bernoulli ("*Ars conjectandi"*): "*Hemos arribado al punto en el que parece que, para realizar una predicción correcta acerca de un evento cualquiera, solamente es necesario calcular exactamente el número de casos posibles y, luego, determinar cuánto más probable es que ocurra un resultado que otro. Pero aquí surge nuestra dificultad principal, ya que este procedimiento es aplicable solamente a un número reducido de fenómenos, casi exclusivamente a aquellos relacionados con los juegos de azar. Los creadores de estos juegos los diseñaron fijando el número de casos que resultarían en ganancia o pérdida y considerándolos conocidos con anticipación y, también, combinando los casos para que todos fueran igualmente probables, de modo que todos los jugadores tuvieran las mismas chances de ganar. Pero ésta no es, de ninguna manera, la situación de la mayoría de los fenómenos gobernados por las leyes de la naturaleza o la voluntad del hombre (...) Ahora bien, me pregunto, qué mortal podría determinar, contando todos los casos posibles, el número de enfermedades que aflijen al cuerpo humano en cada una de sus partes y a cada edad, y decir cuánto mayor que otra es la probabilidad de una enfermedad de ser fatal. Los resultados dependen, en estos casos, de factores que son completamente oscuros y que, por la infinita complejidad de sus interrelaciones, constantemente engañan a nuestros sentidos. Existe, sin embargo, otro camino que nos conducirá a lo que estamos buscando y nos permitirá conocer, al menos 'a posteriori', lo que no podemos determinar 'a priori', es decir, adquirir conocimiento a partir de los resultados observados en numerosas situaciones similares. Debe suponerse en esta circunstancia que, en condiciones similares, la ocurrencia (o no ocurrencia) de un evento en el futuro observará el mismo comportamiento que fue observado para eventos similares, en el pasado"* (p. 226).

[13] $Y_n = \frac{X}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i$ denota una sucesión de sumas parciales. Donde $X_i(1,2,\dots,n)$ define una sucesión de variables aleatorias binomiales que asumirán los valores 1 ó 0 -con probabilidades $\theta$ y $1-\theta$, respectivamente ($X_i\colon b(1,\theta)$)- según que el resultado de la $i$-ésima extracción fuere "bolilla roja" o "bolilla azul", respectivamente.

$$p(|X-n\theta|\leq\varepsilon)=\sum\nolimits_{|X-n\theta|\leq\varepsilon} b(X_i,1,\theta)=p\left(\left|\frac{X}{n}-\theta\right|\leq\varepsilon\right)>\frac{t^2-1}{t^2}$$

A fin de simplificar la demostración de este teorema, e intentando distorsionar lo menos posible la interpretación original de Bernoulli en lo que hace a los alcances de los resultados obtenidos[14], se utilizará aquí la relación conocida como desigualdad de Bienaymé-Chebychev[15]:

$$p[|X-E(X)|\leq t\sigma(X)]\geq 1-\frac{1}{t^2}$$

Aplicando dicha desigualdad a la variable $Y_n$ se obtiene que[16]:

$$p\left(|Y_n-\theta|\leq t\sqrt{\frac{\theta(1-\theta)}{n}}\right)>1-\frac{1}{t^2}$$

y, haciendo $t\sqrt{\frac{\theta(1-\theta)}{n}}=\varepsilon$[17], será:

[14] Para quien desee recrear el escenario en el que se desenvolvía el devenir académico de los probabilistas del siglo XVII, una versión de la demostración original de J. Bernoulli puede ser hallada en Hald (1984).

[15] Bienaymé (1855), Chebychev (1867). Este resultado depende exclusivamente delvalor esperado y la varianza de una variable binomial. En su expresión más general asume la forma $p(|X|\geq\varepsilon)\leq\frac{E(|X|^s)}{\varepsilon^s}(\varepsilon>0,s>0)$. Debe tenerse en cuenta que en la época de la demostración de Bernoulli el concepto de varianza era desconocido.

[16] De acuerdo a las hipótesis enunciadas, $Y_n$ es una variable binomial con distribución de probabilidades de la forma $p(Y_n=y_n)=p\left(\frac{X}{n}=\frac{x}{n}\right)=p\big(X_{(n)}=x\big)=\binom{n}{x}\theta^x(1-\theta)^{n-x}(x=0,1,2,\dots,n)$ y medidas $E(Y_n)=E\left(\frac{X}{n}\right)=\theta$ y $\sigma^2\left(\frac{X}{n}\right)=\frac{\theta(1-\theta)}{n}$.

[17] En la versión original de Bernoulli $\varepsilon$ no está definido como cualquier número positivo arbitrario, sino que asume un valor múltiplo de $(a+c)$. El supuesto básico de la demostración original es que $n=k(a+c)$; de modo que $n\theta=ak$ y $n\varepsilon=k$ sean enteros, única forma posible de obtener la igualdad $\frac{X}{n}=\frac{a}{a+c}$. Esta restricción no significó pérdida de rigurosidad en la demostración original, la cual admite una formulación general posterior.

$$p(|Y_n - \theta| \leq \varepsilon) > 1 - \frac{\theta(1-\theta)}{n\varepsilon^2}$$

Entonces se puede escribir $\theta(1-\theta) = \left(\frac{1}{2} - k\right)\left(\frac{1}{2} + k\right) = \frac{1}{4} - k^2$. De esta forma, el producto $\theta(1-\theta)$ asume su valor máximo cuando $\theta = \frac{1}{2}$. Lo que prueba que la probabilidad que figura en el primer miembro está uniformemente acotada con respecto a $\theta$y a $(1-\theta)$:

$$p(|Y_n - \theta| \leq \varepsilon) > 1 - \frac{1}{4n\varepsilon^2}$$

de lo que se concluye que $\lim_{n\to\infty} p(|Y_n - \theta| \leq \varepsilon) = 1$[18]. Es decir que, cuando $n \to \infty$, el estimador $Y_n$ converge en-probabilidad al valor $\theta$ ($Y_n \xrightarrow{p} \theta$)[19]. De modo que, conocidos $n$ e $Y_n$, se puede resolver la ecuación $n(Y_n, \varepsilon, t) = n$ con respecto a $t$, obteniendo así una aproximación al límite inferior $\frac{t^2-1}{t^2}$ (cota inferior de la "*incertidumbre residual*") correspondiente a la probabilidad de ocurrencia del evento $|Y_n - \theta| \leq \varepsilon$ y, a partir de esta expresión, determinar la probabilidad de que $\theta$ esté incluida en un intervalo de la forma $[Y_n - \varepsilon, Y_n + \varepsilon]$, donde $\theta$ admite cualquier representación no frecuencista.

---

18 De acuerdo con la propia terminología de Bernoulli, queda demostrado que es "*moralmente cierto*" (ver Sec. 13) que, para $n$ suficientemente grande, $\frac{X}{n}$ no se desviará en gran medida de $\theta$. En otros términos, establece lo que hoy se conoce como el *"principio de estabilidad de las frecuencias estadísticas"*(von Mises (1928)): si al cabo de $n$ observaciones (para $n$ suficientemente grande) la frecuencia relativa del resultado "bolilla roja" ha sido $Y_n$, existe un alto grado de probabilidad de que, al cabo de una serie (suficientemente larga) de $n'$ observaciones, la frecuencia relativa del resultado "bolilla roja" sea $Y_{n'} \approx Y_n$. Según lo expresado en sus "*Meditationes*", Bernoulli concluyó esta demostración entre 1688 y 1690. Ver Kolmogorov (1956).

19 El concepto de convergencia –que originalmente se aplicó a sucesiones de números y oportunamente se extendió a sucesiones de funciones- es aplicable a sucesiones de infinitas variables aleatorias, $\{Y_n\}$ (que son sucesiones de funciones, $Y_n = Y_n(w)$, definidas en un espacio $\Omega$). Esta convergencia –denominada *"convergencia estocástica"*- si bien posee ciertas características similares a la convergencia en el sentido del análisis, se verifica a través del operador probabilidad. Esta intervención de la probabilidad caracteriza diferentes definiciones de convergencia estocástica. En particular, se dice que una sucesión de variables aleatorias $\{Y_n\}$ converge "*en-probabilidad*" a una variable aleatoria $Y$, si se verifica que $\lim_{n\to\infty} p(|Y_n - Y| > \varepsilon) = 0\ (\varepsilon > 0)$ (esta condición suele ser denotada como $plim_{n\to\infty} Y_n = Y$). Como corolario se demuestra que, si $Y_n \xrightarrow{p} Y$ entonces, dada una función continua, $g$, se verifica que $g(Y_n) \xrightarrow{p} g(Y)$.

Leibniz planteó varias críticas a esta ley débil de los grandes números (interpretada como un teorema de la teoría de la estimación) que permiten reflexionar acerca de la naturaleza indiscutiblemente metafísica de los supuestos que condujeron a Bernoulli a postular la asociación entre la interpretación clásica de la probabilidad y la filosofía de la inducción. Una se refiere a la condición necesaria (puramente virtual) de contar con un número infinito de observaciones para identificar la probabilidad $\theta$ y la imposibilidad de demostrar la convergencia de una sucesión de frecuencias relativas a partir de un conjunto finito de observaciones (como se mencionó más arriba, la idealización que implica la generación de una sucesión infinita de observaciones admite sólo una convergencia estocástica basada en condiciones inherentes a la lógica de la incertidumbre)[20].

Una segunda objeción considera que, dado que dichas observaciones son asimilables a un conjunto finito de puntos, que los "*resultados ocultos*" son asimilables a una curva que pasa por entre esos puntos y que, para cada conjunto de puntos existen infinitas curvas que cumplen con dicha condición, la probabilidad "a posteriori" obtenida por inferencia es sólo una de las infinitas probabilidades posibles incluidas en el intervalo definido por el teorema.

Una tercera cuestión está relacionada con la exagerada simplificación que implicaba la utilización del esquema de urnas y con el mecanismo combinatorio que suponía la independencia estocástica de los eventos sucesivos.

Sin proponer una metodología general para valores finitos de $n$, la respuesta de Bernoulli con respecto a las primeras cuestiones se basó en el ambiguo supuesto de uniformidad y simplicidad de la naturaleza[21], según el cual la estimación de $\theta$ debía realizarse mediante la selección de la curva más simple[22], de modo que el problema de la inversión quedaba reducido exclusivamente a la selección del estimador más simple de la probabilidad "a priori"[23].

---

[20] Como se verá más adelante, este problema de la convergencia fue tratado rigurosamente, entre otros, por von Mises, Faber, Hausdorff, Cantelli, Kolmogorov, Prokhorov, Khinchin, Doob.

[21] El supuesto pitagórico según el cual Dios había diseñado un universo racional, de acuerdo a principios matemáticos simples.

[22] "*La determinación de una trayectoria a partir de un conjunto finito de observaciones (...) sería bastante débil e incierta si no se tuviera en cuenta que la curva a seleccionar pertenece a la clase de las curvas simples (...) ya que podemos observar que la naturaleza obedece a los comportamientos más simples*" (ver correspondencia de Bernoulli a Leibniz en Gerhardt (ed.)(1962, vol. 3, p. 11)).

[23] Como se verá en la Sec. 2, en su intento de solución del problema de la inversión, de Moivre ("*Approximatio ad summam terminorum binomii $(a+b)^n$ in seriem expansi*" (1733)) reemplazó este

Con respecto al esquema de urnas, debe tenerse en cuenta que, contrariamente a lo que ocurrió con los probabilistas clásicos –que trataron a las loterías, los juegos de dados y las tiradas independientes de una moneda a nivel de aplicaciones prácticas exclusivamente-, como se vio en la sección precedente, Bernoulli utilizó este modelo como la metáfora más adecuada para reproducir los postulados de la filosofía natural de fines del siglo XVII e intentar demostrar una vinculación entre las causas ($\theta$) y los efectos observados ($Y_n$) según la cual combinaciones aleatorias de causas inaccesibles -ocultas para el observador- podían, a partir de sucesiones infinitas de observaciones repetidas en igualdad de condiciones, producir eventualmente efectos regulares, generando una propuesta alternativa al modelo de causación racionalista, el cual, para un observador ideal capaz de acceder a un conjunto de información de tamaño infinito, la relación de causación era esencialmente deductiva y, en consecuencia, no-combinatoria[24].

Fue su postura (más teológica y filosófica que matemática) de convencido militante del determinismo metafísico la que condujo a Bernoulli a la identificación de las causas ignoradas del comportamiento de los fenómenos con el parámetro $\theta$, determinado e invariable y representativo de la naturaleza gobernada por leyes inmutables y, en consecuencia, a una interpretación de su teorema que consideraba a la probabilidad $\theta$ como conocida, limitando sus alcances a la demostración de la convergencia en-probabilidad de la frecuencia relativa $Y_n$ a $\theta$ y, en segundo lugar, a la rigurosa determinación del número de observaciones$n(\theta, \varepsilon, t)$ que se requiere para que la probabilidad de que el valor de $Y_n$ esté incluido en un intervalo $[\theta - \varepsilon, \theta + \varepsilon]$ sea $t^2$ veces mayor que la probabilidad de que no lo esté[25]. Según el ejemplo desarrollado por el mismo

---

supuestode simplicidad por el de "*Orden*" resultante de un "*Diseño Original*" en el comportamiento de los fenómenos naturales. Las distintas interpretaciones matemáticas de ese "*Orden*" generaron una serie de controversias que constituyen uno de los capítulos más interesantes del desarrollo de la teoría de la probabilidad durante el siglo XVIII.

[24] Excepto para los filósofos atomistas como Charleton ("*Physologia epicuro-gassendo-charletoniana"* (1654)), Gassendi ("*Syntagma"* (1659)) y Boyle ("*Origin of forms and qualities"* (1666)), esta concepción debe haber resultado extraña a los científicos de los siglos XVII, XVIII y comienzos del XIX. Recién a fines de este último siglo, a través de los trabajos de Maxwell, Boltzman y Clausius, entre otros, esta "*explicación estadística"* comenzó a ser aceptada por los físicos. Este concepto de "*causas"* determinadas y a la vez combinatorias, regulares y a la vez estocásticamente independientes, tiene su precedente, precisamente, en la teoría corpuscular de Charleton-Gassendi-Boyle.

[25] J. Bernoulli ("*Ars conjectandi"*): "*Este tipo de predicción requiere 'un gran número de observaciones' (...) pero, si bien todos reconocemos esto, la demostración científica de este principio no es simple (...) Existe, además, otro problema a ser tomado en consideración (...) podría ocurrir que al incrementar el número de observaciones también se incrementara la probabilidad de que la proporción registrada entre resultados favorables y no-favorables se*

Bernoulli en el "*Ars conjectandi"*, para una urna que contiene 30 bolillas rojas y 20 azules, un $\varepsilon = \frac{1}{50}$ y $t^2 = 1000$, se necesitarían $n \geq 25.550$ observaciones; es decir, a partir de la información proporcionada por 25.550 observaciones de los resultados del juego (extracciones "con reposición"), se podría asegurar que $p\left(\frac{29}{50} \leq \frac{X}{25.550} \leq \frac{31}{50}\right) > \frac{1000}{1001}$. De la misma forma demostró que, para $t^2 = 10.000$, se necesitarían $n \geq 31.258$ observaciones y para $t^2 = 100.000$, se necesitarían $n \geq 36.966$ observaciones. Estos resultados -cuyas magnitudes resultaban desproporcionadas para la época- indudablemente deben haber sorprendido -y decepcionado- a Bernoulli -y a sus contemporáneos-, y quizás haya sido la sospecha de la existencia de algún error en el tratamiento matemático del problema -y no razones de orden filosófico- la causa fundamental de su negativa a publicar los resultados de sus investigaciones de 20 años[26].

A partir de esta demostración Bernoulli intentó una extensión de su teorema que implicaba un planteo inverso de la forma: si se observa que la frecuencia relativa "*...converge a un valor determinado*", $\theta$, entonces este valor definirá la "*ley*" que gobierna a dicho evento. Pero la circularidad de este esquema de razonamiento en el cual la convergencia en-probabilidad de las frecuencias relativas se verificaba porque los eventos estaban regidos por una ley determinada pero, a su vez, la convicción de que los eventos se regían por una ley determinada se fundaba en el postulado de inversión de la probabilidad según el cual las

---

*aproximara a un 'verdadero cociente', de modo que dicha probabilidad finalmente superara cualquier grado deseado de certeza, o podría ocurrir que el problema presentara una asíntota, lo cual implicaría la existencia de un grado de certeza particular de que el 'verdadero cociente' habría sido hallado, que no podría ser superado no importa cuánto se aumentara el número de observaciones. Para que esta cuestión no sea interpretada en forma imperfecta, debe tenerse en cuenta que el cociente que representa la verdadera relación entre los números de casos, nunca puede ser obtenido con absoluta seguridad (...) El cociente al cual arribamos es sólo aproximado: deben definirse dos límites, pero esos límites pueden ser construidos de modo que se aproximen tanto como deseemos"* (p. 225).

[26] Si bien el objetivo de J. Bernoulli fue el desarrollo de una teoría general de la decisión racional en condiciones de incertidumbre, de los tres primeros capítulos de la Parte IV del "*Ars conjectandi"* se puede concluir en forma inmediata que los conceptos y problemas tratados están íntimamente relacionados con la necesidad de incrementar la importancia relativa de las consideraciones objetivas exógenas en la práctica legal de los contratos aleatorios (un movimiento que se originó entre los juristas del siglo XII en Europa continental, y que tropezó inevitablemente con el problema de la cuantificación). Estos intentos de aplicar la teoría de la probabilidad a cuestiones legales en general, y a la credibilidad de los testimonios en un juicio en particular comenzó, en realidad, con Craig ("*Theologiæ christianæ principia mathematica"*, 1699) y, pasando fundamentalmente por N. Bernoulli ("*De usu artis conjectandi in jure"*, 1709) y J. Bernoulli ("*Ars conjectandi"*, 1713), continuaron hasta Poisson ("*Recherches sur la probabilité des jugements en matière criminelle et en matière civile"*, 1837). Ver Stigler (1986).

frecuencias relativas debían converger a $\theta$, no le permitió obtener una justificación rigurosa de los alcances de su teorema como fundamento de una teoría de la inversión de la probabilidad[27].

J. Bernoulli concluyó su tratado con el siguiente postulado, posteriormente adoptado por muchos científicos (incluyendo a Laplace) como el fundamento de la filosofía determinística[28]: "*Si algo que está destinado a ocurrir no es cierto que ocurra, no resulta claro entonces cómo puede permanecer inquebrantable el elogio de la omnisciencia y omnipotencia del Todopoderoso. Si todos los eventos fueran observados en forma continua, desde ahora y por toda la eternidad (con lo cual la probabilidad se transformaría en certeza), se concluiría que en el mundo todo ocurre por razones definidas y de conformidad con una Ley y que, por lo tanto, estamos obligados, aún en casos que parecen ser accidentales, a suponer una cierta necesidad o fatalismo (...) Dadas su posición, su velocidad y su distancia del tablero, cuando el dado abandona la mano del jugador, indudablemente, no puede caer de una forma distinta de la que, en realidad, cae. En forma similar, para una composición dada del aire y para masas, posiciones, direcciones y velocidad de los vientos, vapor y nubes dados, y dadas, también, las leyes de la mecánica que gobiernan todas estas interacciones, el clima correspondiente al día de mañana no podrá ser diferente del que debería ser en realidad. De modo que estos fenómenos se comportan con una regularidad no menor (non minus necessario) que la de los eclipses de los cuerpos celestes. Es, sin embargo, una práctica usual considerar a un eclipse como un evento regular, mientras que a la caída de un dado o al clima del día de mañana se los considera como eventos aleatorios. La razón de esto radica exclusivamente en el conocimiento insuficiente de las acciones sucesivas en la naturaleza. Y, aún en el caso en que éstas fueran conocidas, nuestro saber en matemáticas y física no está suficientemente desarrollado, por lo que, partiendo de las causas iniciales, no podemos calcular estos fenómenos, en tanto que, a partir de los principios absolutos de la astronomía, es posible precalcular y predecir los eclipses (...) La aleatoriedad depende, fundamentalmente, de nuestro conocimiento (...) el arte de la predicción está definido aquí como el arte de medir las probabilidades de los eventos tan precisamente como sea posible, a fin de lograr que nuestras decisiones y*

[27] Paradójicamente, poco después de su publicación el "*Ars conjectandi"* -que logró el inmenso avance que implica trasladar el eje central del razonamiento de las expectativas a las probabilidades, es decir de los resultados posibles equiprobables a las medidas de incertidumbre- desapareció como instrumento para el estudio de la teoría matemática de la probabilidad y fue sustituido en el ámbito académico por la segunda edición del "*Essai d'analyse sur les jeux de hazard*" (1713) de Pierre de Montmort.

[28] Laplace (1814), si bien coincidió con la posición de Bernoulli, reemplazó al "*Todopoderoso*" por "*...una inteligencia capaz de comprender a todas las fuerzas de la naturaleza*" (p. xi).

*acciones sean las mejores, las más satisfactorias, fáciles y razonables. En esto consiste la sabiduría del filósofo y la prudencia del estadista*" (p. 237)[29].

## 2.- Las generalizaciones de la solución de Bernoulli

En "*Usu artis conjectandi in jure*" (1713), con motivo de la discusión que se había generado acerca de la estabilidad del cociente entre el número de nacimientos masculinos y femeninos en Londres, Nikolaus Bernoulli, a partir fundamentalmente de los resultados obtenidos por Lodewijk Huygens y John Arbuthnot y, en menor medida, por su tío Jakob, logró una demostración de la ley de los grandes números sin restricciones sobre los valores a asumir por $\varepsilon$, suponiendo solamente que debía ser pequeño en comparación con $n$.

En realidad, la primera mención del problema de la definición de un modelo para estimar la proporción de nacimientos masculinos y femeninos se debe a John Graunt ("*Natural and political observations made upon the bills of mortality"* (1662)), quien postuló que la proporción de bautismos masculinos en Londres, era mayor que 1/2 y que la misma no variaba mucho en el tiempo. En 1712, Arbuthnott ("*An argument for Divine Providence, taken from the constant regularity observed in the births of both sexes"*) sugirió que la variación del número de bautismos masculinos podía ser considerada como una variable binomial e intentó diseñar un test (generalizado luego por Willem 'sGravesande, "*Démonstration mathématique du soin que Dieu prend de diriger ce qui se passé dans ce monde, tirée du nombre des garçons et des filles qui naissent journellement"* (1712)[30]) para contrastar la hipótesis de que la probabilidad de que se produjera un nacimiento masculino ($\theta$) era igual a $\frac{1}{2}$, respecto de la alternativa $\theta \neq \frac{1}{2}$, demostrando que, para cualquier número de nacimientos, se verificaba que $p\left[M > F/\left(\theta = \frac{1}{2}\right)\right] \leq \frac{1}{2}$ (donde $M$ y $F$ denotan el número de nacimientos masculinos y femeninos, respectivamente) y luego utilizó este resultado para transformar la hipótesis original en la hipótesis $p(M < F) \leq \frac{1}{2}$, y la alternativa $p(M < F) > \frac{1}{2}$[31].

---

[29] En este punto el "*Ars conjectandi"* concluye abruptamente, sin hacer mención alguna a las aplicaciones de la teoría de la probabilidad a problemas civiles y económicos prometidas en el título de la "*Pars Quarta"*.

[30] Esta edición fue de circulación reducida. Los principales resultados de este trabajo fueron publicados en 1715, por el médico holandés Bernard Nieuwentyt bajo el título "*Het regt gebruik der Wereldbeschouwingen"*.

[31] A partir de estos resultados William Derham publicó "*Physico-theology, or a demonstration of*

La intención de N. Bernoulli fue comparar la distribución de las observaciones con la función de probabilidades binomial, con el objeto de determinar si la variación observada podía ser explicada por dicho esquema (comparó los datos con una distribución binomial con parámetro $\theta = \frac{18}{35}$, y aceptó, equivocadamente, la hipótesis que, para el período al cual corresponden las observaciones, la variación era binomial) y, sobre todo, de lograr una herramienta matemática que permitiera una aproximación a la probabilidad:

$$p(|Y_n - \theta| \le \varepsilon) = p(|X - n\theta| \le \varepsilon^*) = \sum_{|X-n\theta|\le\varepsilon^*} \theta^x (1-\theta)^{n-x} =$$
$$= \sum_{|X-n\theta|\le\varepsilon^*} b(x,n,\theta)$$

que evitara las enormes complicaciones que se presentaban en el cálculo de los coeficientes binomiales cuando los valores de $n$ crecían. Los resultados fueron la obtención del límite inferior:

$$p(|X - n\theta| \le \varepsilon^*) > 1 - max\left(\frac{b(n\theta+\varepsilon^*,n,\theta)}{b(n\theta,n,\theta)}, \frac{b(n\theta-\varepsilon^*,n,\theta)}{b(n\theta,n,\theta)}\right)$$

y la siguiente aproximación para dicho límite inferior:

$$\frac{b(n\theta,n,\theta)}{b(n\theta+\varepsilon^*,n,\theta)} = \left(\frac{n\theta+\varepsilon^*}{n(1-\theta)-\varepsilon^*+1} . \frac{n\theta+1}{n\theta} . \frac{n(1-\theta)}{n\theta}\right)^{\varepsilon^*/2}$$

Tomando $n = 14.000$ como el número de nacimientos anuales en Londres, $\theta = \frac{18}{35} = 0{,}514265$y $\varepsilon^* = n\varepsilon = 163$ (es decir $\varepsilon = 0{,}0116$), la demostración de N.

*the being and attributes of God from his works of creation"* (1716), obra que se transformó en un clàsico del pensamiento de la Royal Society. Posteriormente, este problema fue tratado por Daniel Bernoulli ("*Mensura sortis ad fortuitam successionem rerum naturaliter contingentium applicata"* (1770-1771)), Pierre Simon de Laplace ("*Mémoire sur la probabilité des causes par les èvénements"* (1774), "*Sur les naissances, les mariages et les morts a Paris, depuis 1771 jusqu'en 1784, et dans toute l'étendue de la France, pendant les anées 1781 et 1782"* (1786), "*Mémoire sur les approximations des formules qui sont fonctions de trés grands nombres et sur leur application aux probabilities"* (1810)) y Wilhelm Lexis ("*Das Geschlechverhältnis der Geborenen und die Wahrscheinlichkeitsrechnung"* (1876)), quien demostró que la probabilidad de los nacimientos masculinos varía temporal y geográficamente.

Bernoulli proporciona los siguientes resultados: $\frac{b(n\theta+\varepsilon^*,n,\theta)}{b(n\theta,n,\theta)} \cong 44{,}74$ y $\frac{b(n\theta-\varepsilon^*,n,\theta)}{b(n\theta,n,\theta)} \cong 44{,}58$ y, por lo tanto[32]:

$$p\left(\left|Y_{14.000} - \frac{18}{35}\right| \leq 0{,}0116\right) > 0{,}9776$$

Como corolario de su teorema se obtiene que el resultado de Jakob Bernoulli se verifica para[33]:

$$\varepsilon^*(\theta, 1-\theta, t^2) \geq \frac{ln(t^2+1)}{ln\left(\frac{(a+c)\theta+1}{a+c}\right)} \cdot \frac{(a+c)+1}{(a+c)\theta+1} - \frac{(a+c)(1-\theta)}{(a+c)\theta+1}$$

De acuerdo con esta expresión, el resultado $n \geq 25.550$ obtenido de la aplicación del teorema de J. Bernoulli para $a+c=50$ y $t^2 = 1.000$, se transforma en $n \geq 17.350$.

En una revisión de este resultado (obtenida como una combinación de las demostraciones de Jakob y Niklaus) J.V. Uspensky (1937) concluyó que $\varepsilon^* \geq 1+(a+c+1)ln(t^2+1)$. Resultado que le permitió demostrar que, para $p, \varepsilon$ y $\delta > 0$, su expresión del teorema de J. Bernoulli, $p(|Y_n - p| \leq \varepsilon) > 1-\delta$, se verifica para $n \geq \frac{1+\varepsilon}{\varepsilon} ln\left(\frac{1}{\delta}\right) + \frac{1}{\varepsilon}$. De modo que, para las condiciones fijadas en el ejemplo precedente, se obtendría que $n \geq 17.700$[34].

En 1733, de Moivre, continuando la línea teórica adoptada por N. Bernoulli, obtuvo una nueva aproximación a la distribución binomial. Este resultado, al cual de Moivre asignó gran importancia, fue publicado bajo el título de "*Approximatio ad summam terminorum binomii (a+b)$^n$ in seriem expansi*". En realidad, estetrabajo constituye una continuación de la serie de investigaciones comenzadas en 1721 (publicadas luego en la "*Miscellanea analytica de seriebus*

---

[32] Obsérvese que con los mismos datos, el resultado obtenido por J. Bernoulli habría sido $p > 0{,}1353$, lo cual pone en evidencia el avance producido por N. Bernoulli. Como se verá en la próxima sección, el resultado obtenido por de Moivre, con los mismos datos, utilizando la aproximación Normal, hubiera sido $p = 0{,}9942$.

[33] Los resultados obtenidos por N. Bernoulli se encuentran en su correspondencia con de Montmort.

[34] A.A. Markov (1924) propuso una versión "modernizada" del teorema de J. Bernoulli utilizando los resultados obtenidos por Niklaus, pero omitiendo la referencia bibliográfica. Asimismo, cabe mencionar que la versión "modernizada" de Uspensky de la demostración de Markov omite la referencia bibliográfica referida a éste y a los Bernoulli.

*et quadraturibus" (1730)*)[35] a partir del siguiente problema (propuesto por Alexander Cuming (1721)): Sean $A$ y $B$ dos jugadores empeñados en un juego consistente en una serie de $n$ partidas, tales que las probabilidades de los jugadores de ganar cada partida sean, respectivamente, $\theta$ y $1-\theta$, para toda la serie. Al cabo de las $n$ partidas $A$ deberá pagar a una tercera persona, $C$, tantas fichas como partidas haya ganado sobre la cantidad $n\theta$ (que se supone que puede asumir solamente valores enteros), y $B$ deberá pagar a $C$ tantas fichas como partidas haya ganado sobre la cantidad $n(1-\theta)$. ¿Cuál es el valor esperado del juego para $C$?

La expectativa de $C$ será, obviamente:

$$E(C) = \sum_{x=n\theta+1}^{n} (x-n\theta)b(x,n,\theta) + \sum_{x=0}^{n\theta-1} [(n-x)-n(1-\theta)]b(x,n,\theta)$$

$$= \sum\nolimits_{x=0}^{n} |x-n\theta| b(x,n,\theta)$$

Partiendo de la igualdad $\sum_{x=0}^{n}(x-n\theta)b(x,n,\theta)=0$, entonces se puede escribir:

$$E(C) = 2\sum\nolimits_{x=n\theta}^{n}(x-n\theta)b(x,n,\theta) = 2\sum\nolimits_{y=0}^{n(1-\theta)} yb(y+n\theta,n,\theta) =$$

$$= 2\sum\nolimits_{y=0}^{n(1-\theta)} y\binom{n}{y+n\theta} p^{y+n\theta}q^{n-y-n\theta}$$

Ahora bien, si se tiene en cuenta que $\frac{b(y+n\theta+1,n,\theta)}{b(y+n\theta,n,\theta)} = \frac{n-y-n\theta}{y+n\theta+1}.\frac{\theta}{1-\theta}$, es decir, que:

$$(n-y-n\theta)\theta b(y+n\theta,n,\theta) = (y+n\theta+1)(1-\theta)b(y+n\theta,n,\theta)$$

sumando entre 0 y $n(1-\theta)$, se obtiene que:

---

[35] El "*Approximatio"* -un pequeño trabajo de sólo 7 páginas-, traducido al inglés por el mismo de Moivre con el título "*A method of approximating the sum of the terms of the binomial* $(a+b)^n$ *expanded into a series, from whence are deduced some practical rules to estimate the degree of assent which is to be given to experiments"* constituyó, en principio, un apartado de circulación privada fechado el 12-11-1733. Posteriormente fue incluido como suplemento en las ediciones sucesivas de la "*Miscellanea Analytica"* (título bajo el cual fue publicada la colección de sus investigaciones del período 1721-1730).

$$\theta\sum_{y=0}^{n(1-\theta)} yb(y+n\theta,n,\theta)+(1-\theta)\sum_{y=0}^{n(1-\theta)}(y+1)b(y+n\theta+1,n,\theta)$$

$$=n\theta(1-\theta)\left[\sum_{y=0}^{n(1-\theta)} b(y+n\theta,n,\theta)-\sum_{y=0}^{n(1-\theta)} b(y+n\theta+1,n,\theta)\right]$$

Por lo tanto, será:

$$\sum_{y=0}^{n(1-\theta)} yb(y+n\theta,n,\theta)=n\theta(1-\theta)b(n\theta,n,\theta)$$

Es decir, queda demostrado que:

$$E(C)=2n\theta(1-\theta)b(n\theta,n,\theta)=2n\theta(1-\theta)\binom{n}{n\theta}\theta^{n\theta}(1-\theta)^{n\theta}$$

A fin de facilitar el cálculo de los coeficientes binomiales, de Moivre comenzó buscando una aproximación a la función $ln(n!)$obteniendo la expresión $ln\left(\frac{n^n}{n!}\right)\approx n-\frac{1}{2}ln(n)-\frac{1}{2}ln(2\pi)$[36].Luego, teniendo en cuenta que:

$$b(n\theta,n,\theta)=\frac{n!}{(n\theta)!\,\big(n(1-\theta)\big)!}\theta^{n\theta}=\frac{(n\theta)^{n\theta}}{(n\theta)!}\cdot\frac{\big(n(1-\theta)\big)^{n(1-\theta)}}{\big(n(1-\theta)\big)!}\cdot\frac{n!}{n^n}$$

concluyó en forma inmediata que:

$$ln[b(n\theta,n,\theta)]=ln\left[\frac{(n\theta)^{n\theta}}{(n\theta)!}\right]+ln\left[\frac{\big(n(1-\theta)\big)^{n(1-\theta)}}{\big(n(1-\theta)\big)!}\right]-ln\left(\frac{n!}{n^n}\right)=$$

$$=-ln\left(\sqrt{n^2\theta(1-\theta)}\right)-ln\big(\sqrt{2\pi}\big)+ln\big(\sqrt{n}\big)=$$

$$-ln\left(\sqrt{2\pi n\theta(1-\theta)}\right)$$

Es decir, que $b(n\theta,n,\theta)=\big(2\pi n\theta(1-\theta)\big)^{-1/2}$[37]. Por otra parte, a partir de la relación:

---

[36] de Moivre, basándose en esta aproximación, publicó su notable tabla de 14 dígitos de la función log(n!), para n=10(10)900, como suplemento a la "*Miscellanea analytica*" (1730). Contemporáneamente James Stirling dio a conocer su solución (conocida hoy como "*aproximación de de Moivre-Stirling*") en "*Methodus differentialis*" (1730) (ver Archibald (1926), Chow; Teicher (2003)).

$$\frac{b(n\theta,n,\theta)}{b(n\theta+\varepsilon,n,\theta)}=\frac{(n\theta+\varepsilon)!\,(n(1-\theta)-\varepsilon)!}{(n\theta)!\,\big(n(1-\theta)\big)!}.\frac{\big(n(1-\theta)\big)^{\varepsilon}}{(n\theta)^{\varepsilon}}=$$

$$=\frac{(n\theta+\varepsilon)!(n(1-\theta)-\varepsilon)!}{(n\theta)!(n(1-\theta)-1)!}.\frac{\big(n(1-\theta)\big)^{\varepsilon-1}}{(n\theta)^{\varepsilon}}=$$

$$=\frac{n\theta+\varepsilon}{n\theta}.\frac{\big(n(1-\theta)\big)^{\varepsilon-1}}{(n\theta)^{\varepsilon-1}}.\frac{\frac{(n\theta+\varepsilon-1)!}{(n\theta)!}}{\frac{(n(1-\theta)-1)!}{(n(1-\theta)-\varepsilon)!}}=$$

$$=\frac{n\theta+\varepsilon}{n\theta}.\frac{\frac{(n\theta+1)(n\theta+2)\dots(n\theta+\varepsilon-1)}{(n\theta)^{\varepsilon-1}}}{\frac{(n(1-\theta)-1)(n(1-\theta)-2)\dots(n(1-\theta)-\varepsilon+1)}{\big(n(1-\theta)\big)^{\varepsilon-1}}}=$$

$$=\left(1+\frac{\varepsilon}{n\theta}\right)\prod_{j=1}^{\varepsilon-1}\frac{1+\frac{j}{n\theta}}{1-\frac{j}{n(1-\theta)}}$$

obtuvo que:

$$ln\left(\frac{b(n\theta,n,\theta)}{b(n\theta+\varepsilon,n,\theta)}\right)=$$

$$=ln\left(1+\frac{\varepsilon}{n\theta}\right)+\sum_{j=1}^{\varepsilon-1}\left[ln\left(1+\frac{j}{n\theta}\right)-ln\left(1-\frac{j}{n(1-\theta)}\right)\right]=$$

$$=\left[\frac{\varepsilon}{n\theta}-\frac{1}{2}\left(\frac{\varepsilon}{n\theta}\right)^2+\frac{1}{3}\left(\frac{\varepsilon}{n\theta}\right)^3-\cdots\right]+$$

$$+\sum_{j=1}^{\varepsilon-1}\left[\frac{j}{n\theta(1-\theta)}+\frac{(\theta^3+(1-\theta)^3)}{3\big(n\theta(1-\theta)\big)^3}+\cdots\right]=$$

$$=\left[\frac{\varepsilon}{n\theta}-\frac{1}{2}\left(\frac{\varepsilon}{n\theta}\right)^2+\frac{1}{3}\left(\frac{\varepsilon}{n\theta}\right)^3-\cdots\right]+\frac{\varepsilon(\varepsilon-1)}{2n\theta(1-\theta)}\approx$$

$$\approx\frac{\varepsilon^2}{2n\theta(1-\theta)}-\frac{\varepsilon}{2n\theta(1-\theta)}+\frac{\varepsilon}{n\theta}-\frac{\varepsilon^2}{2(n\theta)^2}+\cdots$$

[37] Nótese que, contrariamente a lo realizado por J. y N. Bernoulli -quienes intentaron una aproximación a $p(|X-n\theta|\leq\varepsilon)$-, de Moivre buscó una aproximación a $b(x,n,\theta)$.

donde $\varepsilon = O(\sqrt{n})$, es decir, tal que $\lim_{n\to\infty}\left(\frac{|\varepsilon|}{\sqrt{n}}\right) < K$. Luego, para valores de $n$ suficientemente grandes, la expresión anterior se redujo a $ln\left[\frac{b(n\theta,n,\theta)}{b(n\theta+\varepsilon,n,\theta)}\right] \approx \frac{\varepsilon^2}{2n\theta(1-\theta)}$, o, lo que es lo mismo,a:

$$b(n\theta+\varepsilon,n,\theta) \approx n\}b(n\theta,n,\theta)e^{-\frac{\varepsilon^2}{2n\theta(1-\theta)}} \approx$$

$$\approx \frac{1}{\sqrt{2\pi n\theta(1-\theta)}}e^{-\varepsilon^2/2n\theta(1-\theta)}$$

A partir de este resultado concluyó, entonces, que:

$$p(|x-n\theta| \leq \varepsilon) = \sum_{|x-n\theta|\leq\varepsilon} b(x,n,\theta) = \frac{2}{\sqrt{2\pi}}\int_0^{\varepsilon/\sqrt{n\theta(1-\theta)}} e^{-y^2/2}dy$$

Donde $f(y)=\frac{1}{\sqrt{2\pi}}e^{y^2/2}dy$ denota la función de densidad[38].

En términos resumidos este teorema -conocido hoy como *"primer teorema central del límite"*[39]- podría ser enunciado de la siguiente forma: Sea $f_N(\cdot)$la función de densidad, para una variable$\theta$ $(0 < \theta < 1)$ y sea una variable acotada

---

[38] De acuerdo con su costumbre, de Moivre omitió la demostración de este teorema. El desarrollo que figura en el texto fue elaborado estrictamente a partir de relaciones definidas en publicaciones previas (ver Hald (1984), Sheynin (1968)). Obsérvese que N. Bernoulli estuvo muy cerca de anticipar la solución de de Moivre: si hubiese utilizado series logarítmicas sobre su aproximación hubiera hallado que:$ln\left[\frac{b(n\theta,n,\theta)}{b(n\theta+\varepsilon,n,\theta)}\right] = \left[ln\left(\frac{n\theta+\varepsilon}{n(1-\theta)+\varepsilon+1}\right) + ln\left(\frac{n\theta+1}{n\theta}\right) + ln\left(\frac{n(1-\theta)}{n\theta}\right)\right] = \frac{\varepsilon}{2}\left[ln\left(1+\frac{\varepsilon}{n\theta}\right) - ln\left(1-\frac{\varepsilon-1}{n(1-\theta)}\right) + ln\left(1+\frac{1}{n\theta}\right)\right] = \frac{\varepsilon^2}{2n\theta(1-\theta)} - \frac{\theta\varepsilon}{2n\theta(1-\theta)} + \frac{\varepsilon^2}{4}.\frac{\theta^2-(1-\theta)^2}{(n\theta(1-\theta))^2} - \frac{\varepsilon^2}{2}.\frac{\theta^2}{(n\theta(1-\theta))^2} + \cdots$ Y, a partir del supuesto que $\varepsilon = O(\sqrt{n})$, hubiera obtenido inmediatamente que $\lim_{n\to\infty} b(n\theta+\varepsilon,n,\theta) = b(n\theta,n,\theta)\varepsilon^{n\theta(1-\theta)}$.

[39] El calificativo "*central*" agregado a la denominación de este teorema se debe a Pólya (1920), quien quiso, de esta manera, destacar el papel central que juega en la teoría de la probabilidad. Algunos probabilistas franceses entendieron que la inclusión del adjetivo central se debía a que el teorema describía el comportamiento de los valores centrales de la distribución de probabilidades de la variable binomial y que, por lo tanto, dicho adjetivo debía calificar al sustantivo "*límite"*, no al sustantivo "*teorema"*. Por esta razón, es posible hallar en la literatura también la denominación "*teorema del límite central"*. La calificación de "*primero"* implica que la denominación de teorema central del límite resume una familia de teoremas que postulan sucesivas generalizaciones y cuya culminación son las denominadas "*condiciones de Feller-Lindeberg"*.

$\frac{X-n\theta}{\sqrt{n\theta(1-\theta)}} = \frac{\sum_{(n)} U-n\theta}{\sqrt{n\theta(1-\theta\theta)}}$ (donde las $U$ son variables del tipo $b(1,\theta)$), se verifica entonces que:

$$\lim_{n\to\infty}\left[\sqrt{n\theta(1-\theta)}b(x,n,\theta)\right] = f_{N(v)}$$

Es decir,que $\frac{\sum_n u-n\theta}{\sqrt{n\theta(1-\theta)}}$ converge en-distribuición a la variable $\theta$ ($\frac{\sum_n u-n\theta}{\sqrt{n\theta(1-\theta)}} \xrightarrow{d} \theta$)[40].

En particular, si el desvío $\varepsilon$ es medido como $t$ veces $\sqrt{n\theta(1-\theta)}$, entonces se puede obtener el valor de $t$ que determina la probabilidad de que la variable $X$ asuma valores comprendidos en un intervalo dado:

$$p\left(|x-n\theta| \leq t\sqrt{n\theta(1-\theta)}\right) = p\left(\left|\frac{x}{n}-\theta\right| \leq t\sqrt{\frac{\theta(1-\theta)}{n}}\right) =$$

$$= \sum_{|x-n\theta|\leq t\sqrt{n\theta(1-\theta)}} b(x,n,\theta) =$$

$$= p\left(n\theta - t\sqrt{n\theta(1-\theta)} \leq X \leq n\theta + t\sqrt{n\theta(1-\theta)}\right) =$$

$$= p\left(\theta - t\sqrt{\frac{\theta(1-\theta)}{n}} \leq Y_n \leq \theta + t\sqrt{\frac{\theta(1-\theta)}{n}}\right) \approx \frac{2}{\sqrt{2\pi}}\int_0^t e^{-u^2/2}du$$

La expresión precedente permite deducir que el grado de confiabilidad de la estimación aumenta en forma proporcional a la raíz cuadrada del número de observaciones independientes realizadas. Este resultado constituye el gran avance de de Moivre con respecto a la solución de Bernoulli: la cuantificación efectiva

---

[40] Sean una sucesión de variables aleatorias independientes entre sí, $\{X_1, X_2, \dots, X_n\}$, con funciones de ditribución $F_{X_j}(x)(j = 1,2,..,n)$ y una variable $X$, con función de distribución $F_X(x)$, se dice que la sucesión $\{X_n\}$ converge en-distribución (o en-ley o débilmente o en el sentido de Bernoulli) a la variable $X$ ($\{X_n\} \xrightarrow{d} X$) si se verifica que $\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$. La condición necesaria y suficiente para poder asegurar que la sucesión $\{X_1, X_2, \dots, X_n\}$ es de variables estocásticamente independientes es que $F_{X_1,X_2,\dots,X_n}(x_1, x_2, \dots, x_n) = \prod_{j=1}^{n} F_{X_j}(x_j)$. Esta condición se verifica también para cualquier sucesión infinita de variables independientes.

del aumento de confiabilidad ante un incremento de la información empírica y una justificación de su modelo implícito de causación combinatoria[41].

A fin de comprender mejor la evolución histórica posterior de los intentos de solución del problema de Bernoulli, cabe realizar algunas consideraciones sobre esta demostración de de Moivre:

**i)** Algunos autores, como por ejemplo, Karl Pearson (1924)(1925), interpretaron al resultado obtenido por de Moivre como el "*nacimiento"* de la función Normal. Pero, es necesario destacar que de Moivre no consideró a esta función como una función de densidad, sino simplemente como una curva útil para aproximar probabilidades binomiales[42].

**ii)** de Moivre intentó resolver el problema de la inversión de la probabilidad por el absurdo. Bernoulli demostró que, suponiendo ""*...una cierta ley determinada"*, la frecuencia relativa de los eventos "*...se aproximará en forma continua a dicha ley"*, a medida que el número de observaciones aumente indefinidamente. A partir de este postulado, el razonamiento de de Moivre puede ser expresado de la siguiente forma: si se observa que la frecuencia relativa "*...converge a un valor determinado"*, $\theta$, entonces quedaría definida la "*ley"* que "*gobierna"* a dicho evento. Por otra parte, si esta ley fuera expresada por otro valor,$\theta'$, el teorema de Bernoulli dictaminaría que la frecuencia relativa debería converger a ese valor. Pero esto violaría la hipótesis de que la frecuencia relativa converge a $\theta$.Este resultado y el supuesto que el principio de la estabilidad de las frecuencias era una prueba incontrovertible de la existencia de una inteligencia superior que regía el comportamiento de los fenómenos naturales (es decir, de la existencia de un "*Diseño original*"), lo condujeron a la convicción de haber demostrado su propia versión inversa del teorema de Bernoulli, la cual tampoco pudo resolver la circularidad del razonamiento. No obstante, este planteo constituyó un argumento de consideración contra el escepticismo radical que sostenía que las causas regulares no necesariamente tenían que producir efectos regulares. De acuerdo con suinterpretación, no sólo debía esperarse que, en el largo plazo, causas regulares produjeran efectos regulares, sino que la observación de los efectos permitía asintóticamente el descubrimiento de las causas bajo el supuesto ontológico de que dichas causas ("*leyes determinadas*") existían[43].

---

41 $\sqrt{n}$ constituye "*...el módulo mediante el cual regulamos nuestra estimación"* ("*Doctrine of chances"* (1718, p. 252)).

42 Si bien de Moivre no demostró que $\int_{-\infty}^{\infty} f(u)du = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-u^2/2}du = 1$, utilizó a $f(u)$ como una función de densidad.

43 Uno de los principales comentaristas a favor de la demostración de de Moivre fue David Hartley (*"Observations on man, his frame, his duty and his expectation"* (1749)). Su asimilación

**iii)** La restricción más importante de la demostración de de Moivre radica en que la convergencia de la frecuencia relativa se justifica sólo en el límite. Es decir, al igual que Bernoulli, tampoco de Moivre logró resolver el problema de la identificación de la frecuencia relativa convergente, a partir de una sucesión finita de observaciones[44]. Sus intentos de determinar unívocamente el valor de $\theta$ a partir de un argumento matemático como el de los "*puntos de condensación"* en una sucesión finita de observaciones, terminaron en fracaso. De acuerdo con T.C. Fry (1928)[45], podría haber demostrado que la aparente convergencia de las frecuencias relativas era compatible con (y, aún, causada por) la aleatoriedad de las observaciones. Pero, obviamente, esta posibilidad no comulgaba con la concepción del determinista de Moivre, que consideraba, simplemente, que las series "*debían"* converger. Esta negación del argumento aleatorista lo condujo a un esquema de razonamiento circular que podría ser expresado de la siguiente forma: las frecuencias relativas convergen a $\theta$ porque los eventos están regidos por una ley determinada pero, a su vez, la convicción de que los eventos se rigen por una ley determinada, se funda en el postulado de la inversión de la probabilidad.

**iv)** A pesar del adelanto que significó en el tratamiento de los coeficientes binomiales, la aproximación Normal de de Moivre no fue utilizada en la resolución del problema planteado por N. Bernoulli acerca de la estabilidad de la proporción de nacimientos. Es más, contrariamente a lo que se hubiera podido esperar, la solución de de Moivre fue recibida con total indiferencia por sus contemporáneos, posiblemente por el carácter eminentemente matemático del trabajo, o porque debido a razones de orden filosófico-teológico imperantes en la época, el autor enfatizó casi exclusivamente los aspectos vinculados a la ley de los grandes números, es decir, a la desaparición de irregularidades en el comportamiento del fenómeno a medida que se incrementaba el número de observaciones, dejando de lado las consecuencias de su notable resultado acerca del significado de $\sqrt{n}$ como unidad de medida del grado de confiabilidad de una estimación[46]. Citando el argumento probabilístico de Arbuthnott y 'sGravesande

de la inducción a la probabilidad matemática lo convirtió en el principal defensor de las consideraciones de de Moivre sobre las implicancias del teorema de Bernoulli. Al igual que de Moivre, Hartley enfatizó que los resultados que proporcionaba la ley de los grandes números constituían sólo una confirmación de los efectos de la razón y postuló que su teorema representaba una evidencia más de la existencia del *"…Orden y Proporción establecidos por la Divinidad que observan los fenómenos naturales"* (p. 331).

[44] Tampoco indicó un método práctico para obtener un intervalo para $\theta$, en función de los valores de $n, Y_n$ y $t$.

[45] Ver Fine (1973).

[46] Las primeras publicaciones sobre la función Normal se produjeron por obra de los matemáticos-astrónomos de fines del siglo XVIII y comienzos del XIX, fundamentalmente Laplace y Gauss,

sobre la existencia de la Divina Providencia, a partir de la proporción entre nacimientos masculinos y femeninos y de acuerdo con los postulados de la teología natural[47], de Moivre (1718) finalizó su trabajo con la siguiente expresión: "*Se puede concluir, entonces, que si bien el azar produce irregularidades -que pueden ser infinitamente grandes-, con el transcurso del tiempo, la proporción de éstas con respecto a la recurrencia del Orden que resulta naturalmente del 'Diseño Original', no tendrá ninguna importancia (...) De esta forma se demuestra que hay, en la constitución de las cosas, ciertas Leyes de acuerdo a las cuales ocurren los Eventos.Que no es menos evidente a partir de la Observación que esas Leyes sirven a propósitos sabios, útiles y beneficiosos para preservar el Orden del Universo, para propagar las diversas especies de seres y para proporcionar a la Clase conciente tales grados de satisfacción como conviene a su Estado. Pero tales Leyes, así como el Diseño y el Propósito original de su existencia deben ser explicados desde su negación; la 'inercia' de la materia y la naturaleza de todos los seres creados hacen imposible que nada pueda modificar su propia esencia o darse a sí mismos o a cualquier otro una determinación o una propensión original. Por lo tanto, si no nos dejamos enceguecer por el polvo de la metafísica, seremos conducidos a través de un corto y obvio sendero, al reconocimiento del gran HACEDOR y GOBERNADOR de todo, infinitamente sabio, infinitamente poderoso e infinitamente bueno*" (p. 254). Esta hipótesis de "*Orden*" (impuesto por un agente externo, no-natural) a la que se refiere de Moivre -"*...un orden constante, general, magnífico, completo y hermoso*"[48]- sustituyó a la de "*simplicidad de la naturaleza*", postulada por J. Bernoulli[49].Debe tenerse en cuenta que, dado que ni éste ni de Moivre se ocuparon de obtener una confirmación empírica de sus observaciones, sus argumentos teológicos resultan bastante débiles.

---

quienes arribaron al mismo resultado en forma independiente y, presumiblemente, sin conocimiento del desarrollo de de Moivre. Puede decirse que el trabajo de de Moivre permaneció ignorado hasta su redescubrimiento por K. Pearson en la biblioteca del University College de Londres, en 1924.

[47] Arbuthnott (1712): "*Entre las innumerables huellas de la Divina Providencia que pueden hallarse en los Trabajos de la Naturaleza, cabe considerar muy especialmente el mantenimiento de un exacto Balance entre el Número de Hombres y Mujeres; esto prueba que la Especie no puede desaparecer nunca, ya que cada Varón puede tener su Mujer, y de una Edad adecuada. Esta igualdad entre Varones y Mujeres no resulta del azar sino de la Divina Providencia trabajando para un buen Fin"* (p. 186).

[48] Süssmilch (1741, vol. 1, p. 49).

[49] De acuerdo con Poisson (1836a) un "*Orden*" interpretable solamente en términos de expectativa.

## 3.- La solución de Simpson

De acuerdo a lo comentado en la sección precedente, tanto Bernoulli como deMoivre (como la mayoría de los miembros de la Royal Society), en su fidelidad a la teología Newtoniana, veían en sus teoremas límite el argumento según el cual la estabilidad de las frecuencias estadísticas demostraba la presencia de la "*Divina Providencia*"[50]. Uno de los autores que traicionó ese principio fue Simpson. Su trabajo se orientó al estudio de las propiedades de los errores de observación en astronomía. En este sentido, propuso una distribución triangular continua, demostró que, de acuerdo con esta distribución, el valor medio estaba menos afectado por errores no-sistemáticos que cualquier observación individual e intentó hallar su distribución en el límite.

Los problemas planteados por la astronomía condujeron gradualmente al principio de la media aritmética -calculada a partir de observaciones realizadas en igualdad de condiciones- y a distintos métodos de estimación basados en modelos paramétricos, cuya culminación fue el método de los cuadrados mínimos. Pero la teoría de los errores de observación, que constituyó su fundamento, recién surgió con un tratamiento sistemático en el siglo XVIII, en "*Opera miscellanea sive æstimatio errorum in mixta mathesiper variationes partium trianguli plani et sphærici"* (1722) de Roger Côtes. Como continuación de este trabajo pionero, Simpson publicó "*On the advantage of taking the mean of a number of observations in practical astronomy"* (1755) y"*Miscellaneous tracts on some curious, and very interesting subjects in mechanics, physical-astronomy, and speculative mathematics"* (1757).

En estas memorias demostró en primer lugar que, supuesta una hipótesis acerca del carácter probabilístico de las observaciones sobre un fenómeno $Y$, el valor promedio de los valores observados era superior, como estimador ($\hat{Y}$) del verdadero valor de $Y$, a cualquier observación individual: "*Al tomar el valor promedio de un número de observaciones, disminuye en forma apreciable la probabilidad de cometer errores pequeños y prácticamente hace desaparecer toda probabilidad de cometer errores de gran magnitud. Esta última consideración, por sí misma, parece suficiente para hacer recomendable el uso de este método, no sólo a los astrónomos, sino a todos los que estén relacionados con cualquier tipo de información obtenida experimentalmente. Y cuanto mayor sea el número de observaciones realizadas, menor será la propensión a cometer errores, suponiendo que las observaciones se realicen en*

[50] Pearson, K. (1925): "*Los matemáticos ingleses post-Newtonianos experimentaron una mayor influencia de la teología Newtoniana que de su matemática*" (p. 202).

*igualdad de condiciones"*[51]. Esta versión de la ley de los grandes números de Simpson no aparece sustentada por ninguna demostración teórica.

Como se mencionó más arriba, la novedad conceptual introducida por Simpson consistió en dirigir la atención sobre los errores (es decir, sobre las diferencias entre las observaciones registradas y el "verdadero valor" del fenómeno)[52] -asumiendo una hipótesis específica acerca de su distribución de probabilidades- y no sobre las observaciones en sí mismas, ni sobre los objetos observados[53]. Este planteo, que le permitió -aún sin poseer una teoría de la inferencia- medir el error de estimación (es decir, la incertidumbre de la predicción en términos de una distribución de probabilidades dada), constituyó indudablemente el avance fundamental de su trabajo respecto al de de Moivre.

El principio planteado por Simpson implicaba que, una vez asignada una distribución de probabilidades a los errores de observación, la inversión de la probabilidad debía darse por añadidura, postulando que la distribución de probabilidades de $Y$ condicionada por$\hat{Y}$, debía ser tal que$p\left(Y / \hat{Y}\right) \propto p\left(\hat{Y} / Y\right)$[54]. En particular, en su publicación de 1755, Simpson supuso que, dado un

[51] Simpson (1755, pp. 92-93).

[52] En realidad, fue Galileo Galilei ("*Dialogo sui massimi sistemi del mondo"* (1632)) quien introdujo en la literatura el problema de los errores de observación. Si bien no obtuvo una solución cuantitativa ni analítica, arribó a ciertas conclusiones que interpretaban profundamente la esencia del problema: que dichos errores son inevitables, que se distribuyen simétricamente, que la probabilidad de cometer un error de medición aumenta a medida que disminuye la medida del error y que la mayoría de las observaciones se agrupan alrededor del "verdadero valor". Obsérvese que, en estos postulados, están contenidas las características de la distribución Normal que, con el tiempo, se constituiría en una de las funciones básicas de la teoría de la probabilidad y de los rudimentos de la teoría de la contrastación de hipótesis. Conclusiones similares recién pueden ser halladas en "*Beyträge zum Gebrauche der Mathematik und deren Anwendung"* (1765) de Johann Heinrich Lambert.

[53] Simpson supuso que la distribución de probabilidades de los errores era siempre conocida, aún cuando la *"verdadera posición"* del objeto observado pudiera considerarse desconocida (debe recordarse que Simpson era astrónomo). Esta hipótesis constituyó el elemento fundamental para la justificación de cualquier método referido a la estimación mediante la utilización de conjuntos finitos de observaciones.

[54] Stigler, S. M. (1986): "*Simpson había comprendido que el concepto de distribución de los errores permitía un acceso a la medida de la incertidumbre por la puerta de servicio. Fue Laplace quien, colándose posteriormente por ella, logró abrir la puerta principal (sólo para hallar que la llave de Bayes ya se encontraba en la cerradura)"* (p. 98).

conjunto de $n$ observaciones independientes, los errores contenidos en cada una de ellas ($\varepsilon_i = Y_i - \widehat{Y}_\iota, i = 1,2,\dots,n$) podían asumir los valores:

$$-\varepsilon, -\varepsilon + 1, \dots, -2, -1, 0, 1, 2, \dots, \varepsilon - 1, \varepsilon$$

con probabilidades proporcionales a la distribución:

$$q^{-\varepsilon}, q^{-\varepsilon+1}, \dots, q^{-2}, q^{-1}, q^{0}, q^{1}, q^{2}, \dots, q^{\varepsilon-1}, q^{\varepsilon}$$

o a la distribución:

$$q^{-\varepsilon}, 2q^{-\varepsilon+1}, \dots, (\varepsilon-1)q^{-2}, \varepsilon q^{-1}, (\varepsilon+1)q^{0}, \varepsilon q^{1}, (\varepsilon+1)q^{2}, \dots, 2q^{\varepsilon-1}, q^{\varepsilon}$$

(quedando definida una distribución de probabilidades para cada $q > 0$)[55].

Simpson desarrolló solamente el caso para $q = 1$, obteniendo una distribución simétrica de los errores. Si bien la selección de las distribuciones de probabilidades de los errores fue tomada de los trabajos de de Moivre, la forma de su utilización fue totalmente novedosa; adelantándose a Lagrange y a Laplace en más de diez años, desarrolló expresiones para la probabilidad de que el valor medio de los errores del conjunto de observaciones no superara una cantidad dada, y comparó esta probabilidad con las correspondientes a las observaciones individuales.

Para la selección de la primera distribución de probabilidades se basó, en particular, en la solución de de Moivre al problema de la definición de la variable aleatoria que representa la suma de los resultados de $n$ tiradas de un dado "clásico" con $k$ caras, numeradas de 1 a $k$, cuyas probabilidades están dadas por los coeficientes del desarrollo de la expresión:

$$\left(1 + q + q^2 + q^3 + \cdots + q^{k-1}\right)^n = \frac{\left(1 - q^k\right)^n}{(1-q)^n}$$

La segunda distribución la obtuvo de sumar dos variables aleatorias que representan errores independientes, cada uno con distribución de probabilidades de la forma:

$$q^{-\varepsilon/2}, q^{-(\varepsilon-1)/2}, \dots, q^{0}, \dots, q^{(\varepsilon-1)/2}, q^{\varepsilon/2}$$

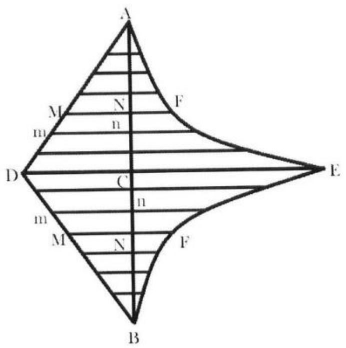
[55] Simpson observó que la suma de los términos de la última sucesión elevados a la potencia $t$ es igual a$q^{-\varepsilon t}(1 + q + q^2 + \cdots + q^{\varepsilon})^{2t} = \frac{1-q^{\varepsilon t+1}}{q^{\varepsilon t}(1-q)^{2t}}$, lo que le permitió concluir que la segunda distribución era asimilable a la primera.

Simpson era conciente (según lo expresó en "*On the advantage of taking the mean"*) del carácter restrictivo de las hipótesis que había asumido acerca de la distribución de los errores, pero también dejó planteada su convicción sobre la posibilidad de generalización de los resultados obtenidos. En 1757, a partir de algunas observaciones realizadas por Bayes a dicha publicación[56], estableció dos principios que caracterizaron los alcances de dicha generalización y que se constituyeron, a partir de ese momento, en las hipótesis implícitas en todos los trabajos sobre teoría del error: **i)** que las probabilidades de ocurrencia de errores de la misma magnitud, por defecto o por exceso, eran iguales; **ii)** que era posible definir límites determinados entre los cuales se podía suponer que estarían incluidos los errores[57].

A partir de este último postulado y remitiéndose a la primitiva distribución de probabilidades discreta proporcional a:

$$1,2,\dots,(\varepsilon+1),..,1$$

introdujo una distribución triangular de los errores (ABD) (la cual sea, posiblemente, la primera publicación de una distribución de probabilidades continua para los errores).




La curva AFEFB representa la función de probabilidades del error que se comete al utilizar el valor medio de los errores. Integrando esta función Simpson obtuvo finalmente que la probabilidad de que el valor medio de $n$

[56] Según se puede deducir de una carta enviada por Bayes a John Canton, fue probablemente "*On the advantage of taking the mean"* la obra que despertó su interés respecto a la teoría de la probabilidad.

[57] Simpson (1757) especificó -arbitrariamente- límites para su distribución de los errores "...*que dependen de las propiedades del instrumento y de la habilidad del observador"* (p. 64).

errores no excediera una proporción dada, $1-\frac{x}{n}$, del error máximo posible (CA) estaba dada por:

$$1-\frac{2}{(2n)!}\left[x^{2n}-2n(x-1)^{2n}+\frac{2n}{1}\frac{(2n-1)}{2}(x-2)^{2n}-\right.$$
$$\left.-\frac{2n}{1}\frac{(2n-1)}{2}\frac{(2n-2)}{3}(x-3)^{2n}+\cdots\right]$$

Una circunstancia que colaboró para que Simpson alcanzara el suceso que no lograron ni los Bernoulli ni de Moivre fue, sin duda, que se independizara de los tradicionales esquemas de urnas (en los que, si bien los argumentos matemáticos necesarios eran simples, los aspectos conceptuales inherentes al problema de la inversión de la probabilidad, dadas su estructura física intrínsecamente discreta y la asimetría existente entre la composición de la urna y el mecanismo de extracción, eran muy complicados) y seleccionara otros modelos -por ejemplo, los que se derivaban de las observaciones astronómicas- en los cuales la cuestión conceptual de la inversión de la probabilidad era más fácil de aprehender.

## 4.- Laplace y la continuación de la obra de Simpson

Laplace aplicó los resultados obtenidos en la inversión de la probabilidad al problema de la selección del mejor valor promedio para un conjunto de observaciones, dada una distribución "a posteriori" de los errores ($p\left(Y\ /\ \hat{Y}\right)$). En principio Laplace propuso dos opciones: **i)** adoptar el valor de $\varepsilon = Y-\hat{Y} = \varepsilon_{me}$ (la mediana de la distribución "a posteriori" de los errores, "*media de la probabilidad*" en la nomenclatura de Laplace) tal que la probabilidad de que el verdadero valor de $\varepsilon$ esté por encima o por debajo de él, sea la misma, **ii)** adoptar el valor de $\varepsilon$ que minimice la suma de los valores absolutos de los desvíos respecto a un valor$\varepsilon_0$, multiplicados por sus correspondientes probabilidades, es decir, el adoptar el valor $\varepsilon_0$ tal que:

$$\int_{-\infty}^{\infty}|\varepsilon-\varepsilon_0|f(\varepsilon)d\varepsilon = mínimo$$

y luego demostró que ambos valores son iguales ( $\varepsilon_{me}=\varepsilon_0$).

Formalizó, asimismo, el principio Simpsoniano que establece que la cuestión central del problema de la selección del valor medio radica en la definición "a priori" de la distribución de probabilidades de los errores: Sea $Y$ el verdadero valor de la medida que se está calculando y sean $\left\{\hat{Y}_i\right\}$ $(i = 1,2,\dots,n)$ las observaciones a realizar sobre dicho valor. Si se supone que los errores $\varepsilon_i = Y-$

$\hat{Y}_i$ poseen un comportamiento aleatorio de acuerdo con una distribución de probabilidades dada, $f(\varepsilon_i)$, entonces la probabilidad de que las observaciones asuman los valores $\hat{Y}_1, \hat{Y}_2, \dots, \hat{Y}_n$ estará dada por:

$$p[(\hat{Y}_1, \hat{Y}_2, \dots, \hat{Y}_n) \,/\, Y] = f(Y - \hat{Y}_1)f(Y - \hat{Y}_1)f(Y - \hat{Y}_2) \dots f(Y - \hat{Y}_n)$$
$$= f(\varepsilon_1)f(\varepsilon_2) \dots f(\varepsilon_n)$$

y la distribución de probabilidades inversa de $Y$ condicionada por $\hat{Y}$ ($p(Y \,/\, \hat{Y})$, de acuerdo con lo demostrado en el teorema de Bayes, debe cumplir la condición de proporcionalidad, $p(Y \,/\, \hat{Y}) \propto p(\hat{Y} \,/\, Y)$ o, más estrictamente, suponiendo una distribución "a priori" de $Y$ uniforme, debe ser:

$$p[Y \,/\, (\hat{Y}_1, \hat{Y}_2, \dots, \hat{Y}_n)] = \frac{f(\varepsilon_1)f(\varepsilon_2) \dots f(\varepsilon_n)}{\int f(\hat{Y}_1 - y)f(\hat{Y}_2 - y) \dots f(\hat{Y}_n - y)dy}$$

Este problema de la selección de la distribución de los errores fue tratado por Laplace fundamentalmente en dos trabajos: "*Mémoire sur le probabilité des causes par les èvénements*" (1774) y "*Mémoire sur le probabilités*" (1781). Laplace comenzó aplicando el principio de la razón insuficiente a las probabilidades correspondientes a todos los valores de $\varepsilon$, con lo que obtuvo como solución la curva $f(\varepsilon) = constante$ y la desestimó porque contradecía el criterio Galileano-Simpsoniano-Newtoniano[58] de que es menos probable cometer errores "grandes" que errores "pequeños". Luego, aplicando el principio de la razón insuficiente al decrecimiento de $f(\varepsilon)$, $\frac{d}{d\varepsilon}f(\varepsilon) = constante$, obtuvo la distribución triangular y también la desestimó porque "*Podemos ver fácilmente que esta disminución no puede ser constante; debe ser menor a medida que las observaciones se aparten más del verdadero valor*" (1774, p. 650).

En su tercera propuesta Laplace llevó el principio de la razón insuficiente mucho más allá que sus predecesores al postular que no existía ninguna razón para suponer que las ordenadas de la curva debían decrecer en forma distinta a sus decrementos (aplicación del principio de la razón insuficiente a las tasas de variación de la función):

---

[58] Debe tenerse en cuenta que, a partir de 1755 todo análisis dirigido a la determinación "a priori" de la forma de la curva representativa de la probabilidad de los errores está "obligado" a cumplir las siguientes condiciones insoslayables en un esquema post-Galileano, post-Simpsoniano y post-Newtoniano: **i)** que la curva sea simétrica con respecto a cero; **ii)** que la curva decrezca hacia cero a medida que el valor del error se aparte de cero en cualquiera de los dos sentidos; **iii)** que el área comprendida bajo la curva sea igual a uno; iv) que se cumpla el principio de la razón insuficiente, es decir, que la curva sea en algún sentido reducible a una descripción en términos de casos igualmente probables.

$$\frac{df(\varepsilon + d\varepsilon)}{df(\varepsilon)} = \frac{f(\varepsilon + d\varepsilon)}{f(\varepsilon)}$$

Es decir, $\frac{df(\varepsilon)}{d\varepsilon} = -af(\varepsilon)$. El resultado obtenido fue la función doble exponencial $f(\varepsilon) = \frac{a}{2}e^{-a|\varepsilon|}$ $(-\infty \leq \varepsilon \leq \infty)$.

Esta solución significó un avance muy importante en la medida que constituyó el primer tratamiento analítico de los errores de observación. Contrariamente a lo realizado por Simpson y Legendre –quienes desarrollaron un procedimiento para la selección y evaluación del valor medio "a priori" de las observaciones-, Laplace consideró la selección del valor medio, condicionado por el conjunto de observaciones.

En la "*Mémoire sur les probabilités*"(1781), Laplace, basándose en una partición aleatoria del intervalo unitario (método que garantiza el cumplimiento del principio de la razón insuficiente, en la medida que no privilegia "a priori" a ninguna distribución de probabilidades) y preservando el cumplimiento de las hipótesis de simetría y unimodalidad, obtuvo la función:

$$f(\varepsilon) = \frac{1}{2a} log\left(\frac{a}{|\varepsilon|}\right) (|\varepsilon| \leq a)$$

(donde $a$ denota el límite de los posibles errores)[59]. Desafortunadamente esta solución también resultó poco apropiada en las aplicaciones[60].

D. Bernoulli ("*Diiudicadtio maxime probabilis plurium observationem discrepatium atqu verosimillina inductio inde formanda*" (1778)) –influido por el principio Simpsoniano de que la curva representativa de la distribución de los errores debe ser de rango finito y debe decrecer "abruptamente" hacia sus extremos- propuso un semicírculo de diámetro conocido cuyo centro era el punto estimado mediante el método de máxima verosimilitud y, en "*Specimen philosophicum de compensationibus horologicis, et veri mensura temporis*"

[59] La explicación de Laplace acerca de las razones para la selección de esta curva puede ser resumida en el siguiente pasaje de la "*Mémoire*": "*Es natural pensar que los mismos errores, en más y en menos, son igualmente probables y que su facilidad de ocurrencia es tanto menor cuanto más grandes sean; si no se tiene otra información con relación a su facilidad de ocurrencia se cae evidentemente en el caso precedente: es necesario, por lo tanto, suponer que en tal caso la probabilidad, tanto del error positivo $x$, como del error negativo $-x$, es igual a $(1 / 2a) log(a / x)$ y que ésta es la ley de probabilidades que hay que tomar como punto de partida para la selección del promedio de los resultados de un conjunto de observaciones*" (p. 166).

[60] Dale (1991): "*Laplace expresa que su objeto ha sido más bien hacer conocer que laluz de la teoría de la probabilidad puede iluminar este tipo de cuestiones, que presentar a los observadores un método práctico y de fácil utilización*" (p. 152).

(1780) utilizó la distribución de probabilidades binomial-simétrica y su aproximación Normal para describir los errores aleatorios en la medición de los tiempos en un reloj de péndulo. Esta fue, probablemente, la primera aplicación de la función Normal como representación de la distribución de los errores, basándose en la hipótesis que cada error puede ser considerado como la agregación de un gran número de "*errores elementales*".

### 5.- El método de la combinación de observaciones

Como una continuación de los trabajos de Simpson y asociada a cuestiones generadas en el ámbito de la astronomía y la geodesia, surgió el método conocido como de la combinación de observaciones, que tuvo su culminación con la aparición, en 1805, del principio de los cuadrados mínimos de Legendre el cual, dotado por Laplace y Gauss de connotaciones probabilísticas proporcionadas por la teoría de los errores de observación, originó la que se dio en llamar la "síntesis de Gauss-Laplace".

En 1722, R. Côtes en su trabajo sobre la estimación de los errores en mediciones trigonométricas ("*Æstimatio errorum in mixta mathesi, per variationes partium trianguli plani et spherici*") –sin mencionar las razones que lo condujeron a tal conclusión- desarrolló la siguiente solución para la determinación de la ubicación de un punto, $x$, en un plano: Dadas cuatro observaciones, $x_1, x_2, x_3$ y $x_4$, realizadas sobre dicho punto y, dadas las respectivas ponderaciones, $d_1, d_2, d_3$ y $d_4$ (inversamente proporcionales a las distancias entre las $x_i$ y $x$), "*la ubicación más probable del punto $x$ es la proporcionada por el promedio aritmético ponderado de las observaciones*" (p. 107)[61].

De acuerdo con Laplace, la primera aplicación que se conoce del resultado de Côtes se debe a Euler (1749) al introducir el hoy conocido como "método de los promedios", tomando como punto de partida una ecuación de la forma[62]:

$$\varepsilon = Y - a_1X_1 - a_2X_2 - \cdots - a_6X_6$$

Donde $\varepsilon$ denotaba el "error de observación aleatorio", los elementos $Y$ y $X_j (j = 1,2, \ldots,6)$eran obtenidos a partir de la observación y los coeficientes $a_j (j = 1,2, \ldots,6)$constituían las incógnitas. A partir de las observaciones obtuvo 75 ecuaciones para estimar 6 incógnitas. Intentó, entonces, resolver el problema de la sobreestimación del sistema formando tantos conjuntos de ecuaciones como incógnitas tenía que resolver. Estos conjuntos estaban estructurados a partir de una adecuada selección en pequeños conjuntos de ecuaciones tomadas bajo

---

[61] La referencia bibliográfica corresponde a la traducción de Gowing (1983).

[62] El método natural de estimación consistía en construir las denominadas "*ecuaciones de condición*", igualando los valores estimados y observados y resolviendo los valores de los coeficientes a partir de un sistema con tantas ecuaciones como coeficientes a estimar.

condiciones similares de modo que proporcionaran, esencialmente, coeficientes aproximadamente iguales y tales que la diferencia entre dos ecuaciones cualesquiera produjera la eliminación de muchos términos. A partir de las 6 ecuaciones resultantes, estimó los coeficientes $a_j$ como aquellos valores que minimizaban el mayor de los errores.

Al año siguiente de la publicación de Euler, apareció "*Abhandlung über die Umwalzung del monds um seine Axe ind die Scheinbare Bewengung der Mondsflecten*" de T. Mayer (1750) quien, a partir de la expresión:

$$\varepsilon = Y - a_0 - a_1 X_1 - a_2 X_2$$

y de las observaciones con que contaba para los elementos $Y, X_1$ y $X_2$, generó un sistema de 27 ecuaciones lineales en las incógnitas $a_0, a_1$ y $a_2$, a las que luego agrupó, basándose en los valores que asumían los coeficientes de las incógnitas, en tres conjuntos de 9 ecuaciones cada uno. Sumando las ecuaciones incluidas en cada grupo, definió 3 ecuaciones lineales para las 3 incógnitas, cuyas estimaciones fueron obtenidas minimizando el error máximo.

Una vez resuelto el problema del agrupamiento de las ecuaciones, Mayer planteó el problema de la medida de la confiabilidad de las estimaciones obtenidas. Para ello formuló una teoría de los errores que, utilizando notación moderna, podría ser resumida de la siguiente forma: Sea$\varepsilon$ el error que se comete al tomar como valor de un coeficiente $a$ a la estimación $\hat{a}_1$, obtenida como solución de un sistema de $n_1$ ecuaciones y sea $\hat{a}_2$ la estimación del mismo coeficiente obtenida de un sistema de ecuaciones de tamaño $n_2$. Partiendo de la hipótesis de que el error en cada una de las estimaciones varían en forma inversamente proporcional al número de observaciones[63], concluyó que:

$$\frac{\varepsilon}{1 / n_1} = \frac{|\hat{a}_2 - \hat{a}_1| + \varepsilon}{1 / n_2}$$

Es decir, que$\varepsilon = \frac{n_2}{n_1 - n_2} |\hat{a}_2 - \hat{a}_1|$.

Como se puede apreciar, la aproximación estadística de Mayer a la solución del problema de las observaciones realizadas en condiciones diferentes, es totalmente diferente a la solución matemática de Euler. La hipótesis de Mayer de que la combinación de observaciones aumentaba la confiabilidad de los resultados proporcionalmente al número de ecuaciones combinadas –contraria al criterio de Euler, que considera que el error aumenta con la agregación-, se

[63] De acuerdo a lo demostrado por de Moivre, esta hipótesis es errónea: la confiabilidad de una estimación varía en forma directamente proporcional a la raíz cuadrada del número de observaciones, suponiendo que las observaciones se hayan realizado en igualdad de condiciones. No obstante es necesario destacar la originalidad de la idea que precedió en mucho tiempo alos desarrollos de Laplace y Gauss.

basaba en el concepto estadístico que supone la compensación de los errores aleatorios, de acuerdo con los principios Simpsonianos-Newtonianos.

En 1755 R.J. Boscovich, planteó una solución basada en cinco ecuaciones del tipo:

$$Y = a_0 + a_1 X_1 + \varepsilon$$

El gran avance de Boscovich con respecto al trabajo de Mayer fue la formulación de un principio objetivo general de acuerdo con el cual los coeficientes $a_0$ y $a_1$ debían ser estimados de acuerdo con las siguientes condiciones:

**i)** $\sum_i \varepsilon_i = \sum_i (Y_i - a_0 - a_1 X_i) = \frac{1}{n}\sum_i Y_i - a_1 \sum_i X_i = 0$

**ii)** $\sum_i |\varepsilon_i| = \sum_i |Y_i - a_0 - a_1 X_i| = min$

A partir de la primera de estas condiciones que, como se vio, después de los trabajos de Simpson y Newton, se convirtió en obligada en todo análisis "a priori" de la distribución de probabilidades de los errores y que es la consecuencia formal de la noción intuitiva de que los errores positivos y negativos deben ser igualmente probables, se obtiene que:

$$a_0 = \frac{1}{n}\sum_i Y_i - a_1 \frac{1}{n}\sum_i X_i$$

Con respecto a la segunda condición, a partir de la cual, reemplazando el valor de $a_0$ se obtiene que:

$$\sum_i \left| \left( Y_i - \frac{1}{n}\sum_i Y_i \right) - a_1 \left( X_i - \frac{1}{n}\sum_i X_i \right) \right| = min$$

Boscovich interpretó que la única solución analítica posible era mediante la diferenciación con respecto a $a_0$ y $a_1$, por lo que eligió desarrollar una solución geométrica. Fue Laplace (1786) quien estableció un tratamiento algebraico del algoritmo de Boscovich(al cual denominó "*método de situación*".

Como continuador inmediato de la línea teórica de Simpson merece ser mencionado J.H. Lambert (1760)(1765). En particular en "*Beyträge zum Gebrauche der Mathematik und deren Anwendung*" (1765), enunció condiciones específicas sobre las distribuciones de probabilidades de los errores y desarrolló métodos de estimación que, si bien no fueron ampliamente adoptados, ejercieron cierta influencia sobre la obra de Laplace y Gauss.

## 6.- La síntesis de Gauss-Laplace

Las dos corrientes teóricas expuestas en la sección precedente (Mayer-Laplace y Boscovich-Laplace) –originadas, según se vio, en la premisa de que la combinación de observaciones era esencial para la interpretación de la realidad a partir de la experiencia- tuvieron su culminación en la obra del matemático francés A.M. Legendre. En 1805 este autor dio a conocer "*Nouvelles méthodes pour la détermination des orbites des comètes*", en cuyo apéndice –titulado "*Sur la méthode des moindres carrés*"- expuso el método de los cuadrados mínimos, según el cual "*En muchas investigaciones, en la que el objetivo es deducir resultados a partir de observaciones en la forma más confiable posible, nos encontramos con un sistema de ecuaciones de la forma:*

$$y_i = ax_i + bz_i + cw_i + \cdots + \varepsilon_i$$

*donde* $y_i, x_i, z_i, w_i, \ldots$ *denotan coeficientes de valor conocido, que varían de una ecuación a otra y* $a, b, c, \ldots$ *denotan cantidades desconocidas a ser determinadas de acuerdo con la condición de que todos los valores de* $\varepsilon_i$ *(que representan a los errores de obervación) se anulen, o se reduzcan a una cantidad muy pequeña*" (p. 77).

La solución que propuso consistía "*…en minimizar la suma de los cuadrados de los errores*". Legendre obtuvo esta minimización inmediatamente, resolviendo el sistema de ecuaciones lineales que resultaba de derivar dicha suma de los cuadrados de los errores:

$$f(a, b, c, \ldots) = \sum_i \varepsilon_i^2 = \sum_i (y_i - ax_i - bz_i - cw_i - \cdots)^2$$

con respecto a $a, b, c, \ldots$[64] :

$$\sum_i x_i y_i = a \sum_i x_i^2 + b \sum_i x_i z_i + c \sum_i x_i w_i + \cdots$$

[64] En realidad, Legendre no proporciona ningún tipo de demostración teórica. Sólo comenta que "*…para definir la ecuación del mínimo con respecto a cada una de las incógnitas, es necesario multiplicar todos los términos de cada ecuación por el coeficiente de dicha incógnita en esa ecuación tomado con su propio signo, y luego hallar la suma de todos estos productos*" (p. 77).

$$\sum_i z_i y_i = a \sum_i z_i x_i + b \sum_i z_i^2 + c \sum_i z_i w_i + \cdots$$

$$\sum_i w_i y_i = a \sum_i x_i w_i + b \sum_i z_i w_i + c \sum_i w_i^2 + \cdots$$

Según se vio en las secciones precedentes, durante el siglo XVIII se habían desarrollado, en forma completamente independiente, dos metodologías dirigidas a resolver el problema de la inferencia. Por una lado, una corriente que evolucionó a partir del análisis de la distribución de probabilidades binomial y que culminó con los trabajos de Laplace referidos al establecimiento de los principios para la selección del valor medio de un conjunto de observaciones y de las distribuciones de probabilidades representativas de los errores de observación. Y, por otro, una línea basada en los trabajos de Mayer-Laplace y Boscovich-Laplace, sobre combinación de observaciones, que derivó en el método de los cuadrados mínimos de Legendre y que se desarrolló atendiendo exclusivamente a los aspectos matemáticos del problema, sin ninguna consideración formal sobre sus aspectos probabilísticos, ni sobre las cuestiones referidas a la cuantificación de la incertidumbre, que son inherentes a la inferencia[65].

Fue Gauss el encargado de conectar ambas corrientes teóricas. En 1809 se publicó su "*Theoria motus corporum coelestiorum in sectionibus conicis solum ambientium*", la cual contiene una sección dedicada a su interpretación del problema de la combinación de observaciones, en la que presenta la siguiente exposición del método de los cuadrados mínimos: Sean $n$ funciones lineales de la forma:

$$\begin{cases} \hat{Y}_1 = a_1 X_{11} + a_2 X_{21} + a_3 X_{31} + \cdots \\ \hat{Y}_2 = a_1 X_{12} + a_2 X_{22} + a_3 X_{32} + \cdots \\ \hat{Y}_3 = a_1 X_{13} + a_2 X_{23} + a_3 X_{33} + \cdots \\ \qquad \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \end{cases}$$

(donde $X_1, X_2, X_3, \ldots$ son observables y $a_1, a_2, a_3, \ldots$ representan $n$ coeficientes desconocidos cuyos valores pueden ser obtenidos a partir de un conjunto de observaciones $Y_1, Y_2, Y_3, \ldots$; $X_{11}, X_{12}, X_{13}, \ldots$; $X_{21}, X_{22}, X_{23}, \ldots$) que integran el

---

[65]Stigler, S.M. (1975): "*El método de los cuadrados mínimos proporcionó resultados que pudieron ser considerados 'los mejores', ya que minimizaban la suma de los cuadrados de los errores y producían un atractivo equilibrio mecánico; pero, dado que la naturaleza estocástica de las observaciones no fue tenida en cuenta, es decir, no se consideró la cuantificación de la incertidumbre, no fue posible dar una respuesta a la pregunta ¿Cuán bueno es 'el mejor'?*"

sistema de "*ecuaciones normales*" (de acuerdo con la nomenclatura de Gauss (1823))[66].

Si se supone que los errores $\varepsilon_i = Y_i - \hat{Y}_i (i = 1,2,3, \ldots)$ son variables iidcon función de densidad $(f(\varepsilon_i) = f(\varepsilon)(i = 1,2,3, \ldots)$, entonces el sistema de valores "más probable" corresponderá al conjunto de coeficientes $a_1, a_2, a_3, \ldots$ que maximiza, de acuerdo con el método de optimización de máxima verosimilitud, la función $\varphi = f(\varepsilon_1) f(\varepsilon_2) f(\varepsilon_3) \ldots$. Estos valores pueden ser obtenidos a partir de la resolución del sistema de ecuaciones $\frac{\partial \ln \varphi}{\partial a_i} = 0 \ (i = 1,2,3, \ldots)$. Ahora bien, esta solución implica necesariamente la definición previa de la función $f(\varepsilon)$. Pero la forma de esta distribución no puede ser determinada "a priori", sólo se le pueden atribuir las ya mencionadas características Galileanas-Simpsonianas-Newtonianas de carácter general. A partir de estas hipótesis y asumiendo como axioma el principio de que la media aritmética de un conjunto de observaciones realizadas en igualdad de condiciones sobre una medida, constituye el valor más probable de la misma, se demuestra que, en el caso en que se verifique que $Y_i = a + \varepsilon_i$, entonces la media aritmética de las observaciones:

$$\bar{Y} = \frac{1}{n}(Y_1 + Y_2 + Y_3 + \cdots)$$

maximiza la función $\varphi$ si y sólo si:

$$f(\varepsilon) = \frac{h}{\sqrt{\pi}} e^{-h^2 \varepsilon^2}$$

donde $h > 0$ denota la medida de la precisión de las observaciones.

A partir de la versión de Laplace del teorema de Bayes (suponiendo, de acuerdo con el principio de la razón insuficiente, una distribución uniforme de las probabilidades "a priori"), Gauss demostró que la distribución "a posteriori" de las incógnitas cumple la siguiente condición de proporcionalidad:

$$p\{[(a_1 < \hat{a}_1 < a_1 + da_1) \cap (a_1 < \hat{a}_1 < a_1 + da_1) \cap \ldots] \\ / \, [(Y_1 = y_1) \cap (Y_2 = y_2) \cap \ldots]\} =$$

$$= k\varphi da_1 da_2 \ldots = \frac{\varphi da_1 da_2 \ldots}{\int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} \varphi da_1 da_2 \ldots} \approx e^{h^2 \sum \varepsilon^2}$$

[66] Gauss no proporciona ninguna explicación acerca de los detalles de esta demostración.

y que esta probabilidad se maximiza cuando se minimiza la función $\sum_i \varepsilon_i^2$. Es decir, cuando se verifica el principio de los cuadrados mínimos[67].

En su "*Bestimmung der Genauigkeit der Beobachtungen*" (1816) Gauss introdujo la función:

$$F(u) = \frac{2}{\sqrt{\pi}} \int_{-\infty}^{u} e^{-u^2} du$$

y el coeficiente $r = \frac{\theta}{h}$ (donde $\theta = 0{,}4769363$ es el valor tal que $F(\theta) = \frac{1}{2}$) representa el "*error probable para la función* $F(hu)$".

Analizó, por otra parte, la estimación de $h$ a partir del resultado de $n$ observaciones independientes. Diferenciando la función de densidad correspondiente al sistema de las $n$ observaciones:

$$f(\varepsilon) = ch^n exp\left(-h^2 \sum_i \varepsilon_i^2\right)$$

obtuvo que:

$$f'^{(\varepsilon)} = ch^{n-1}\left(n - 2h^2 \sum_{i=1}^{n} \varepsilon_i^2\right) exp\left(-h^2 \sum_i \varepsilon_i^2\right)$$

y concluyó que el valor de $h$ correspondiente al máximo de la función $f(\varepsilon)$ estaba definido por:

$$h = \sqrt{\frac{n}{2\sum_{i=1}^{n} \varepsilon_i^2}}$$

[67] Desde el primer momento se planteó una ardua disputa entre Gauss y Legendre por la paternidad del método de los cuadrados mínimos. De acuerdo con su propio testimonio Gauss (según la correspondencia de Gauss a H.M.W. Olbers, a B.A. vonLindenau y a F.X. von Zach de 1802) había utilizado esta metodología desde 1795. Pero, aparentemente, no adivinó (o no logró comunicar) la importancia del resultado. Su conclusión fue que no era posible lograr ningún progreso en términos de estimaciones "*más probables*" de las incógnitas, si no se definía la forma de la distribución de probabilidades de los errores, por lo que decidió adoptar el camino inverso y buscar la distribución que convirtiera a la media artmética en el valor más probable. Ahora bien, más allá de la cuestión de la prioridad, es necesario tener en cuenta que la contribución de Legendre se limitó a la definición del principio de los cuadrados mínimos y que fue Gauss el encargado de desarrollar la teoría basada en dicho principio.

Este resultado le permitió demostrar que el valor más probable de $r$ era:

$$r = \frac{\theta}{h} = \theta\sqrt{\frac{2\sum_i \varepsilon_i^2}{n}}$$

Posteriormente consideró el caso más general para varianzas no-constantes ($\sigma^2(\varepsilon_i)$, $i = 1,2,\ldots$), lo que lo condujo al método de los "cuadrados mínimos ponderados", en el que la función a minimizar era de la forma $\sum_i \frac{\varepsilon_i^2}{\sigma_i^2}$ y amplió los desarrollos anteriores a funciones lineales que involucraban más de una incógnita, demostrando que el vector de los coeficientes $a = [a_i]$ poseía una distribución de probabilidades Normal multidimensional del tipo $N(a, (XX')\sigma_\varepsilon^2)$.

El razonamiento de Gauss fue exactamente el inverso del propuesto en su momento por Simpson y Laplace quienes, basándose en el principio de la razón insuficiente, establecieron condiciones para determinar la forma "a priori" de la curva $f(\varepsilon)$ y,a partir de esta definición, seleccionar promedio más adecuado para utilizar como verdadero valor del error. Por el contrario, el argumento de Gauss pude ser resumido de la siguiente forma: **i)** la media aritmética de un conjunto de observaciones referidas a una medida desconocida constituye el valor más probable de dicha medida, sólo si los errores se distribuyen de acuerdo a una función Normal;**ii)** por su parte, la media aritmética es generalmente reconocida como la mejor forma de "*combinar*" observaciones, por lo que se puede concluir que los errores deben ser interpretados como Normalmente distribuidos; **iii)** pero la hipótesis de Normalidad de los errores permite la maximización de la función $\varphi$, cuando se minimiza la forma cuadrática $\sum \varepsilon_i^2$ y esto se verifica cuando los valores de $a_i$ son obtenidos de acuerdo con el principio de los cuadrados mínimos[68].

Aquí hizo su aparición una vez más Laplace para enmendar la falla epistemológica que implica la aparente circularidad de este razonamiento. En el suplemento a su "*Mémoire sur les approximations des formules qui sont fonctions de très grands nombres et sur leur application aux probabilités*" (1786), partir del supuesto que los errores a los que se refería la formulación de Gauss estaban formados por la agregación de "*errores elementales*" y, vinculando su versión del teorema central del límite con la estimación lineal, demostró que era legítimo suponer que dichos errores poseyeran una distribución asintóticamente Normal. Por otra parte, basándose en su "*Mémoire sur la probabilité des causes par les èvénements*" (1774), concluyó que las

[68] Una derivación alternativa –prácticamente desconocida-de la función Normal como "*ley para la distribución de los errores aleatorios*" fue propuesta por Adrain (1809).

estimaciones "*más probables*" obtenidas por el método de los cuadrados mínimos minimizaban, a su vez, el valor esperado "a posteriori" del error.

En 1811 Laplace publicó la "*Mémoire sur les intégrales définies et leur application aux probabilités, et spécialement á la recherche du milieu qu'il faut choisir entre les résultats des observations*", que contiene una formulación algebraica de su generalización del método de Mayer: Sea un sistema de ecuaciones inconsistentes de la forma $\varepsilon_i = aX_i - Y_i\,(i = 1,2,3,\dots)$. Donde los valores de $X_i$ e $Y_i$ se obtienen de la observación repetida en igualdad de condiciones, $a$ es una incógnita y los $\varepsilon_i$ denotan variables aleatorias que representan a los errores no-sistemáticos y simétricamente distribuidas con respecto a cero. Laplace construyó tantas combinaciones lineales de las ecuaciones como incógnitas a resolver:

$$\sum_i k_i\varepsilon_i = a\sum_i k_iX_i - \sum_i k_iY_i$$

De modo que:

$$a = \frac{\sum_i k_iY_i}{\sum_i k_iX_i} + \frac{\sum_i k_i\varepsilon_i}{\sum_i k_iX_i} = \frac{\sum_i k_iY_i}{\sum_i k_iX_i} + u$$

Donde $u$denotaba la variable aleatoria que representaba el error que surgía cuando la función lineal de los errores ($\sum k_i\varepsilon_i$) no era exactamente igual a cero.

Si el número de observaciones era suficientemente grande, de acuerdo con una presentación ligeramente diferente del teorema central del límite, esta variable tendría una distribución aproximadamente Normal con valor esperado proporcional a $f(k_i) = \frac{\sqrt{\sum k_i^2}}{\sum k_iX_i}$. A partir de la relación:

$$\frac{df}{dk_i} = \frac{\sum_i k_i \sum_i k_iX_i - \sum_i k_i^2 \sum_i X_i}{\sqrt{\sum_i k_i^2}\,(\sum_i k_iX_i)^2} = 0$$

Laplace concluyó que el "*error esperado*" asumía su valor mínimo cuando las ponderaciones $k_i$ eran proporcionales a $X_i$ y, en ese caso se verificaba que $a = \frac{\sum_i X_iY_i}{\sum_i X_i^2}$ la cual era, precisamente, la estimación que se obtenía de aplicar el método de los cuadrados mínimos; es decir, de minimizar la forma cuadrática $(a\sum_i X_i - \sum_i Y_i)^2$ con respecto a $a$. En otras palabras, Laplace demostró: **i)** que todos los estimadores de $a$ definidos como funciones lineales de los valores observados $Y_i$ poseían una distribución de probabilidades aproximadamente Normal y **ii)** que entre esos estimadores, los que se obtenían de aplicar el método de los cuadrados mínimos, eran los que poseían menor error esperado.

En esa misma memoria Laplace generalizó este resultado a casos que incluían más de una incógnita:

$$\varepsilon_i = a_1 X_{i1} + a_2 X_{i2} - Y_i (i = 1,2,3, \ldots)$$

y demostró que las estimaciones que minimizaban el error esperado eran las que proporcionaba el método de los cuadrados mínimos, $\hat{a} = (X^T X)^{-1} X^T Y$ (donde $X = [X_{ij}]$ e $Y = [Y_i]$). Es decir las que minimizaban la expresión $\sum_i (a_1 X_{i1} + a_2 X_{i2} - Y_i)^2$.

Por su parte, en 1823, Gauss publicó la primera parte de su "*Theoria combinationis observationum erroribul minimis obnoxiæ*" en la que, abandonando su razonamiento circular de 1809, expuso una generalización del método de Laplace. Demostró que entre todas las estimaciones insesgadas de los coeficientes, aquellas obtenidas de acuerdo con el criterio de los cuadrados mínimos poseían varianza mínima. En la segunda parte de esta obra –publicada en 1823- generalizó los resultados anteriores a cualquier combinación lineal de los coeficientes. Halló la fórmula para determinar la suma de los cuadrados de los residuos ($\hat{\sigma}_\varepsilon^2$) a partir de las estimaciones de los $a$. Desarrolló un procedimiento para permitir el agregado al modelo de un $a$ adicional sin tener que recalcular los coeficientes ya estimados. Demostró que, para series de $n$ observaciones en modelos con $k$ incógnitas ($n > k$), se verificaba que $E(\hat{\sigma}_\varepsilon^2) = (n-k)\sigma_\varepsilon^2$.

El núcleo central alrededor del cual se desarrolló esta segunda versión de la teoría de los cuadrados mínimos podía resumirse en el siguiente principio: Cuando se utiliza un estimador$\hat{a}$de$a$, se comete un error $a - \hat{a}$ que implica una pérdida. Se debe elegir, entonces un estimador que minimice dicha pérdida. Si se supone que la pérdida es proporcional al error cuadrático $(a - \hat{a})^2$, el estimador debe ser el que minimice el error medio cuadrático $E[(a - \hat{a})^2]$.

Gauss demostró que si se acepta la hipótesis de que los errores: i) son lo suficientemente pequeños como para que sus potencias de orden mayor que uno pueden ser despreciadas (es decir, si se consideran solamente los estimadores lineales, ii) están independientemente distribuidos con valor esperado nulo ($E(\varepsilon) = 0$) y la misma varianza desconocida ($E(\varepsilon\varepsilon^T) = \sigma_\varepsilon^2 I$), entonces los estimadores obtenidos a partir de la aplicación del principio de los cuadrados mínimos, son los que minimizan el error medio cuadrático $(X^T X)^{-1} \sigma_\varepsilon^2$ .

## 7.- Modelos dinámicos y procesos estocásticos

A partir de 1821 la síntesis de Gauss-Laplace había alcanzado el grado de madurez necesario como para afirmar que el problema de la inferencia se encontraba resuelto y en condiciones de ser aplicado a ramas de la ciencia muy distintas de la astronomía y la geodesia en las que se había generado. No obstante y a pesar de la innegable importancia en el análisis de los fenómenos (ampliamente reconocida desde la época de Bernoulli), la teoría de la inferencia necesitó más de un siglo y el desarrollo de los conceptos de correlación y regresión para que las aplicaciones se extendieran a las ciencias sociales, fundamentalmente por obra de J. Fourier, F. Galton, G. Weldon, F.Y. Edgeworth, K. Pearson, E.S. Pearson, J. Neyman, G.U. Yule y R.A. Fisher.

Esta demora en la adopción de una interpretación probabilística del comportamiento de los fenómenos de las ciencias sociales, parece confirmar el profundo carácter filosófico de este cambio intelectual y, fundamentalmente, el peso del determinismo en la historia de las ciencias fácticas. Resulta sumamente curioso que autores como A. Quetelet, A. Cournot y W.S. Jevons, quienes publicaron obras sobre teoría de la probabilidad, no utilizaran técnicas probabilísticas en las ciencias sociales.

Esta actitud de los humanistas de la primera mitad del siglo XIX puede ser atribuida principalmente a la tradicional identificación de los conceptos de probabilidad y frecuencia relativa y, en consecuencia, al problema de categorización de las observaciones en grupos homogéneos que permitían construir los “colectivos”. Esta categoría, que en astronomía puede suponerse “a priori”, en las aplicaciones económicas sólo puede aceptarse eventualmente “a posteriori” con un grado de complicación, obviamente, mucho mayor.

Como consecuencia de esta modificación en los fundamentos de la inferencia, se produjo el abandono del teorema de Bayes, no sólo como un rechazo a nivel teórico, sino también debido a que esta nueva forma de evaluación de la probabilidad no permitía, salvo en casos muy particulares, la determinación de las probabilidades “a priori”. Esta circunstancia se vio agravada por la negación del principio de indiferencia y, en consecuencia, de la hipótesis de distribución uniforme de las probabilidades “a priori”.

Esta modificación dogmática no implicó solamente el rechazo de la distribución uniforme sino, también, la pérdida de la hegemonía de la distribución Normal. En contradicción a las propuestas de Quetelet (1848) y Galton (1889) –a quienes se puede considerar los artífices del redescubrimiento de la distribución Normal en el comportamiento de gran número de fenómenos naturales- K. Pearson (1893)(1894)(1895)(1896)(1978) concluyó que la adaptación de la curva Normal a los resultados experimentales era mucho menor de lo que se había supuesto. Los resultados experimentales revelaron a Pearson (1895) que en las distribuciones empíricas era común observar leves asimetrías, circunstancia que lo condujo a definir una familia formada por 12 curvas de frecuencia que utilizó asimilándolas a distribuciones de probabilidades.

En lo que hace al ámbito particular de la teoría económica, el desarrollo de la economía matemática –iniciado a comienzos del siglo XIX- no adoptó a la probabilidad como una herramienta útil en las investigaciones hasta 1930, a pesar de que la teoría de la probabilidad –que ya había revolucionado los métodos de investigación de otras ciencias (como la biología)- se desarrolló fundamentalmente como un instrumento para analizar el comportamiento de los fenómenos de las ciencias sociales, en especial, de la economía y de la ciencia actuarial.

Entre los autores de este período de “economía sin probabilidad” merecen ser mencionados A.A. Cournot (1838) (quien puede ser considerado el primer modelista de la economía), L. Walras (1874) (quien trató a la teoría económica según los cánones de la mecánica racional, basándose en conceptos puramente determinísticos), W.S. Jevons (1865), F.Y. Edgeworth (1910), J.M. Keynes (1921).

Una explicación a esta demora en la introducción de la probabilidad en la modelización económica puede derivarse del hecho que todos estos autores desarrollaron un concepto de teoría económica basado en la interpretación determinística del comportamiento de los fenómenos de la física clásica de fines del siglo. Fue la influencia que ejercieron los trabajos de Maxwell y Gibbs (en particular sobre Edgeworth (1881), Samuelson (1947) y Bowley (1901)) y el cambio de paradigma en las investigaciones biológicas (en particular por la influencia de los trabajos de Galton y Pearson) lo que logró relacionar la teoría económica con los métodos estadísticos para cuantificar leyes y relaciones entre factores. Este avance se debió fundamentalmente a los resultados obtenidos en la estadística macroeconómica surgida de las contribuciones de De Bruyn Kops (1869), D. Baxter (1870) y F. Fellner (1901)(1905), continuadas por los por los trabajos de Marshall y Edgeworth referidos a la verificación estadística de conceptos de teoría económica y por el desarrollo de la teoría de los números índice por obra de E.A.T. Laspeyres (1864a)(1864b), H. Paasche (1871), M.W. Drobisch (1871),W.S. Jevons (1865), H.L. Wetergaard (1890), I. Fisher (1892)(1911)(1922) y F.Y. Edgeworth (1896).

Debe tenerse en cuenta, además, que esta introducción de los métodos estadísticos en la ciencia económica dio origen a la econometría de los fenómenos dinámicos la cual, basándose en el argumento que las observaciones sobre los fenómenos económicos no eran en igualdad de condiciones y, en consecuencia, no cumplían las condiciones de independencia y homogeneidad exigidas por la inferencia clásica (surgida de la síntesis de Gauss-Laplace), no sólo ignoró las nociones probabilísticas, sino que las rechazó.

Esta circunstancia condujo un método alternativo de inferencia (basado en conjuntos de observaciones obtenidas en condiciones diferentes) en el que un fenómeno dinámico $Y(t,w)$ se consdera asimilable a un proceso estocástico que evoluciona en el dominio del tiempo ($t \in T$) y cuya configuración varía en el dominio de los estados (o de las fases o de las variables) ($w \in \Omega(Y)$).

La interpretación determinística del comportamiento de $Y(t,w)$ (al menos a nivel macroscópico) se basa en ciertas premisas de orden metafísico[69]: **i)** que el ámbito al que pertenecen los fenómenos es real; **ii)** que existen leyes objetivas que rigen su comportamiento y **iii)** que estas leyes son inherentes a los fenómenos, racionales y asintóticamente cognoscibles. En otros términos, cada estado $w(t)$ se supone definido por la realización simultánea, en el momento $t$, de las infinitas variables aleatorias que forman su estructura causal:

$$\Omega\big(Y(t)\big) = \{Y(t-j), X_1(t-h_1), X_2(t-h_2), X_3(t-h_3), \dots\}$$

(para $j, h_1, h_2, \dots \in \mathbb{R}, j > 0, h_i \geq 0, i = 1,2, \dots$) y la sucesión temporal de los estados determina su trayectoria. De modo que el fenómeno $Y(t,w)$ queda definido como una entidad que evoluciona en un ámbito espacio-temporal caracterizado por el principio de causalidad.

Este supuesto principio de solidaridad universal que relaciona causalmente a los fenómenos y que hace que la naturaleza de $Y(t)$ aparezca como infinitamente complicada, permite concluir que la información con que cuenta el observador ($\Omega^*\big(Y(t)\big) \subset \Omega\big(\mathrm{Y(t)}\big)$) siempre será insuficiente y que, en consecuencia, una parte importante de su comportamiento permanecerá ignorada para sí, de modo que, en ciertas condiciones de estacionariedad, se puede escribir $Y(t) = f\big[\Omega^*\big(Y(t)\big)\big] + \varepsilon(t)$, donde: **i)** $f\big[\Omega^*\big(Y(t)\big)\big]$ denota la representación del comportamiento de $Y(t)$ a partir del conjunto de información $\Omega^*\big(Y(t)\big)$, es decir, el comportamiento que debería observar $Y(t)$ si los factores incluidos en $\Omega^*\big(Y(t)\big)$ fueran sus únicas causas y no estuvieran

[69]. A las cuales no es posible atribuir ningún fundamento (ni inductivo, ni deductivo) y que, según Daston (1988), en la interpretación de Hume (1718), constituyen "...*una necesidad psicológica un precepto casi involuntario instituido por la caritativa naturaleza para compensar las deficiencias de la razón humana*" (p. 202).

afectados por errores de medición, y $f[\cdot]$ fuera la función que representara la verdadera relación causal invariante en el tiempo entre estos factores e $Y(t)$; **ii)** $\lim_{\Omega^*(Y(t))\to\Omega(Y(t))} f\left[\Omega^*\left(Y(t)\right)\right] = Y(t)$ y **iii)** $\boldsymbol{\varepsilon(t)}$ denota el componente azar-ignorancia $\left(\varepsilon(t) = Y(t) - f\left[\Omega^*\left(Y(t)\right)\right]\right)$ y, en consecuencia, es tal que $\lim_{\Omega^*(Y(t))\to\Omega(Y(t))} \varepsilon(t) = 0$).

En este paradigma, para un observador ideal que contara con un conjunto de información no afectado por errores de medición, que abarcara la totalidad de la estructura causal de un fenómeno, su trayectoria admitiría una representación en términos de mecánica clásica reversible en el dominio del tiempo, en la que éste transcurriría uniformemente sin ninguna relación con ningún elemento externo y, en consecuencia, la diferencia entre el pasado y el futuro no poseería ningún significado, en la que, conocido el estado presente del fenómeno, se podría reproducir su pasado y calcular su futuro en forma determinística; es decir, en la que el presente no sería sino un punto que separa el pasado del futuro[70].

La insuficiencia de este modelo en términos de mecánica clásica para explicar "*...un mundo inestable que conocemos a través de una ventana finita*" (Prigogine; Nicolis (1977, p. 16), en el que el estado natural de los fenómenos es de no-equilibrio -un no-equilibrio constructivo que, como consecuencia de su propiedad fundamental de auto-organización, genera nuevos estados y nuevas estructuras complejas que sólo son imaginables en el ámbito de la irreversibilidad temporal- dio origen a una nueva formulación termodinámica -aleatorista-, cuya diferencia con la dinámica clásica radicó esencialmente: **i)** en la postulación del concepto de estado del proceso en un instante dado como resultante de una evolución orientada en el tiempo en la que, a diferencia del pasado y del presente, el futuro está formado por una sucesión de variables aleatorias no-observables vinculadas causalmente; **ii)** en la concepción de $f\left[\Omega^*\left(Y(t)\right)\right]$ como la representación de ciertas regularidades locales observadas y en la sustitución de la interpretación clásica de $\varepsilon(t)$ como azar-ignorancia (epistemológico), generado por los "errores" en la medición de las variables por la interpretación como azar-absoluto (ontológico), generado por dichos errores más las innovaciones a que están sometidos los factores incluidos en el sistema $\Omega\left(Y(t)\right)$.

[70]. Si se tiene en cuenta que la información infinita es indiscernible de ininteligibilidad, se puede concluir que, aún en una situación ideal en la que el observador contara con un conjunto de información de tamaño infinito, el problema de la inexplicabilidad del comportamiento de los fenómenos permanecería.

## 8.- La autorregresividad

A partir de esta asimilación entre fenómenos dinámicos y procesos estocásticos, la econometría evolucionó adoptando alternativamente interpretaciones determinísticas y aleatoristas de las perturbaciones. La econometría se basó inicialmente en una interpretación determinística del comportamiento de los fenómenos económicos, cuya hipótesis fundamental consistió en suponer que un modelo económico ($f[\Omega^*(Y_t)]$ $(t = 1,2,\dots)$) constituye una representación incompleta del comportamiento de un fenómeno y que$\varepsilon_t$representa una variable aleatoria no-observable definida por la agregación de los errores de medición que afectan a los factores incluidos en $\Omega^*(Y_t)$ y de las influencias que ejercen sobre dicho fenómeno los infinitos factores de su entorno económico no incluidos en $\Omega^*(Y_t)$[71].

Este punto de partida condujo con el tiempo a un proceso de formalización basado en el principio que la descripción de los aspectos cuantitativos de los fenómenos económicos a partir de conjuntos finitos de observaciones requería de la representación de sistemas estocásticos de variables interrelacionadas y, por lo tanto, de la teoría de la probabilidad para la estimación de dichas representaciones; es decir, para la construcción de modelos econométricos entendidos como el "*...vehículo útil para comparar la teoría económica con las observaciones, especificado de una forma medible y testeable*" (Qin (1993, p. 37) (ver Staehle (1933))[72].

Este proceso de formalización dio origen a varios problemas referidos: **i)** a la posibilidad de asegurar la adecuacidad de la representación a las relaciones teóricas; **ii)** a su identificación, es decir a la determinación de las condiciones que permitieran asegurar la unicidad de los coeficientes estimados y su interpretación, independientemente de las distribuciones de probabilidades consideradas y **iii)** a la modificación de la interpretación del término $\varepsilon_t$ asociado a la representación $f[\Omega^*(Y_t)]$y, en consecuencia, de la componente residual ($\varepsilon_t$) asociada con el modelo correspondiente[73].

---

71. El cambio de nomenclatura se debe a que, dado el carácter discreto de la información, las representaciones deben ser entendidas como aproximaciones discretas al comportamiento de los fenómenos en un espacio-tiempo continuo-continuo.

72. De acuerdo con la tradición el término "modelo" –entendido como estimador de una representación- fue introducido por Frisch en el Congreso de la Econometric Society en 1931 y fue utilizado por primera vez por Tinbergen, en 1936 (Magnus; Morgan (1987), Morgan (1989)).

73. El desarrollo de métodos de estimación de los coeficientes de representaciones conceptualmente válidas constituyó una cuestión fundamental en la evolución de la econometría.

Durante el período pre-modelístico las aplicaciones econométricas consistieron en el cálculo y la predicción de relaciones económicas, sin considerar la confiabilidad estadística de los resultados, como ya se comentó, dada la autorregresividad de las variables involucradas en la representación y la imposibilidad de asegurar la invariancia temporal de las condiciones que afectan el comportamiento del sistema, la teoría de la probabilidad era considerada como un argumento inapropiado para el análisis de los fenómenos económicos, de modo que, hasta la implementación del proceso de modelización, se utilizaron métodos de estadística matemática "sin probabilidad".

De acuerdo con esta concepción no-probabilística, las primeras interpretaciones atribuyeron la presencia de la componente aleatoria exclusivamente a "errores" en la medición de los factores incluidos en la representación ("*errores en las variables*", según la denominación de Schultz (1925) y Frisch (1934))[74]. Ahora bien, esta interpretación resultaba insuficiente para explicar los errores provenientes de la omisión de variables (en general, de mayor importancia que los errores en las variables) que requería el agregado de otra componente aleatoria, lo cual implicaba la adopción de métodos estadísticos más avanzados, cuyo desarrollo implicaba una identificación desagregada del origen y las características de los errores. En principio, en el análisis de los ciclos económicos, esta insuficiencia fue resuelta mediante la introducción del concepto de "*errores en las ecuaciones*" y una mayor profundización en el estudio de la naturaleza de la aleatoriedad inherente al comportamiento de los fenómenos.

Originalmente el análisis de los ciclos se debe a Moore (1914)(1923), quien propuso una representación utilizando análisis harmónico y series de Fourier. Alternativamente Persons (1916)(1922-23) desarrolló un método conocido como el "*barómetro de Harvard*", basado en la correlación entre series desfasadas un período (un modelo conocido hoy como "*indicador líder*"). Estos métodos perdieron vigencia ante las críticas de Slutsky (1927) y Yule (1921) (1926)(1927) referidas a la posible falta de significado conceptual de la correlación entre series cronológicas autorregresivas.

La propuesta de Yule se basó en el estudio de series perturbadas. Demostró que un esquema autorregresivo definido por un péndulo sometido a perturbaciones estocásticas genera un proceso "*armónico irregular*" y propuso una forma de

---

[74]. La idea de la posibilidad de existencia de errores de medición en las variables fue introducida por Gini (1921).

representación utilizando cuadrados mínimos en base a una ecuación en-diferencias de la forma $Y_t = \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2}$[75] más un "shock" aleatorio formado por la combinación de los errores y las perturbaciones que afectaban al comportamiento de $Y_t$. El hecho que los componentes de esta combinación fueran no-separables permite concluir que la solución de Yule no resolvía las cuestiones referidas a la forma en que las perturbaciones eran "absorbidas" por el sistema, ni a su interpretación conceptual.

## 9.- La descomposición estructural

Basándose en los trabajos de Wicksell (1907), Hotelling (1927) y Walker (1931), Frisch (1933)(1936) propuso el método de descomposición estructural[76], cuyo objetivo era sistematizar, en ciertas aplicaciones, las investigaciones sobre la estimación de representaciones utilizando series cronológicas mediante métodos de estadística matemática basados en argumentos probabilísticos. Introdujo el concepto de "*shocks erráticos*" para completar la caracterización de las variables $\varepsilon_t$. Las consideró como "*...desvíos del comportamiento estructural (...) como algo nuevo espontáneo agregado a la estructura del proceso*" (p. 408) que dan origen a la modificación aleatoria de las regularidades locales observables en su comportamiento y constituyen "*...la fuente de energía que generan los ciclos económicos*", incorporando de esta forma, parcialmente[77], la teoría de la probabilidad al pensamiento econométrico. Propuso, además, una clasificación de los "shocks" -de acuerdo con la forma en que son absorbidos por el sistema- en "*aberraciones*" y "*estímulos*", definiendo como estímulo a "*... una perturbación cuyo efecto perdura sobre estados sucesivos del sistema*" y como una aberración a "*... una perturbación que influye sobre el proceso solamente en el momento que se produce*" (p. 410), generada por el método de estimación utilizado[78].

---

75. Ecuación que puede ser considerada como el origen de las representaciones $AR$ .

76. Esta propuesta fue planteada por Frisch como una alternativa crítica al método utilizado habitualmente de la "*descomposición mecánica*" (en las componentes tendencia, estacionalidad, ciclo y variaciones no-sistemáticas).

77. Parcialemente en la medida que Frisch no estaba de acuerdo con la aplicación indiscriminada de laoría de la probabilidad en economía.

78. En particular, en este caso, como generadas por su método de estimación-identificación, al que denominó "*análisis de confluencia*", cuya finalidad era intentar resolver los problemas generados por la no consideración de relaciones ocultas para el observador, debido a los errores de medición en todas las variables (no solamente en la variable "dependiente", como postulaba el método de regresión utilizado habitualmente) y por la multicolinealidad (ver Hendry; Morgan (1989)).

De acuerdo con la tesis Fechneriana de la "*indeterminación por novedad*"[79] se puede considerar a la postulación de Frisch como el primer intento de interpretación de los estímulos como verdaderas innovaciones[80].

Por otra parte, a partir de los resultados de Slutsky acerca de la capacidad de los procesos aleatorios puros de generar procesos autorregresivos, concluyó "*...que las regularidades observadas en el comportamiento de $Y_t$ podían ser consideradas como derivadas de uncaos generado por elementos inconexos, por su propia inconexión*" (Frisch (1936, p. 107)) y clasificó a dichos elementos (a los cuales denominó en forma poco feliz como "*factores caóticamente aleatorios*")[81] en "*coherentes*" e "*incoherentes*" según que fueran o no autorregresivos y demostró -generalizando los trabajos de Yule (1921) sobre "*sumas ponderadas móviles*" de variables aleatorias independientes- que un proceso coherente podía ser aproximado como una combinación lineal de variables incoherentes de la forma:

$$Y_t = \theta_0 \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} (t = 1,2, \dots)$$

(donde las variables $Y_t$ son centradas y $\{\varepsilon_t\}$ es un ruido blanco.

J. Tinbergen (1935)(1937)(1938)(1939) realizó una notable síntesis de los métodos propuestos por R.A. Fisher y Koopmans y puede ser considerado el continuador inmediato de la obra de Frisch. Si bien coincidió con éste en las propiedades de las representaciones formadas por ecuaciones lineales ponderadas, abandonó su forma estructural combinada, compuesta por ecuaciones diferenciales y en-diferencias, a favor de ecuaciones discretas del tipo de las denominadas "*de rezagos distribuidos*" de la forma[82]:

---

[79]. Fechner (1866)(1871)(1906). Esta tesis se basa en el principio que en la evolución de los fenómenos se van generando nuevas condiciones iniciales las cuales, de acuerdo con una ley general, conducen a efectos no ocurridos previamente y, por lo tanto, no permiten la recurrencia de comportamientos idénticos. Pero que, no obstante, cada conjunto de dichas condiciones conserva alguna similitud con los anteriores por lo que, si bien todo fenómeno posee cierta dosis de aleatoriedad objetiva, su comportamiento no es completamente libre. Ver Landro (2010).

[80]. Frisch (1938): "*...los estímulos pueden ser considerados como condicionantes de la evolución posterior del fenómeno, es decir como causantes de una suerte de cambio permanente de sus condiciones iniciales*" (p. 408).

[81]. La calificación de "poco feliz" se basa en la aparente asimilación de los conceptos de caótico y aleatorio, de complejidad y aleatoriedad.

[82]. El concepto de "*rezagos distribuidos*" fue introducido por I. Fisher en 1920, en un trabajo sobre oferta monetaria (ver Alt (1942)).

$$Y_t = a_0 + b_0X_t + b_1X_{t-1} + \varepsilon_t^*$$

(donde los "shocks" $\varepsilon_t^*$ representan los estímulos que preservan la dinámica del sistema).

La generalización del pensamiento probabilístico que generó la propuesta pionera de Frisch -la cual influyó más en los aspectos metodológicos que en los epistemológicos de la teoría de modelos- se vio impulsada por los avances de la teoría de la inferencia, la axiomatización de la teoría de la probabilidad por obra de Kolmogorov (1933) y, posteriormente, por el desarrollo de la teoría de la decisión en condiciones de incertidumbre, la teoría del riesgo, el criterio de optimización mediante la maximización de la utilidad esperada y sus extensiones y por la aparición de la interpretación personalista del concepto de probabilidad por obra de Ramsey y de Finetti (ver Rowley; Hamouda (1987), Hamouda; Rowley (1988))[83]. En este sentido, merecen ser destacados los trabajos de Marschak (1937) -quien propuso la consideración del comportamiento de los operadores económicos como parte de la teoría de los juegos y, en consecuencia, la asimilación de los conceptos de función de preferencias y utilidad.

## 10.- La revolución probabilística

La culminación de la aproximación probabilística a la econometría se produjo con la llamada "*revolución Haavelmiana*" (Morgan (1987)(1989)). A partir de la propuesta esencialmente estocástica de Slutsky, de la postulación de Moore (1914) -reconsiderada por Schultz (1930)(1939)- sobre la utilización del análisis harmónico y de la aplicación rigurosa de la teoría de la probabilidad por Koopmans (1937), Haavelmo (1938)(1939)(1940a)(1940b)(1943)(1944) postuló que las ecuaciones estructurales debían ser interpretadas como leyes de comportamiento sólo en un sentido estadístico y que los modelos econométricos debían asumir la forma de las "*ecuaciones estructurales estocásticas*" de Frisch, incluyendo "*coeficientes estructurales estocásticos*". Esto lo condujo a la conclusión que la formalización de un procedimiento de modelización estocástica consistente y generalizado (basado en observaciones no-repetibles) requería inevitablemente de los métodos de máxima verosimilitud de Fisher y de la teoría de los tests de hipótesis de Neyman-Pearson, que permitieran decidir acerca de la correspondencia entre la interdependencia de las variables económicas y su tratamiento estadístico[84].

---

[83]. Ver Landro; González (2013)(2014).

[84]. En general la literatura considera que el punto de partida de la revolución Haavelmiana se

Haavelmo recurrió así al concepto de distribución conjunta de todas las variables observables como un argumento para justificar la forma de dicha interdependencia estadística y concluyó que para que esta solución fuera posible, era necesario atribuirle a los términos representativos de los errores ciertas distribuciones de probabilidades.

Esta forma de especificación mediante la definición de la distribución de probabilidades conjunta es considerada, en general, como el fundamento de la revolución probabilística Haavelmiana y de los modelos de ecuaciones simultáneas y determinó la diferencia "*...entre la construcción de modelos econométricos y la formulación matemática de la teoría económica*" (Qin (1993, p. 58)). Es decir determinó la diferencia entre los modelos econométricos y los modelos económicos.

Inmediatamente Mann; Wald (1943) aplicaron la especificación mediante la definición de la distribución conjunta y demostraron la consistencia y la Normalidad asintótica de los estimadores máximo-verosímiles en las ecuaciones lineales de la forma:

$$a_0 Y_t + a_1 Y_{t-1} + \cdots + a_j Y_{t-j} = \varepsilon_t$$

(donde $\{Y_t\}: I(0)$ y $\varepsilon_t$ denota un vector de perturbaciones no-autocorrelacionadas, idénticamente distribuidas y con momentos de orden superior finitos).

Como corolario de esta aplicación surgió el problema práctico de cómo asegurar la completitud del sistema de ecuaciones desde un punto de vista estadístico. Con este fin Koopmans (1950) redefinió el concepto de completitud de Tinbergen y, a partir de la clasificación de las variables en exógenas y endógenas, postuló que un modelo se puede considerar completo cuando el número de ecuaciones coincide con el número de variables endógenas del sistema. Definió como variable exógena aquella que no está afectada por las variables endógenas ni por las perturbaciones e introdujo el concepto de variable predeterminada que incluye a las variables desfasadas y a las variables exógenas correspondientes al período$t$. A partir de estas definiciones propuso una transformación del método de identificación basado en la distribución conjunta de las perturbaciones, en otro basado en el producto de la distribución de las perturbaciones condicionada por las variables

---

encuentra en la respuesta de Haavelmo (1943) (fundamentalmente en Schumpeter (1939) y en el debate Keynes-Tinbergen) a los prejuicios planteados con respecto a la naturaleza no-experimental de las observaciones económicas y, en consecuencia, a la capacidad de la inferencia estadística basada en la probabilidad para verificar las supuestas verdaderas relaciones propuestas por la teoría económica.

exógenas multiplicada por la distribución marginal de dichas variables exógenas.

Como una excepción a esa corriente de pensamiento que consideraba a la teoría de la probabilidad exclusivamente como el fundamento de los métodos estadísticos a aplicar en la construcción de los modelos estructurales, entendidos estos como representaciones (incompletas) de las verdaderas trayectorias de los fenómenos económicos más una componente representativa de los errores aleatorios, cabe mencionar los trabajos de Wold y Tintner.

## 11.- El teorema de la descomposición predictiva

El teorema de la descomposición predictiva de Wold (1938) -su tesis doctoral- resultó el hito fundamental en el análisis del tratamiento causal moderno de los fenómenos dinámicos que condujo a la solución de Kolmogorov (1941b)-Wiener (1942) al problema de la predicción lineal. Consideró a las series cronológicas como la porción observable de un proceso estocástico, es decir, como "*una realización de una distribución de infinitas dimensiones*" (p. 4) y completó las propuestas de Yule y Slutsky demostrando que todo proceso $\{Y_t\}: I(0)$ puede ser desagregado en una componente estructural aproximable por un $AR(p)$ $(p \gg 0)$ y una combinación lineal de infinitos "*errores*" incorrelacionados, asimilables a variables (latentes) no-observables:

$$Y_t = \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q}$$

(donde las variables $Y_t$ son centradas).

Cabe destacar que más allá que la condición de aleatoriedad pura implicaba suponer que las variables$\varepsilon_t$se distribuían de acuerdo con una función de densidad simétrica con valor esperado nulo y la condición de estacionariedad implicaba suponer que su varianza era finita e, ignorando la aproximación asintótica a la Normalidad que proporcionaba la versión de Laplace del teorema central del límite, Wold consideró "*insoportablemente arbitrario*" (Epstein (1987), p. 161) asignar una distribución de probabilidades particular a las variables $\varepsilon_t$.

Como corolario de este trabajo liminar sobre la teoría pura de los procesos de parámetro discreto y del análisis de regresión de H. Cramér y en oposición a los modelos de ecuaciones simultáneas, Wold desarrolló una aproximación recursiva de la forma[85]:

$$Y_t^{(i)} = F\left(Y_{t-1}^{(i)}, \dots, Y_{t-1}^{(i-1)}, \dots, Y_{t-2}^{(i)}, \dots, Y_{t-1}^{(n)}, Y_{t-2}^{(n)}, \dots\right) + \varepsilon_t^{(i)}$$

(Bentzel; Wold (1946)) y demostró (Wold (1949)(1954)) que este modelo proporcionaba una interpretación causal y que, en general, todo conjunto de series cronológicas puede ser representado formalmente como un sistema de encadenamiento causal ((1948)(1951))[86].

Sea $\{Y_t\}$ $(t = 1,2,\dots)$ un proceso estocástico unidimensional estacionario de segundo orden con valor esperado nulo y varianza $\sigma_Y^2$[87] y tal que, dado un valor fijo de $t$ y un $p$ entero, admite una representación lineal de la forma[88]:

$$\begin{aligned} Y_t &= a_1 Y_{t-1} + a_2 Y_{t-2} + \cdots + a_p Y_{t-p} + a_{p+1}\varepsilon_{t-p-1} + a_{p+2}\varepsilon_{t-p-2} + \cdots \\ &= Y_t^{(p)} + \varepsilon_t^{(p)} \end{aligned}$$

Donde:

**1)** $\left\{Y_t^{(p)}\right\}$ representa una aproximación finita al comportamiento que se debería esperar que asumiera $\{Y_t\}$ si dependiera exclusivamente de sus $p$ realizaciones pasadas (una aproximación finita óptima de orden $p$ a su estructura autorregresiva):

$$Y_t^{(p)} = E\left(Y_t / Y_{t-1}, Y_{t-2}, \dots, Y_{t-p}\right) = a_1 Y_{t-1} + a_2 Y_{t-2} + \cdots + a_p Y_{t-p}$$

Esta aproximación es única para cada sucesión $\left\{Y_{t-1}, Y_{t-2}, \dots, Y_{t-p}\right\}$ y es tal que converge en $L_2$ a la estructura autorregresiva:

$$\lim_{p\to\infty} E\left[\left(Y_t^{(\infty)} - Y_t^{(p)}\right)^2\right] = 0$$

---

85. Con respecto a esta ecuación, debe tenenerse en cuenta que, en el planteo dinámico de Wold las relaciones simultáneas no son justificables en la medida que no permiten establecer el sentido de la causalidad.

86. Wold (como un producto de la posición adoptada por la escuela sueca de economía) rechazó el paradigma Mashalliano-Walrasiano del equilibrio general como un estado persistente de la dinámica económica.

87 Dado el carácter discreto de la información que da origen a la definición débil de $\{Y_t\}$, las representaciones que se obtengan deben ser entendidas como aproximaciones discretas al comportamiento de los fenómenos en una espacio-tiempo continuo-continuo.

88 Es decir, un proceso caracterizado completamente por sus momentos de primero y segundo orden.

donde $Y_t^{(\infty)} = E(Y_t/Y_{t-1}, Y_{t-2}, \dots)$. Lo cual permite concluir que, a partir de una interpretación determinística de la autorregresividad de $\{Y_t\}$, el planteo de Wold considera que existe una representación $AR$ de orden finito que describe su verdadera estructura autorregresiva y que la representación $AR(\infty)$ surge sólo como un corolario del teorema. Es decir que, de acuerdo con una lectura estricta de sus postulados, el teorema conduce a la tesis de Church-Turing, la cual se funda en la hipótesis que la estructura autorregresiva de $\{Y_t\}$ involucra más vínculos que los que puede contener cualquier expresión sintáctica de su comportamiento, en otros términos, que su "tiempo de coherencia" es infinito y, por lo tanto, que su "verdadera" representación autorregresiva es de orden infinito (y que la única representación verdadera de orden finito se verificará si, y sólo si $\{Y_t\}: AR(0)$)[89].

**2)** $\{\varepsilon_t\}$ define, en la nomenclatura de Wod, el denominado proceso de los "errores, que representa la parte de $Y_t$ "no explicada" por la combinación lineal de las variables $Y_{t-1}, Y_{t-2}, \dots, Y_{t-p}$ y satisface las siguientes condiciones:

**i)** $E\left(\varepsilon_t^{(p)}\right) = 0 \ (t = 1,2, \dots)$

**ii) de ortogonalidad**:

$$E\left(Y_{t-j}\varepsilon_t^{(p)}\right) = E\left[Y_{t-j}\left(Y_t - Y_t^{(p)}\right)\right] = E\left(Y_{t-j}Y_t - \sum_{i=1}^{p} a_i Y_{t-i} Y_{t-j}\right)$$

$$= \gamma_j(Y) - \sum_{i=1}^{p} a_i \gamma_{|j-i|}(Y) = 0 \qquad (j = 1,2, \dots, p)$$

Obsérvese que se podría haber arribado al mismo resultado de una forma exclusivamente conceptual, teniendo en cuenta que, si no se cumpliera esta condición, es decir, si existiera algún tipo de relación entre las variables $\varepsilon_t$ y las variables $Y_t$ anteriores, entonces parte del comportamiento de las perturbaciones podría ser "explicado" por duchas variables pero, en ese caso los coeficientes $a_j$ no cumplirían la condición de minimizar $E\left[\left(Y_t - Y_t^{(p)}\right)^2\right]$ y el comportamiento estructural $E\left(Y_t/Y_{t-1}, Y_{t-2}, \dots, Y_{t-p}\right)$ perdería la propiedad de unicidad.

[89] Ver González; Landro (2018).

**iii)** $E\left(\varepsilon_{t-j}^{(p)}\varepsilon_t^{(p)}\right) = \gamma_j\left(\varepsilon^{(p)}\right) = E\left[\left(Y_{t-j} - Y_{t-j}^{(p)}\right)\varepsilon_t^{(p)}\right] = E\left(Y_{t-j}\varepsilon_t^{(p)}\right) - E\left(Y_{t-j}^{(p)}\varepsilon_t^{(p)}\right) =$

$$= E\left(Y_{t-j}\varepsilon_t^{(p)}\right) - E\left[\left(a_1Y_{t-1} + a_2Y_{t-2} + \cdots + a_pY_{t-p}\right)\varepsilon_t^{(p)}\right] = 0$$

**iv)** $E\left[\left(\varepsilon_t^{(p)}\right)^2\right] \equiv \sigma^2\left(\varepsilon_t^{(p)}\right) = \sigma_\varepsilon^2.$

**v)** $\varepsilon_t^{(p)}$ converge en $L_2$ a $\varepsilon_t^{(\infty)} = Y_t - Y_t^{(\infty)}$, $\lim_{p\to\infty} E\left[\left(\varepsilon_t^{(\infty)} - \varepsilon_t^{(p)}\right)^2\right] = 0.$

De las condiciones **i)**, **ii)** y **iv)** se sigue que el espectro de $\left\{\varepsilon_t^{(p)}\right\}$ es constante en el intervalo $(-\pi, \pi)$ y que, por lo tanto, es un WN-débil[90]. En este sentido, cabe destacar que, más allá que el reconocimiento que la condición de aleatoriedad pura implica suponer que las variables $\varepsilon_t$ se distribuyen de acuerdo con un función de densidad simétrica con valor esperado nulo y que la condición de estacionariedad implica suponer que su varianza es finita, curiosamente Wold no tuvo en cuenta la aproximación asintótica a la Normalidad que proporciona la versión de Laplace del teorema central del límite, lo que lo condujo a considerar como "*insoportablemente arbitrario*" (Epstein (1987), p. 161) asignarles una distribución de probabilidades particular.

En su críptica nomenclatura original Wold (1938) denominó a $\left\{Y_t^{(p)}\right\}$ como la "*componente singular*" y a $\left\{\varepsilon_t^{(p)}\right\}$ como la "*componente regular*", ignorando el concepto de "*shock errático*" introducido por Frisch (1933)(1936), quien consideró a las variables $\varepsilon_t$ como "*desvíos el comportamiento estructural (...) como algo nuevo, espontáneo agregado a la estructura* (autorregresiva) *del proceso*" (p. 408).

Tampoco consideró la tesis Fechneriana de la "*indeterminación por novedad*", según la cual cada innovación $\varepsilon_t$ representa una nueva información que actualiza el valor del proceso para cada valor de $t$ y que, por lo tanto, replantea para cada $t$ nuevas condiciones iniciales, las cuales conducen a efectos no ocurridos previamente y, por lo tanto, no permiten la recurrencia de comportamientos idénticos. Una interpretación que marca la diferencia entre un proceso autorregresivo y una ecuación en diferencias.

---

[90] Si bien la calificación de $\left\{\varepsilon_t^{(p)}\right\}$ como un "ruido blanco" pertenece al ámbito de los procesos que admiten representación lineal, puede admitir una estructura no-lineal y la descomposición propuesta por el teorema de Wold es aplicable aún a procesos caóticos de baja dimensión.

**3)** Los coeficientes $a_j$ se obtienen de minimizar la expresión:

$$E\left[\left(\varepsilon_t^{(p)}\right)^2\right] = E\left[\left(Y_t - Y_t^{(p)}\right)^2\right]$$
$$= E\left(\sum_{i=1}^{p}\sum_{j=1}^{p} a_i a_j Y_{t-i} Y_{t-j} - 2\sum_{i=1}^{p} a_j Y_t Y_{t-j} + Y_t^2\right)$$

Es decir, suponiendo que la solución exista, los coeficientes son de la forma:

$$a = \left[\Gamma_p(Y)\right]^{-1}\gamma(Y)$$

Siendo:

**i)** $a' = \left[a_1 \quad a_2 \quad \cdots \quad a_p\right]$

**ii)** $\gamma'(Y) = [\gamma_0(Y) \quad \gamma_1(Y) \quad \cdots \quad \gamma_{p-1}(Y)]$

**iii)** $\gamma_j(Y) = E\left(Y_{t-j}Y_t\right) = a_1\gamma_{|j-1|}(Y) + a_2\gamma_{|j-2|}(Y) + \cdots + a_p\gamma_{|j-p|}(Y) \qquad (j = 0,1,2,\dots,p)$

**iv)** $\Gamma_p(Y) = \begin{bmatrix} \gamma_0(Y) & \gamma_1(Y) & \cdots & \gamma_{p-1}(Y) \\ \gamma_1(Y) & \gamma_0(Y) & \cdots & \gamma_{|p-2|}(Y) \\ \cdots & \cdots & \cdots & \cdots \\ \gamma_{p-1}(Y) & \gamma_{|p-2|}(Y) & \cdots & \gamma_0(Y) \end{bmatrix}$ denota la matriz de autocovarianzas de Toeplitz. La cual, dado que se ha supuesto que el proceso $Y_t$ es estacionario, es simétrica con respecto a ambas diagonales.

Debe tenerse en cuenta que suponer que $\gamma_j(\varepsilon) \neq 0$ (es decir, suponer que la condición **iii)** no se cumple) implicaría reconocer que $\{\varepsilon_t\}$ posee una estructura autorregresiva:

$$E\left(\varepsilon_t / \varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_{t-p}\right) = c_1\varepsilon_{t-1} + c_2\varepsilon_{t-2} + \cdots + c_p\varepsilon_{t-p}$$

Pero, en ese caso, los coeficientes $a_j$ no cumplirían la condición de minimizar $E(\varepsilon_t^2)$.

Suponiendo el cumplimiento de estas condiciones, con un razonamiento similar al de la página precedente, se puede escribir:

$$Y_{t-1} = b_2 Y_{t-2} + b_3 Y_{t-3} + \cdots + b_{p+1} Y_{t-p-1} + \varepsilon_{t-1}^{(p)} = Y_{t-1}^{(p)} + \varepsilon_{t-1}^{(p)}$$

Donde los coeficientes $b_j$ son tales que minimizan la expresión $E\Big[\Big(Y_{t-1} -$

$Y_{t-1}^{(p)}\Big)^2\Big]$ y se verifica que: **i)** $E\left(\varepsilon_{t-1}^{(p)}\right) = 0$; **ii)** $E\left(Y_{t-j}\varepsilon_{t-1}^{(p)}\right) = 0 \;\; (j = 2,3,\dots)$ y **iii)**

$$E\left(\varepsilon_{t-1}^{(p)}\varepsilon_t^{(p)}\right) = E\left[\left(Y_{t-1} - Y_{t-1}^{(p)}\right)\varepsilon_t^{(p)}\right] = E\left(Y_{t-1}\varepsilon_t^{(p)}\right) - E\left(Y_{t-1}^{(p)}\varepsilon_t^{(p)}\right) = 0$$

A partir de estas condiciones se puede escribir:

$$Y_t = \varepsilon_t^{(p)} + a_1\left(Y_{t-1}^{(p)} - \varepsilon_{t-1}^{(p)}\right) + a_2Y_{t-2} + \cdots + a_pY_{t-p} =$$

$$= \varepsilon_t^{(p)} + a_1\varepsilon_{t-1}^{(p)} + a_1\left(b_2Y_{t-2} + b_3Y_{t-3} \dots + b_pY_{t-p}\right) + a_2Y_{t-2} + \cdots + a_pY_{t-p} =$$

$$= \varepsilon_t^{(p)} + a_1\varepsilon_{t-1}^{(p)} + a_2^*Y_{t-2} + \cdots + a_p^*Y_{t-p} + a_1\varepsilon_{t-1}^{(p)} + Y_{t-2}^{(p)}$$

Operando en forma iterativa se concluye que:

$$Y_t = \varepsilon_t^{(p)} + b_1^*\varepsilon_{t-1}^{(p)} + b_2^*\varepsilon_{t-2}^{(p)} + \cdots + b_j^*\varepsilon_{t-j}^{(p)} + Y_{t-j-1}^{(p)} = \eta_{t-j}^{(p)} + Y_{t-j-1}^{(p)}$$

Donde:

**i)** $Y_{t-j-1}^{(p)} = a_jY_{t-j-1}^{(p)} + a_{j+1}Y_{t-j-2}^{(p)} + \cdots + a_pY_{t-p-1}^{(p)}$

**ii)** $\gamma\left(\varepsilon_{t-h}^{(p)}\varepsilon_t^{(p)}\right) = E\left[\varepsilon_t^{(p)}\left(Y_{t-h} - a_1Y_{t-h-1} - \cdots - a_pY_{t-h-p}\right)\right] = 0 \quad (h > 0)$

**iii)** $E\left[\left(\varepsilon_t^{(p)}\right)^2\right] = \sigma_\varepsilon^2(p)$

**iv)** $\gamma\left(\varepsilon_{t-h}^{(p)}Y_{t-j-1}^{(p)}\right) = 0 \quad (h \le j)$

**v)** $E\left[\left(Y_{t-i}^{(p)}\right)^2\right] \ge E\left[\left(Y_{t-j}^{(p)}\right)^2\right] \quad (i > j; i,j = 0,1,2,\dots)$

**vi)** $b_j^* = \frac{E\left(Y_t\varepsilon_{t-j}^{(p)}\right)}{E\left[\left(\varepsilon_{t-j}^{(p)}\right)^2\right]}$

De acuerdo con la condición de ortogonalidad se demuestra, entonces, que:

$$E\left(\eta_{t-j}^{(p)}Y_{t-j-1}^{(p)}\right) = 0 \qquad (j = 0,1,2,\dots,p-1)$$

y que:

$$E\left[\left(Y_t-\eta_{t-j}^{(p)}\right)^2\right]=E\left[\left(Y_t-\sum_{j=0}^{q}b_j^*\varepsilon_{t-j}^{(p)}\right)^2\right]$$

$$=E(Y_t^2)+E\left(\sum_{j=0}^{q}b_j^{*2}\varepsilon_{t-j}^{(p)2}\right)-2\sum_{j=0}^{q}b_j^*E\left(Y_t\varepsilon_{t-j}^{(p)}\right)=$$

$$=\ E(Y_t^2)+\sigma_\varepsilon^2\sum_{j=0}^{q}\frac{1}{\sigma_\varepsilon^4}\left[E\left(Y_t\varepsilon_{t-j}\right)\right]^2-2\,\sigma_\varepsilon^2\sum_{j=0}^{q}\frac{1}{\sigma_\varepsilon^4}\left[E\left(Y_t\varepsilon_{t-j}\right)\right]^2=$$

$$=\ E(Y_t^2)+\sigma_\varepsilon^2\sum_{j=0}^{q}b_j^{*2}\geq 0$$

Dado que $\sigma_Y^2<\infty$ es condición necesaria y suficiente para poder asegurar que $\{Y_t\}\!:I(0)$, se puede concluir que $\sigma_\varepsilon^2\sum_{j=0}^{q}b_j^{*2}\leq\sigma_Y^2$ (para todo $q$) y, por lo tanto, que $\lim_{q\to\infty}\sum_{j=0}^{q}b_j^{*2}<\infty$. En consecuencia, para todo $q^*>q$, será:

$$E\left[\left(\sum_{j=0}^{q}b_j^*\varepsilon_{t-j}^{(p)}-\sum_{j=0}^{q^*}b_j^*\varepsilon_{t-j}^{(p)}\right)^2\right]=\sum_{j=q+1}^{q^*}b_j^{*2}\varepsilon_{t-j}^{(p)2}=\sigma_\varepsilon^2\sum_{j=q+1}^{q^*}b_j^{*2}$$

De modo que siempre es posible elegir un valor de $q$ suficientemente grande como para que $\sum_{j=q+1}^{q^*}b_j^{*2}$ sea tan pequeño como se quiera y, en consecuencia, se puede afirmar que $\eta_t^{(q)}$ converge en $L_2$ a $\eta_t$.

De todo lo anterior se puede concluir que el proceso $\{Y_t\}$ admite una expresión como la suma de dos procesos estacionarios ortogonales, $Y_t=\eta_t+Y_{t-\infty}^{(p)}$ (donde $\{\eta_t\}$ es una combinación lineal de innovaciones puramente no-determinísticas y $\left\{Y_{t-\infty}^{(p)}\right\}$ representa la estructura autorregresiva que depende del pasado remoto del proceso[91].

---

[91] De acuerdo con la generalización del teorema de Wold debida a Cramér (1961), se demuestra que aún si los coeficientes $b_j$ son variables en el tiempo ($b_j(t)$) y se verifica la condición $\sum_{j=0}^{\infty}b_j^2(t)<\infty$ (para todo $t$), cualquier proceso $\{Y_t\}$ admite una descomposición única de la forma $Y_t=\eta_t+Y_{t-\infty}^{(p)}$. Además, para procesos multidimensionales se demuestra que un proceso $Y_t=[Y_{1t}\quad Y_{2t}\quad \dots\quad Y_{kt}]$ admite una descomposición de la forma $Y_t=b(B)\varepsilon_t$, donde $b(B)$ denota una matriz de orden $(k\times k)$ cuyos elementos son $b_{ij}(B)=\sum_{h=0}^{\infty}b_{ij,h}B^h$ y $\varepsilon_t=[\varepsilon_{1t}\quad\varepsilon_{2t}\quad\dots\quad\varepsilon_{kt}]$ denota un vector WN tal que $E(\varepsilon_{jt})=0$ ($j=1,2,\dots,k$) y $E(\varepsilon_{jt}\varepsilon_{j,t-h})=0$

Se puede asegurar, además, que esta descomposición aditiva es única en la medida que existe sólo una sucesión $\{b_j\}$ y una sucesión de variables aleatorias $\{\varepsilon_t\}$ que satisfacen las condiciones analizadas precedentemente.

Por otra parte, teniendo en cuenta que la predicción del proceso $\{Y_t\}$ con un horizonte $k$, utilizando un predictor definido como la convolución:

$$f^{(k,t-k)} = E(Y_t/Y_{t-k}\,, Y_{t-k-1}, \dots) = \sum_{j=k}^{\infty} a_j Y_{t-j}$$

minimiza la varianza del error de predicción:

$$\sigma^2[e(k,t)] = E\left\{\left[Y_t - f^{(k,t-k)}\right]^2\right\} = \frac{1}{2\pi}\int_{-\pi}^{\pi} \left|e^{iwk} - K(w)\right|^2 dA(w)$$

(donde $A(w)$, que representa la función espectral, es una función real, monótona no-decreciente y tal que $A(0) = 0$ y $A(\pi) = \pi$), se puede concluir que la estructura autorregresiva puede ser predicha mediante una aproximación arbitraria en $L_2$. Lo que generaliza la ´propuesta de Wold respecto de la optimización por una estructura $AR$ de orden infinito.

En particular, si $\{Y_t\}$ es un proceso ergódico, su función de autocovarianzas:

$$\gamma_j(Y) = \lim_{n\to\infty} \frac{1}{2n+1} \sum_{j=-\infty}^{\infty} Y_{t+j} Y_t$$

Es definida no-negativa y, en consecuencia, admite una representación mediante una transformación de Fourier de la forma:

$$\gamma_j(Y) = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{iwj} dA(w) = \frac{1}{\pi}\int_{0}^{\pi} cos(wj) dA(w)$$

y si $\sum_{j=-\infty}^{\infty}\left|\gamma_j(Y)\right| < \infty$, entonces $A(w)$ es absolutamente continua y está definida por la fórmula de inversión:

$$A'^{(w)} = \sum_{j=-\infty}^{\infty} \gamma_j(Y) e^{iwj} = \gamma_0(Y) + 2\sum_{j=1}^{\infty} \gamma_j(Y) cos(wj)$$

Asimismo, que su distribución espectral sea absolutamente continua es

---

(para $h = 1,2, \dots$) (ver Hannan (1970)).

condición necesaria y suficiente para poder asegurar que admite una representación $MA(\infty)$ de la forma[92]:

$$Y_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} = \Psi(B)\varepsilon_t \qquad (\psi_0 = 1)$$

Es decir, para poder asegurar que $\{Y_t\}$ pueda ser considerado como el resultado de la aplicación de un filtro con función de transferencia $\Psi(B)$, en la cual la sucesión $\{\psi_0, \psi_1, \psi_2, \dots\}$ define la función de respuesta del proceso a una innovación y $\{\varepsilon_\tau\}: WN$ $(\tau = 0,1,2, \dots, t)$ constituye una sucesión de innovaciones no-autocorrelacionadas, cuya contribución sobre la realización $Y_t$ será de magnitud $\psi_j \varepsilon_{t-j}$ y $\Psi(B)\varepsilon_t$ define la contribución agregada de todas la innovaciones $\varepsilon_\tau$ $(\tau = 0,1,2, \dots, t)$ a la formación de la trayectoria $\{Y_t\}$.

En términos más generales, se demuestra que todo proceso $\{Y_t\}: I(0)$ cuyo espectro sea una función racional en $z = e^{-iw}$, es un híbrido entre un proceso $MA(q)$ y un proceso $AR(p)$ con función de densidad espectral de la forma:

$$A'^{(z)} = \frac{\left|\sum_{j=0}^{p} g_j e^{-iwj}\right|^2}{\left|\sum_{j=0}^{q} h_j e^{-iwj}\right|^2} = \frac{G(w)G(-w)}{H(w)H(-w)}$$

(donde los polinomios $G(w)$ y $H(w)$ no poseen factores comunes, $H(w)$ posee sólo raíces en módulo mayores que la unidad y $G(w)$ posee raíces en módulo mayores o iguales que la unidad).

Luego, se puede concluir que un proceso $\{Y_t\}: ARMA(p,q)$ puede ser considerado como el resultado de someter a $\{\varepsilon_\tau\}: WN - débil$ a un filtro lineal con función de transferencia $B(w) = \frac{G(w)}{H(w)}$. Equivalente a someter a $\{\varepsilon_\tau\}$ a un filtro $G(w)$ lineal y someter el resultado de esta operación a un filtro lineal $[H(w)]^{-1}$ de una representación $AR(p)$ con función de respuesta a una innovación $\{f_0, f_1, f_2, \dots\}$, la cual es la inversa del operador $\{h_0, h_1, h_2, \dots\}$. De modo que la descomposición predictiva de $Y_t$ puede ser expresada de la siguiente forma:

$$Y_t = \Psi(B)\varepsilon_t = \sum_{j=0}^{\infty} f_j \sum_{h=0}^{\infty} g_h \varepsilon_{t-j-h} = \sum_{j=0}^{\infty} \left( \sum_{h=0}^{\infty} g_h f_{j-h} \right) \varepsilon_{t-j}$$

[92] Ver Kolmogorov (1941a), Doob (1953).

o, lo que es lo mismo, como:

$$\sum_{j=0}^{p} \phi_j Y_{t-j} = \sum_{j=0}^{q} \theta_j \varepsilon_{t-j} \qquad (\phi_0 = \theta_0 = 1)$$

(donde $\{\varepsilon_\tau\}: WN - débil$ representa el "input" del sistema y $\{Y_t\}$ representa el "output"[93].

La importancia fundamental de estos resultados está evidenciada en el hecho que todo el desarrollo posterior de la teoría econométrica de los fenómenos dinámicos se basó en corolarios y generalizaciones del teorema de Wold.

## 12.- La teoría de modelos después del teorema de Wold

Contrariamente a la posición de Wold -quien justificó la utilización de la teoría pura de la probabilidad en las representaciones de fenómenos económicos basándose en elementos abstractos inherentes a la definición clásica, Tintner (1938a)(1938b) planteó la posibilidad de utilizar a la probabilidad, de acuerdo con una interpretación más amplia y más flexible, como un argumento para vincular "*...la teoría económica de las expectativas y la teoría estadística de los errores*" (1938b, p. 154)[94]. Consideró a los errores aleatorios como generados por la imposibilidad de obtener "*...las representaciones óptimas de los factores controlables*" y de predecir "*...los factores no-controlables*", clasificándolos como errores en las variables y errores debidos a la influencia de fluctuaciones aleatorias en la descomposición de los sistemas de series cronológicas y postuló que los primeros debían ser tratados por el método de la diferencia variable y los segundos por los métodos postulados por Wold. En particular, su propuesta se basó en la interpretación Carnapiana, la cual admite dos conceptos posibles de probabilidad: $P_{r_1}$relacionado con el "*grado de confirmación*" y $P_{r_2}$, relacionado con la "*frecuencia empírica*" y concluyó que la primera definición podía constituir el fundamento de una teoría de los tests de hipótesis apropiada para la econometría[95].

En desacuerdo con la propuesta de Wold, según la cual $\{\varepsilon_t\}: WN - débil$,

---

[93] Doob (1944) generalizó este teorema al caso multidimensional (ver Hannan (1970)).

[94]. Debe tenerse en cuenta que, dado que Wold no consideró la necesidad de testear las representaciones propuestas respecto de su adecuacidad a las correspondientes posiciones teóricas, no reparó en las condicióones de aplicabilidad del concepto de probabilidad. Su propuesta postulaba que los sistemas recursivos no estaban afectados por los problemas de identificación.

[95]. Ver Landro (2010).

Marschak (1953) justificó la posible presencia de autocorrelaciones entre las perturbaciones y postuló que"...*deben ser consideradas como parte del comportamiento estructural, en la medida que no hay ninguna razón económica para descartar la posibilidad de que formen un proceso estocástico en el que cada shock depende de uno o más de sus predecesores*" (p. 21). Esta propuesta hizo que la investigación econométrica se dirigiera (fundamentalmente por obra de Cochran; Orcutt (1949), Durbin (1960), Sargan (1959)(1961)) al estudio de las implicaciones producidas por las autocorrelaciones entre las perturbaciones sobre las propiedades de los estimadores de los coeficientes estructurales.

Muth, en su trabajo fundamental de 1961, propuso un retorno a la interpretación de Frisch (1933) de los estímulos en el comportamiento estructural de un sistema como una caracterización de los efectos dinámicos de los "shocks" aleatorios autocorrelacionados (pero económicamente no especificados), débilmente estacionarios de segundo orden, asimilables a variables exógenas y representables como una combinación lineal de variables$\varepsilon_t$no-autocorrelacionadas con valor esperado nulo y varianza constante ($Y_t = f[\Omega^*(Y_t)] + \varepsilon_t^*$ donde $\{\varepsilon_t\}: WN$)[96].

En este punto la hipótesis de la ortodoxia econométrica original, que sostenía que siempre era posible especificar un modelo teórico que explicara la verdadera ley que rige el comportamiento de un fenómeno[97], comenzó a debilitarse y surgió la necesidad de modificaciones "ad hoc" en las especificaciones de los modelos estructurales para representar las eventuales regularidades locales, de interpretar a los modelos como hipótesis generales sobre las principales relaciones causales sugeridas por la teoría económica, testeables a partir de las observaciones. En este contexto de un naciente paradigma aleatorista, Theil (1957)(1958), a partir del supuesto que "...*los modelistas en general no conocen la verdadera especificación de una representación*" (Theil (1958, p. 215)) (cabría agregar, suponiendo que ésta exista), propuso una definición alternativa de "*especificación*" y, por lo tanto, de capacidad de predicción, basada en la minimización de la varianza residual

[96].A diferencia de Frisch (1937) -quien utilizó una representación $MA$ para descomponer una variable económica- Muth utilizó una representación $MA$ para descomponer un "shock" aleatorio.

[97]. Entendida la especificación como "...*la selección de la forma matemática de la población*" (Koopmans (1937, p. 3)).

como criterio de optimización para la selección de un modelo[98], pero sin asignarle ninguna interpretación conceptual a los residuos.

En general en el paradigma estructuralista los principales esfuerzos estuvieron dirigidos a la identificación y estimación de los coeficientes estructurales a partir de las formas reducidas, dedicándole muy poca atención a la verificación del cumplimiento de las condiciones impuestas a los "shocks" exógenos (Granger; Newbold (1977): "*Los residuos son tratados habitualmente por los econometristas como meras molestias de poca importancia. Muchos textos de econometría introducen a los modelos como un conjunto de relaciones determinísticas entre variables y adicionan descuidadamente términos representativos del 'error' a las ecuaciones para explicar cosas como errores en la especificación del modelo y en la medición de las variables. La utilización de denominaciones como 'residuos' y 'errores' implica juicios de valor sobre la importancia de dichos términos. Si bien 'error' es una denominación adecuada en el contexto de la predicción, una denominación más apropiada podría ser 'innovaciones*'" (p. 8)).

Un breve retorno al determinismo en este transcurso hacia el aleatorismo lo constituye la obra de Leamer (1978), quien reinterpretó la propuesta de Theil sobre el problema de la especificación desde una perspectiva Bayesiana[99]. Consideró a las leyes económicas como el fundamento teórico de las representaciones estructurales y propuso considerar a los errores como independientes de esa "verdadera" formulación teórica generados por la agregación de elementos "*no-observables*" pero "*manejables*", debidos exclusivamente a deficiencias en la especificación de la representación[100] y concluyó, en consecuencia, que una representación "completa" implicaría la eliminación de las perturbaciones[101].

---

[98]. Es necesario tener en cuenta que "*...dado que este criterio no siempre es concluyente ni factible, deben considerarse dos criterios adicionales de naturaleza más subjetiva: plausibilidad y simplicidad*" (Theil (1958, p. 208)).

[99]. La diferencia fundamental entre Theil y los econometristas clásicos y Leamer y los econometristas Bayesianos radica en que estos estaban más interesados en el análisis de las propiedades de los estimadores de los coeficientes que en la minimización de la varianza residual de los clásicos.

[100]. Ver Qin; Gilbert (2001).

[101]. Una interpretación que puede considerarse resumida en la expresión atribuida a Tukkey,

Dadas las debilidades evidenciadas por el estructuralismo en la formulación dinámica de las representaciones y a partir de extensiones del teorema de Wold, Sargent (1976)(1977), Sargent; Sims (1977) y Sims (1980) impulsaron una aproximación a la explicación del comportamiento de los fenómenos económicos mediante modelos puros de series cronológicas. En particular, propusieron la utilización de representaciones de la forma $\Phi(B)Y_t = \varepsilon_t$, en las cuales el orden $p$ del operador $\Phi(B)$ fuera tal que $\{\varepsilon_t\}$ pudiera ser considerado un proceso de innovaciones no-autocorrelacionadas, de modo que esta condición de $AR(0)$ del $\{\varepsilon_t\}$ procesose verificara por construcción y no por una hipótesis "a priori" y la representación $\Phi(B)Y_t = \varepsilon_t$ constituyera una caracterización económicamente válida del comportamiento de $Y_t$.

La representación inversa, $Y_t = [\Phi(\mathrm{B})]^{-1}\varepsilon_t$ retorna, en una interpretación aleatorista, a la propuesta de Slutsky-Wold de representación lineal de un proceso "*coherente*" ($\{Y_t\}$) en términos de un proceso "*incoherente*" ($\{\varepsilon_t\}$). En la que el operador $[\Phi(\mathrm{B})]^{-1}$, como función que describe la medida del impacto de los "shocks" aleatorios, constituye el argumento más importante para considerar a las perturbaciones como innovaciones y comprender cómo estas innovaciones son asimiladas por el proceso $\{Y_t\}$.

## 13.- Conclusiones: A modo de resumen

La primera conclusión a la que permiten arribar esta escueta reseña sobre el desarrollo de la teoría de la probabilidad y su influencia en la explicación del comportamiento de los fenómenos económicos, es que el origen de la teoría de los modelos estocásticos dinámicos se remonta a una época muy anterior a la que propone, en general, la literatura econométrica. Que su génesis puede considerarse asociada: **i)** a los postulados del teorema de la inversión de la probabilidad de J. Bernoulli (1713), que planteó la necesidad de hallar el nexo entre las probabilidades "a priori" (definidas a partir de un razonamiento que va de las causas a los efectos) y las probabilidades "a posteriori" (definidas a partir de un razonamiento que va de los efectos a las causas) y, en particular, **ii)** a la novedad conceptual de la solución propuesta por Simpson, relacionada con la definición de una distribución de probabilidades de los errores de observación aleatorios y la obtención, por añadidura, de la inversión de la probabilidad.

---

según la cual "*El hombre construye* $f\left[\Omega^*(Y_t)\right]$ *y Dios nos proporciona* $\varepsilon_t$ ".

La solución de Simpson condujo al criterio de los cuadrados mínimos de Legendre el cual, provisto de una interpretación probabilística tuvo su culminación en la síntesis de Gauss-Laplace.

La adopción de este criterio de optimización basado en la minimización de la suma de los cuadrados de los errores aleatorios y en la Normalidad de su distribución, y el surgimiento de un nuevo método de inferencia estructurado a partir de conjuntos de observaciones no homogéneas, dio origen a la aplicación de la teoría de los modelos estocásticos a la econometría.

A partir de los postulados de la versión de Laplace del teorema central del límite y de la asimilación de fenómenos dinámicos a procesos estocásticos la econometría evolucionó adoptando, alternativamente, interpretaciones determinísticas y aleatoristas de las perturbaciones.

El análisis de la interpretación econométrica del concepto de error aleatorio comenzó con Yule, quien propuso una forma de representación a partir de un esquema autorregresivo e introdujo el concepto de "shock" aleatorio formado por la combinación de errores y perturbaciones que afectaban el comportamiento del sistema. Pero, que los componentes de esta combinación fueran no-separables, permitió arribar a una segunda conclusión: que esta solución no logró resolver el problema de la interpretación conceptual de las perturbaciones ni a la forma en que eran "absorbidas" por el sistema

Fue Frisch quien propuso una primera solución a esta cuestión considerando a los "shocks" como "*algo nuevo y espontáneo agregado a la estructura del proceso*", que da origen a la modificación aleatoria de las regularidades locales observables y constituye "*...la fuente de energía que generan los ciclos económicos*", incorporando de esta forma, parcialmente, la teoría de la probabilidad al pensamiento econométrico. Esta interpretación condujo a una tercera conclusión: que, de acuerdo con la tesis de la "*indeterminación por novedad*" se puede considerar a la postulación de Frisch como el primer intento de interpretación de los "estímulos" como verdaderas innovaciones Fechnerianas.

En oposición a la propuesta de Haavelmo referida a la construcción de modelos de ecuaciones simultáneas, cabe mencionar los trabajos de Wold y Tintner, basados en un sistema de encadenamiento causal. En particular, el teorema de la descomposición predictiva de Wold permitió arribar a una cuarta conclusión: que esta propuesta constituye la piedra angular sobre la que se construyó todo el análisis del tratamiento moderno de los fenómenos dinámicos, cuyas generalizaciones generaron todo el desarrollo posterior de la teoría econométrica.

El estudio de la evolución de los métodos econométricos demuestra que el teorema de la descomposición predictiva constituyó la base sobre la cual se desarrolló todo el tratamiento causal de los fenómenos dinámicos y, en particular, la solución de Kolmogorov-Wiener del problema de la predicción lineal.

Como complemento a los aspectos teóricos que involucra este teorema y en reconocimiento a ese rol protagónico en la construcción de la moderna teoría económica, en este Cuaderno se intentó un análisis de la interpretación de Wold de las representaciones $ARMA$ y de la naturaleza de las perturbaciones. Esta revisión permitió concluir: **i)** que, a partir de una interpretación determinística de la autorregresividad de $\{Y_t\}$, el planteo de Wold considera que existe una representación $AR$ de orden finito que describe su verdades estructura autorregresiva y que la representación $AR(\infty)$ surge como un corolario del teorema, es decir que, a partir de una interpretación estricta, el teorema contradice la tesis de Church-Turing; **ii)** que, curiosamente, Wold no adoptó el concepto de "*shock errático*" introducido por Frisch, quien consideró a las variables $\varepsilon_t$ como "*desvíos del comportamiento estructural (...) como algo nuevo, espontáneo agregado a la estructura* (autorregresiva) *del proceso*"; **iii)** que no consideró la tesis Fechneriana de la "*indeterminación por novedad*", según la cual cada innovación $\varepsilon_t$ representa nueva información que actualiza el valor del proceso para cada valor de $t$ y que, por lo tanto, replantea para cada $t$ nuevas condiciones iniciales y **iv)** que, más allá que el reconocimiento que la condición de aleatoriedad pura implica suponer que las variables $\varepsilon_t$ se distribuyen de acuerdo con una función de densidad simétrica con valor esperado nulo y que la condición de estacionariedad implica suponer que su varianza es finita, Wold no tuvo en cuenta la aproximación asintótica a la Normalidad que proporciona la versión de Laplace del teorema central del límite, lo que lo condujo a considerar como "*insoportablemente arbitrario*" asignarles una distribución de probabilidades particular.

La quinta y última ("*last but not least*") conclusión a la que se arribó es: **i)** que, diez años después de la aparición de la propuesta de Wold, el progresivo debilitamiento de la hipótesis determinista de la ortodoxia econométrica generó la necesidad de introducir modificaciones "ad hoc" en las especificaciones de los modelos estructurales para representar las eventuales regularidades locales, y de interpretar a los modelos como hipótesis generales sobre las principales relaciones causales sugeridas por la teoría económica, testeables a partir de las observaciones y **ii)**que este cambio en la teoría de modelos hacia un paradigma aleatorista se fundamenta en la obra de H. Theil.

## Referencias

Adrain, R. (1809): “Research concerning the probabilities of the errors wich happen in making observations”. The analyst, vol. 1 (93-109). Reeditado en Stigler, S., 1980.

Alt, F.L. (1942): “Distributed lags”. Econometrica, vol. 10 (113-128).

Arbuthnot, J. (1710-1712): “An argument for Divine Providence, taken from the regularity observed in the birth of both sexes”. Philosophical Transactions of the Royal Society, vol. 27 (186-190).

Archibald, R.C. (1926): “A rare pamphlet of de Moivre and some of his disconveries”. Isis, vol. 8, n° 4 (671-684).

Arnauld, A.; Nicole, P. (1662): “*La logique ou l’art de penser*”. Reeditado por Flammarion, 1970.

Baxter, D. (1870): “National income and taxation of the United Kingdom”. Congrés International de Statistique à la Haye (137-144).

Bentzel, R.; Wold, H. (1946): “On statistical demand analysis from the viewpoint of simultaneous equations”. Skandinavisk Aktuarietidskrift, vol. 29 (95-114).

Bernoulli, D. (1770-1771): “*Mensura sortis ad fortuitam successionem rerum naturaliter congentium applicata*”. Reeditado en “*Die Werke von Daniel Bornoulli*”, Birhäuser (1982).

Bernoulli, D. (1778): “Diiudicatio maxime probabilis plurium observationem discrepatium atque verosimillima induction inde formanda”. Acta Academiæ Scientiarum Imperialis Petropolianæ. Reeditado en “*Die Werke von Daniel Bernoulli*”, Birkhäuser.

Bernoulli, J. (1975): “*Meditationes*” (1684-1690). En “*Die Werke von Jakob Bernoulli*”. Birkhäuser, 1975.

Bernoulli, J. (1713): “*Ars conjectandi*”. Thurnisiorum.

Bernoulli, N. (1709): “*De usu artis conjectandi in jure*”. Basilea. Reeditado en “*Die Werke von Jakob Bernoulli*”. Birkhäuser, 1975.

Bienaymé, I.J. (1855): “Communication sur un príncipe que M. Poisson avait cru decouvrir et qu’il avait appelé loi des grandes nombres”. Séances et Travaux de l’Academie des Science Morales et Politiques, Serie 3, vol. 11 (379-389).

Boscovich, R.J.; Maire, C. (1755): "*De Litteraria Expeditione per Pontificiam ditionem ad dimetiendas duas Meridiani gradus*", Palladis.

Bowley, A.L. (1921): "Elements of statistics". P.S. King.

Boyle, R. (1666): "*Origin of forms and quantities*". En"*The works of the Honorable Robert Boyle*", Londres (1772).

Boyle, R. (1675): "*Some considerations about the reconcileableness of reason and religion*". En"*The works of the Honorable Robert Boyle*", Londres (1772).

Cardano, G. (1539): "Pratica aritmeticæ et mensurandi singularis". En "*Opera omnia*". Traducción al inglés por Stuttgart-Bad Cannstat (1966).

Charleton, W. (1654): "*Physiologia Epicuro-Gassendi- Charletoniana*", Londres. Reeditado en NuevaYork-Londres (1966).

Chebychev, P.L. (1867): "Des valeurs moyennes". Journal de Mathématiques pures et appliqués, vol. 12 (177-184).

Chow, Y.S.; Teicher, H. (2003): "*Probability theory: Independence, interchangeability, martingales*". Springer.

Cochran, D.; Orcutt, G. (1949): "Application of least squares regression to relationships containing autocorrelated error terms". JASA, vol. 44 (32-61).

Côtes, R. (1722): "Opera miscellanea sive æstimatio errorum in mixta mathesi, per variationes partium trianguli plani et sphærici". En Smith, R. (ed.).

Cournot, A.A. (1838): "*Recherches sur les principles mathématiques de la théorie des richesses*". París.

Cournot, A.A. (1843): "*Exposition de la théorie des chances et des probabilités*" Hachette. Reeditadoen "*Cournot, A.A. OEvres completes*". Vrin, 1973-1984.

Craig, J. (1699): "*Theologiæ christianæ ptincipia mathematica*", Londres. Traducción al ingles como "*Craig's rules of historical evidence*". En "*History and theory*", Beiheft 4, The Hage (1964).

Cramér, H. (1961): "On some classes of non-stationary stochastic processes". En Neyman (ed.).

Dale, A.I. (1991): "*A history of inverse probability: From Thomas Bayes to Karl Pearson*". Springer.

Daston, L. (1980): “Probabilistic expectation and rationality in classical probability theory”. Historia Mathematica, vol. 7 (234-260).

Daston, L. (1988): “*Classical probability in the enlightenment*”. Princeton University Press.

de Moivre, S.D. (1718): “*The doctrine of chances: or a method of calculating the probability of events in play*”. W.- Pearson.

de Moivre, S.D. (1730): “*Miscellanea analítica de seriebus et quadraturis*”. Tonson & Watts.

de Moivre, S.D. (1733): “Approximatio ad summam terminorum binomii $(a+b)^n$ in seriem expansi”. Reproducciónen Archibald, R.C. (1926).

de Montmort, P.R. (1713): “*Essai d’analyse sur les jeux de hazard*”. Quillau. Reeditado por Chelsea (1980).

de Morgan, A. (1838): “On a question in the theory of probabilities”. Transactions of the Cambridge Philosophical Society, vol. 6 (423-430).

Derham, W. (1716): “*Physico-theology or ademonstration of the being and attributes of God from his works of creation”*. Londres. Reeditado, Nueva York (1977).

Doob, J.. 1944): “The elementary Gaussian processes”. Annals of Mathematical Statistics, vol. 15 (229- ).

Drobisch, M.W. (1871): “ÜberMittelgrössen und die Anwendberkeitderselben auf die berechnung des Steigens und Sinkens des geldwertes”. JarbücherfürNationalökonomieunsStatistik (1871).

Durbin, J. (1960a): “The fitting of time-series models”. Review of the International Statistical Institute, vol. 28 (233-243).

Durbin, J. (1960b): “Estimation of parameters in time-series regression models”. JRSS, Serie B, vol. 22 (139-153).

Edgeworth, F.Y. (1881): “*Mathematical psychics: An essay on the application of mathematics to the moral sciences*”. C. Keegan. Reeditado por A.M. Kelley, 1967.

Edgeworth, F.Y. (1896): “Eine Erwiderung”.JahrbücherfürNationalökonomie und Statistik, vol. 12 (838-845).

Edgeworth, F.Y. (1910): “On the applicacion of the calculus of probabilities to statiscs”. Bulletin de l’Innstitut Internationale de Statistique, vol. 18 (505-537).

Epstein, R.J. (1987): “*A history of econometrics*”. North-Holland.

Euler, L. (1749): “*Recherches sur la question des inegalités du mouvement de Saturneet et Jupiter, sujet propose por le prix de l’aneé, par l’Academie Royale des Sciences de Paris*”. Turici.

Fechner, G.T. (1866): “*Elemente der Psycophysics*”. Versióneninglés, Holt-Rinehart-Winston (1966).

Fechner, G.T. (1871): “Zúr experimentalen Ästhetik”. Abhandlungen der Könglich säschischen gesellchaft der Wissenschaften, vol. 9.

Fechner, G.T. (1906): “*Send-Avesta, oderüber die Dinge des Himmelsund des Jenseits*”. 3ra. edición, Leopol Voss.

Fellner, F. (1901): “L’evaluation dela richesse nationale”. Bulletin de l’Institute International de Statistique, vol. 13, part 2 (96-136).

Fellner, F. (1905): “Die Scätzung der Volkseinkomments”. Bulletin de l’Institute International de Statistique, vol. 14, part 3 (109-120).

Fine, T.L. (1973): “*Theories of probability: an examination of foundations*”. Academic Press.

Fisher, I. (1892): “Mathematical investigations in the theory of value and prices”. Kelley.

Fisher, I. (1911): “The purchasing power of money”. Kelley.

Fisher, I. (1922): “The making of index numbers: A study of the varieties, tests and reliability”. Kelley.

Fisher, R.A. (1926): “*The mathematical theory of probability and its applications to frequency curves and statistical methods*”. Oliver & Boyd.

Franklin, W.J. (2001): “*The science of conjecture: Evidence and probability before Pascal*”. The John Hopkins University.

Frisch, R. (1933): “Propagation problems and impulse problems in dynamic economics”. En “*Economic essays in honour of Gustav Cassel*”. Allen-Unwin.

Frisch, R. (1934): “*Statistical confluence analysis by means of complete regression systems*”.Universitets Økonomiske Institutt.

Frisch, R. (1936): "Time series and business cycle analysis: Economic macro dynamics". En "Report of the work done under the direction of Professor I. Wedervang, at the U. Institute of Economics, Oslo". Enero 1932-Junio 1936, Rockefeller Archive Centre.

Frisch, R. (1938): "Statistical versus theoretical relations in economic macrodynamics". Memorandum, Oslo. Reproducido en Hendry; Morgan (1995).

Galilei, G. (1632): "*Dialogo sui massimi sistema del mondo*". En "*Opere*", editado por A. Favaro, 1968.

Gassendi, P. (1659): "*Syntagma*". Londres.

Gauss, C.F. (1809): "*Theoria motus corporum coelestiorum*" Partes et Besser. Reimpreso por Dover, 1963.

Gauss, C.F. (1816): "Bestimmung dert Genauigkeit der Beobachtungen". Reeditado en "*Carl Friederich Gauss*". Königlichen Gesellschaft der Wissenschaften, 1880.

Gauss, C.F. (1823): "*Theoria combinationis observationum erroribus minimis obnoxiæ*". Dietrich.

Gerhardt, C.I. (ed.)(1962): "*G. Leibniz Mathematische Schriften*". G. Olms.

Gini, C. (1921): "*Sull' interpolazione di una retta quando i valori della variable indipendente sono affetti da errori accidentali*". Metron, vol. 1 (63-82).

González, M.L.; Landro, A.H. (2013): "Acerca de la interpretación del concepto de perturbación en los procesos discretos de parámetro finito". IV Seminario de Docencia, Investigación y Transferencia en las Cátedras de Matemáticas para Economistas.

Gowing, R. (1983): "*Roger Cotes-Natural philosopher*". Cambridge University Press.

Granger, C.W.J.; Newbold, P. (1977): "*Forecasting economic time series*". Academic Press.

Graunt, J. (1662): "*Natural and political observations mentioned in a following index and made upon the bills of mortality*". Londres.

Gravesande, G.J. (1712): "*Démonstration matématique du soin que dieu prend de diriger ce qui se passé dans ce monde, tirée du nombre des garçons et des* *filles qui naissent journellement*". En "*Oeuvres philosophiques et mathématiques de Mr. G.J. 'sGravesande*". Marc Michel Rey (1774).

Hagstroem, K, -G. (1938): "Pure economics as a stochatical theory". Econometrica, vol. 6 (40-47).

Haavelmo, T. (1938): "The method of supplementary confluent relations. Illustrated by a study of stock prices". Econometrica, vol. 6 (203-218).

Haavelmo, T. (1939): "Statistical testing of dynamic systems if the series observed are shock cumulants". Report of the 5th Annual Reseaerch Conference on Economics and Statistics.

Haavelmo, T. (1940a): "The inadequacy of testing dynamic theory by comparing the theoretical solutions and observed cycles". Econometrica, vol. 8 (312-321).

Haavelmo, T. (1940b): "The problem of testing economic theories by means of passive observations". Report of the 6th Annual Reseaerch Conference on Economics and Statistics.

Haavelmo, T. (1943): "The statistical implications of a system of simultaneous equations". Econometrica, vol. 11 (1-12).

Haavelmo, T. (1944): "*The probability approach in econometrics*". Suplemento a Econometrica, vol. 12.

Hald, A. (1984): "A. De Moivre: '*De mensura sortis*' or '*On the measurement of chance*'". International Statistical Review, vol. 52 (229-236).

Halley, E. (1693): "An estimate of the degrees of the mortality of mankind, drawn from curious tables of the births and funerals at the city of Breslau wit an attempt to escertyain the price of annuities upon lives". Philosophical Trensactions of the Royal Society, vol. 17 (596-610).

Hammond, N. (ed.)(2003): "The Cambridge companion of Pascal". Cambridge University Press.

Hamouda, O.; Rowley, J.C.R. (1988): "*Expectation, equilibrium and dynamics*". Hemel Hempstead.

Hannan, E.J. (1970): "*Multiple time series*". Wiley.

Hartley, D. (1749): "*Observations on man, his frame, his duty and his expectations*". Londres.

Hendry, D.F.; Morgan, M.S. (1989): “A re-analysis of confluence analysis”. Oxford Economic Journal, vol. 41 (35-52).

Hotelling, H. (1927): “Differential equations subjet to error”. JASA, vol. 22 (283-314).

Hume, D. (1718): “*An inquiry concerninghuman understanding*”. Reeditado por Handel (1955).

Jevons, W.S. (1871): “*The theory of practical economy*”. MacMillan.

Kendall, M.G.; Plackett, R.L. (eds.)(1977): “*Studies in the history of statistics and probability*”. Griffin.

Keynes, J.M. (1921): “*A treatise of probability*”. MacMillan, 1963.

Keynes, J.M. (1993): “*The general theory of employment, interest and money*”. En “*Collected writings of John Maynard Keynes*”. MacMillan, Cambridge University Press for the Royal Economic Society.

Kolmogorov, A.N. (1933): “*Grundbegriffe der Wahrcheinlichkeitsrechnung*”. Springer. Traducción al inglés: “*Foundations of the theory of probability*”, Chelsea, 1950.

Kolmogorov, A.N. (1965): “Three approaches to the quantitative definition of information”. Problems of Information Transmission, vol. 1 (1-7).

Kolmogorov, A.N. (1941a): “Stationary sequences in Hilbert space”. Bulletin Moscow State University, Mathematics, vol. 2.

Kolmogorov, A.N. (1941b): “Interpolation und extrapolation von stationaren zufalligen Folgen”. Bolletin de la Academia de Ciencias de la URSS, Serie Matemática, vol. 5 (3-14).

Koopmans, T.C. (1937): “*Linear regression analysis of economic time series*”. Netherlands Economic Institute.

Koopmans, T.C. (1950): “*Statistical inference in dynamic economic models*”. Cowles Commission, Monografía nº 10.

Kops, (1869): “Revenu annuel de la nation”

Lambert, J.H. (1760): “Photometria, sice de mensura et gradibusluminis, colorum et umbræ”. Detleffsen.

Lambert, J.H. (1765-1772): “BeyträgezumGebrauche der Mathematik und derenAnwendung”. Verlage des Buchladens des Realschule. Vols. 1-3.

Landro, A.H. (2010): “*Acerca de la probabilidad*”. Ediciones Cooperativas.

Landro, A.H.; González, M.L. (2013): “Acerca de la interpretación económica de los conceptos de esperanza matemática y esperanza moral y sus contraejemplos. Parte I: Las soluciones clásicas”. XIII Jornadas de Tecnología Aplicada a la Educación Matemática Universitaria.

Landro, A.H.; González, M.L. (2014): “Acerca de la interpretación económica de los conceptos de esperanza matemática y esperanza moral y sus contraejemplos. Parte II: Las soluciones frecuencista, logicista y subjetivista”. IV Seminario de Docencia, Investigación y Transferencia en las Cátedras de Martemática para Economistas.

Laplace, P.S. (1774): “Mémoire sur la probabilité des causes par les événements”. Mémoires de l’Académie Royal des Sciences Presentés. Savants ètrangers., vol. 6 (621-656).

Laplace, P.S. (1781): “Mémoire sur les probabilités”. Mémoires de l’Académie Royal des Sciences de Paris (227-332).Acta academica Scientiarum Imperialis Petropolitanæ (3-23). Traducción al inglés en Kendall (1961)(3-13).

Laplace, P.S. (1786): “Mémoire sur les approximations des formules qui sont fonctions de très grands nombres”. Mémoires de l’Académie Royal des Sciences de Paris (1-88).

Laplace, P.S. (1811): “Mémoire sur les intégrales définies et leur application aux probabilités, et spécialement á la recherche du millieu qu’il faut choisir entre les résultats des observations”. Mémoires de l’Académie Royal des Sciences de Paris, Primera Serie, vol. 11 (279-437).

Laplace, P.S. (1814): “*Essai philosophique sur les probabilités*“. En “*Œuvres completes*“, Académie des Sciences Sciences. Traducción al inglés como “*A philosophical essay on probabilities*“, Nueva York (1951).

Laspeyres, A.Th. (1864a): “Hamburger Warenpreise und die californisch-australischen Goldentdeckungenseit 1848”. Jahrbücherfür Nationalökonomie und Statistik (81-220).

Leamer, E.E. (1978): “*Specification searches: Ad hoc inference with non-experimental data*”. Wiley.

Legendre, A.M. (1805): “Nouvelles methods pour la determination des orbites des comètes”. Firmin Didot.

Leibniz, G.W. (1678): “*De incertiæstimatiome*”. Berlin.

Lexis, W. (1876) : “Das Gesachlechtsverhältnis der Geborenen und die Wahrscheinlichkeitsrechnung”. Jarbücher für Ntionalökonomie und Statistik, vol. 27 (209-245).

Magnus, J.R-; Morgan, M.S. (1987): “The ET interview: Professor J.Tinbergen”. Econometric Theory, vol. 4 (187-209).

Mann, H.B.; Wald, A. (1943): “On the statistical treatment of linear stochastic difference equartions”. Econometrica, vol. 11 (173-220).

Markov, A.A. (1924): “*The calculus of probabilities*”. Moscú.

Marschak, J. (1937): “Utility and probability in human choice”. En Report of 3rd. Annual Research Conference on Economics and Statistics. Cowles Commission.

Marschak, J. (1953): “Economic measurements for policy and prediction”. En Hood; Koopmans (eds.).

Mayer, T. (1750): “Abhandlungüber die Umwalzung des Monas um seine Axe und die ScheinbareBewegung der Mondsflecten”

Moore, H.L. (1914): “*Economic cycles: Their law and cause*”. Nueva York.

Moore, H.L. (1923): “*Generating economic cycles*”. Nueva York.

Morgan, M.S. (1987): “Statistics without probability and Haavelmo´s revolution in econometrics”. EnKrüger; Gingerezer; Morgan (eds.).

Morgan, M.S. (1989): “*The history of econometric ideas*”. Cambridge University Press.

Nieuwentyt, B. (1715): “*Het Regt gebruik der wereldeschouwingen*”. Traducido al ingles como “*The religious philosopher: or the right use of contemplating the works of the Creator*”. Londres.

Paasche, H. (1871): “Über die Preisentwicklung derletzen Jahre nach den Hamburger Börsenentwicklungen”. JahrbücherfürNationalökonomie und Statistik.

Pearson, K. (1893a): “Asymetrical frequency curves”. Nature, vol. 48 (20-35).

Pearson, K. (1893b): “Contributions to the mathematical theory of evolution, I”. Proceedings of the Royal Society, vol. 54 (211-242).

Pearson, K. (1895): “Contributions to the mathematical theory of evolution, II”. Philosophical Transactions of the Royal Society, Sereie A, vol. 186 (343-414).

Pearson, K. (1896): “Contributions to the mathematical theory of evolution, III”. Philosophical Transactions of the Royal Society, Sereie A, vol. 187 (174-227).

Pearson, K. (1896): “*The history of statistics in the 17th and 18th centuries, against the changing background of intellectual scientific and religious thought. Lectures given at the University College of London, during the academis sessions 1921-1923*”. Griffin.

Pearson, K. (1924): “Historical note on the origin of the normal curve of errors”. Biometrika, vol. 16 (402-404).

Pearson, K. (1925): “James Bernoulli theorem”. Biometrika, vol. 17 (201-210).

Persons, W.M. (1916): “Construction of a business barometer”. American Economic Review, vol. 6 (739-769).

Persons, W.M. (1922-23): “Correlation of time series”. JASA, vol. 18 (713-726).

Poisson, S.D. (1836): “Note sur le calcul des probabilités“. Comptes Rendus Hebdomadaires des Séances de l’Académie des Sciences, vol. 2 (395-400).

Poisson, S.D. (1837): “Recherches sur la probabilité des jugements en matière criminelle et en matière civile, précédés des règles générales du calcul des probabilités. Bachelier.

Pólya, G. (1920): “Über den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung und das Momentproblem”. Mathematische Zeitschrift, vol. 8 (171-181).

Prigogine, I.; Nicolis, G. (1977): “*Self-organizatgion in non-equilibrium system, from dissipative structures to order to fluctuztions*”. Wiley.

Qin, D. (1993): “*The formation of econometrics. A historical perspective*”. Clarendon.

Qin, D.; Gilbert, C.L. (2001): “The error term in the history of time series econometric”. Econometric Theory, vol. 17 (424-450).

Quetelet, A. (1848): “Sur la statistique morale et les príncipes qui donent en former la base”. Nouveaux Mémoires de l’Académie Royal des Sciences et Belles Lettres de Belgique, vol. 21 (67-114).

Rowley, J.C.R.; Hamouda, O. (1987): “Troublesome probability and economics”. Journal of Post Keynesian Economics, vol. 10 (44-64).

Sargan, J.D. (1959): “The estimation of relationships with autocorrelated residuals by the use of instrumental variables”. JRSS, Serie B, vol. 21 (91-105).

Sargan, J.D. (1961): “The maximum likelihood estimation of economic relationships with autorregressive residuals”. Econometrica, vol. 29 (414-426).

Sargent, T.J. (1976): “A classicalmacroeconomic model for the United States”. Journal of Political Economy, vol. 84 (207-237).

Sargent, T.J. (1977): “The demand for money during hyperinflations under rational expectations”. International Economic Review, vol. 18 (59-82).

Sargent, T.J. ; Sims, C.A. (1977): “Business cycle modelling without pretending to have too much ‘a priori’ economic theory”. En Sims, C.A. (ed.).

Schultz, H. (1925): “The statistical law of demand”. Journal of Political Economy, v ol. 33 (481-504 y 577-637).

Schultz, H. (1930): “The meaning of statistical demand curves”. Veröffentlichungen der Frankfurter GessellschaftfürKonjunkturforaschung, Leipzig.

Schultz, H. (1939): “*The theory and measurement of demand*”. Chicago.

Schumpeter, J. (1939): “*Business cycles*”. Nueva York.

Sent, E,-M. (1998): “*The evolving rationality of rational expectations: An assessment of Thomas Sargent’s achievements*”. Cambridge University Press.

Shea, W. (2003): “*Designing experiments and games of chance*”. Reidel.

Sheynin O.B. : “On the early history of the law of large numbers”. Biometrika, vol. 55, n°3 (459-467).

Simpson, T. (1755): “On the advantage of taking the mean of a number of observations in practical astronomy. A letter to the Right Honorable George Earl of Macclesfield. President of the Royal Society”. Philosophical Transactions of the Royal Society, vol. 49 (82-93).

Simpson, T. (1757a): “*Miscellaneous tracts on some curious, and very interesting subjects in maechanics, physical-astronomy and speculative mathematics*”. J. Nourse.

Simpson, T. (1757b): “An attempt to show the advantage arising by taking ther mean of number of observations in practical astronomy”. En Simpson, T. (1757).

Sims, C.A. (ed.)(1977): "*New methods in business cycle research*". Federal Reseerve Bank of Minneapolis.

Sims, C.A. (1980): "Macroeconomic and reality". Econometrica, vol. 48 (1-48).

Slutsky, E. (1937): "The summation of random causes as the source of cyclic processes". Econometrica, vol. 5 (105-146). Original en ruso, 1927.

Smith, R. (ed.) : "*Opera miscellanea*". Cambridge, 1722.

Staehle, H. (1933): "Henry L. Moore and statistical economics". Econometrica, vol. 1 (73-86).

Stigler, S.M. (1975): "Napoleonic statistics: The work of Laplace". Biometrika, vol. 62

Stigler, S.M. (1986a): "Laplace 1774 memoir on inverse probability". Statistical Science, vo. 1 (359-378).

Stigler, S.M. (1986b): "*The history of statistics: The measurement of uncertainty before 1900*". Harvard University Press.

Stirling, J. (1730): "*Methodus differentialis: sive tractatus de summation et interpolation serieruim infinitarium*". Bouyer.

Süsmilch, J.P.: "*Die göttliche Ordnung in der Veränderugen des menslichen Gesechlechs*". Reeditado como "*L'ordre divin aux origins de la démographie*". Institut National d'Études Démographiques (1979).

Theil, H. (1957): "Specification errors and the estimation of economic relationships". Review of International Statistical Institute, vol. 25 (41-51).

Theil, H. (1958): "*Economic forecast and policy*". North-Holland.

Tinbergen, J. (1935): "Annual survey: Suggestion on quantitatyive business cycle theory". Econometrica, vol. 3 (241-308).

Tinbergen, J. (1936): "*Grondproblem der TheoretischeStatistiek*". Bohn.

Tinbergen, J. (1937): "*Econometric approach to business cycles*". París.

Tinbergen, J. (1938): "On the theory of business-cycle control". Econometrica, vol. 6 (22-29).

Tinbergen, J. (1939): "*Statistical testing of business-cycle theories*". Ginebra.

Tintner, G. (1938a): "A note on economic aspects to the theory of errors in time series". Quarterly Journal of Economics, vol. 53 (141-149).

Tintner, G. (1938b): "The maximization of utility over time". Econometrica, vol. 6 (154-158).

Uspensky, J.V. (1937): "*Introduction to mathematical probability*". McGraw-Hill. Traducido como "*Matemática de las probabilidades*", Nigar (1954).

Von Mises, R. (1920): "*Wahrscheinlichkeit, Statistik und Wahrheif*". Springer. Traducción al ingles como "*Probability, statistics and truth*", MacMillan (1928).

Walker, G. (1931): "On periodicity in series related terms". Philosophiocal Transactions of the Royal Society, Serie A, vol. 131 (519-532).

Walras, L. (1874): "Economique et mécanique"

Westergaard, .L. (1890): "Grunzüge der Theorie des Statistik". En Fisher, G. (ed.).

Wicksell, K. (1907): "KrisernasGåta".StatøkonomistTedsskrift (255-286).

Wiener, N. (1942): "The extrapolation, interpolation and smoothing of stationary time series with engineering applications". MIT DIC Contract 6037, National Defence Research Copuncil, Section D2. Reeditado en Wiley, 1949.

Wold, H. (1938): "*A study in the analysis of stationary time series*". Almqvist-Wiksells.

Wold, H. (1948): "One prediction in stationary time.series". Annals of Mathematical Statistics, vol. 19 (558-567).

Wold, H. (1949): "Statistical estimation of economic relationships". Econometrica, vol. 17 (1-22).

Wold, H. (1951): "Dynamic systems of the recursive type: Economic and statistical aspects". Sankhyá, vol. 11 (205-216).

Wold, H. (1954): "Causality and econometrics". Econometrica, vol. 22 (162-177).

Yule, G.U. (1921): "On the time-correlation problems, with special reference to the variate-difference correlation method". JRSS, vol. 84 (495-526).

Yule, G. U. (1926): "Why do we sometimes get nonsense correlations between time-series? A study in sampling and the nature of time-series". JRSS, vol. 89 (1-64). Reeditado en Stuart; Kendall (1971).

Yule, G. U. (1927): “On a method investigating periodicities in disturbed series, with special application to Wolfert’s sun spot numbers”. Philosophical Transactions of the Royal Society, Serie A, vol. 226 (267-298).