PROBABILIDAD E INCERTIDUMBRE EN KEYNES. UNA REVISIÓN DE TIPO BAYESIANO*

Juan Esteban Jacobo[a]

* DOI: https://doi.org/10.18601/01245996.v23n45.07. Recepción: 03-03-2021, modificación final: 15-04-2021, aceptación: 02-06-2021. Sugerencia de citación: Jacobo, J. E. (2021). Probabilidad e incertidumbre en Keynes. Una revisión de tipo bayesiano. *Revista de Economía Institucional*, *23*(45), 137-162.

[a] Doctor en Economía, profesor de la Universidad Externado de Colombia, Bogotá, [juan.jacobo@uexternado.edu.co], [https://orcid.org/0000-0001-7078-0266].

Probabilidad e incertidumbre en Keynes. Una revisión de tipo bayesiano

Resumen Este articulo analiza el concepto de probabilidad en la obra de Keynes y propone un metodo para formalizar la nocion de la incertidumbre en la *Teoría general* utilizando el teorema de Bayes y el principio de maxima entropia. Una de las principales conclusiones es que, a pesar de compartir su rechazo del enfoque frecuentista de la estadistica, es poco razonable pensar que no se pueden determinar probabilidades numericas, aunque se cumpla algun criterio de objetividad.

Palabras clave: entropia, teorema de Bayes, incertidumbre, Keynes; JEL: B16, B31, C11

Uncertainty and probability in Keynes. A Bayesian-type review

Abstrac This paper analyzes the concept of probability in Keynes' work and proposes a method for formalizing the notion of uncertainty in the *General Theory* using Bayes' theorem and the principle of maximum entropy. One of the main conclusions is that, despite sharing his rejection of the frequentist approach to statistics, it is unreasonable to think that numerical probabilities cannot be determined, even if some criterion of objectivity is met.

Keywords: entropy, Bayes' theorem, uncertainty, Keynes; JEL: B16, B31, C11

Probabilidade e incerteza em Keynes. Uma revisão do tipo bayesiano

Resumo Este artigo analisa o conceito de probabilidade na obra de Keynes e propõe um método para formalizar a noção de incerteza na Teoria Geral usando o teorema de Bayes e o princípio da entropia máxima. Uma das principais conclusões é que, apesar de compartilhar sua rejeição à abordagem frequentista da estatística, não é razoável pensar que as probabilidades numéricas não podem ser determinadas, mesmo que algum critério objetivo seja atendido.

Palavras-chave: entropia, teorema de Bayes, incerteza, Keynes; JEL: B16, B31, C11

Antes de ser economista, John M. Keynes fue matemático. Su tesis doctoral, *Un tratado sobre la probabilidad,* fue su primera gran obra; presentada en 1907, revisada en 1908 y publicada en 1921, cuando ya era conocido por *Las consecuencias económicas de la paz.*

El *Tratado sobre la probabilidad* fue un gran esfuerzo intelectual que hizo algunos aportes originales y contribuciones relevantes para la interpretación filosófica de la probabilidad. De manera similar a Bernoulli, Laplace y Poincaré, Keynes interpretó la probabilidad como una representación de grados de creencias y, así, como una extensión de la lógica deductiva. Sin embargo, a diferencia de esos autores, Keynes introdujo una cláusula de objetividad, según la cual debe ser cierto que personas igualmente informadas deben llegar a los mismos grados de creencia, independientemente de sus opiniones.

Como veremos, esta cláusula de objetividad es la principal debilidad en el análisis de Keynes y, quizá, fue la razón para que la teoría matemática de la probabilidad se desarrollara con independencia de los argumentos del *Tratado.* Conforme a dicha restricción, Keynes niega que las probabilidades sean siempre comparables y que sea posible asignar valores numéricos a los grados de creencia. Este argumento es de gran importancia por la relación con sus obras económicas, en especial con la *Teoría General*, en la que el concepto de incertidumbre es un eje central (Keynes, 1973, p. 113).

Este artículo, que conmemora los cien años de la publicación del *Tratado*, analiza el concepto de probabilidad en las obras de Keynes y propone un método para formalizar la noción de incertidumbre en la *Teoría General* utilizando el teorema de Bayes y el principio de máxima entropía. Una de sus principales conclusiones es que, a pesar de compartir el rechazo de Keynes al enfoque frecuentista de la estadística, es poco razonable pensar que no es posible asignar probabilidades numéricas, aunque sean determinadas de una manera objetiva.

Una abundante literatura aborda diversos aspectos del *Tratado* y su influencia en las ideas económicas de Keynes, pero no parece existir una revisión detallada de los fundamentos que separan el enfoque bayesiano de la probabilidad y el de Keynes, ni propuestas para interpretar sus objeciones a la medición numérica de la probabilidad. Algunos trabajos –como los de Carabelli (1988) y O'Donnell (1989)– son muy informativos acerca del método de Keynes y su desarrollo filosófico desde el *Tratado* hasta la *Teoría General*. La relación entre el enfoque bayesiano y el de Keynes se discute parcialmente en Bateman (1987), Roncaglia (2009) y Feduzi, Runde y Zappia (2013). Este artículo complementa esa literatura en cuanto formaliza la coherencia lógica del enfoque bayesiano y propone un método para revisar el concepto de incertidumbre en la obra económica de Keynes.

LA VISIÓN KEYNESIANA DE LA PROBABILIDAD

Keynes concibió la teoría de la probabilidad como una rama de la lógica e independiente del concepto de frecuencia. Desde esa perspectiva, es una extensión de la lógica deductiva, una lógica de la inferencia probable. En el desarrollo de esta idea, Keynes describió la probabilidad como una relación entre premisas y conclusiones, que se establece *racionalmente* y da lugar a grados de creencia limitados por los extremos de imposibilidad y certeza (Keynes, 1921, p. 3).

A diferencia del enfoque matemático de la probabilidad, Keynes definió los grados de creencia con respecto a proposiciones y no a eventos. Esta distinción, en apariencia sutil, no solo permite representar eventos –algunos de los cuales pueden verse como realizaciones en el tiempo– sino también afirmaciones que pueden ser verdaderas o falsas, pero que por falta de información se consideran inciertas[1]. Así, la teoría de la probabilidad es de carácter subjetivo, porque los grados de creencia se basan en el conocimiento individual. A este aspecto subjetivo, añadió el concepto de creencia racional, el vínculo que convierte la probabilidad en un asunto objetivo.

> Una proposición no es probable porque pensemos que lo es. Una vez están dados los hechos que determinan nuestro conocimiento, lo que es probable o improbable en estas circunstancias ha sido fijado objetivamente, y es independiente de nuestra opinión. La teoría de la probabilidad es entonces lógica, porque se refiere al grado de creencia que es *racional* mantener en unas condiciones dadas, y no meramente a las creencias actuales de individuos particulares, que pueden ser o no racionales (ibíd., p. 4).

Este pasaje es muy citado, pero rara vez se interpreta en detalle para entender por qué la probabilidad puede ser objetiva. La clave es la definición de creencia racional y conocimiento. Cuando una persona cree algo por una razón absurda o sin ningún motivo, esa creencia –aun si llega a ser verdad– no se puede considerar racional. En cambio, esa creencia es racional si existen razones que la soportan –aun si llega a ser falsa–. El caso límite del grado de creencia racional, al que Keynes llamó creencia racional cierta, es conocimiento (ibíd., p. 21).

[1] "Todas las proposiciones son verdaderas o falsas, pero el conocimiento que tenemos de ellas depende de nuestras circunstancias; y aunque a veces es conveniente hablar de proposiciones ciertas o probables, esto expresa una relación estricta en la que se basa nuestro *corpus* de conocimiento, real o hipotético, y no una característica de las proposiciones en sí mismas" (ibíd., p. 4).

Teniendo en mente las definiciones anteriores, supongamos que p es una proposición acerca de la cual existe un grado de creencia dado el conocimiento de una hipótesis o premisa h. Y, además, que existe una proposición secundaria q que establece la relación de probabilidad entre p y h. Los grados de creencia de p se consideran racionales solo si conocemos la proposición q, es decir, es necesario tener certeza sobre la relación de probabilidad entre la premisa (h) y la conclusión (p) para que la creencia sea racional. En cambio, si no se conoce con certeza la proposición secundaria q, no puede existir un grado de creencia racional acerca de p dado el conocimiento de h[2]. Este principio implica que dos individuos que tienen la misma información no pueden llegar a creencias racionales diferentes.

Así, es evidente que no siempre será posible determinar probabilidades y menos aún asignarles un valor numérico[3]. Uno de los ejemplos que utiliza Keynes es el de los seguros, con el cual ilustra que no se puede asignar un valor numérico a los grados de creencia porque no se conoce con certeza el proceso de fijación de la prima de riesgo. Y con el cual descarta la existencia de un valor numérico de la probabilidad porque dos corredores de seguros inteligentes con igual información pueden no llegar al mismo resultado (ibíd., p. 23).

Ramsey (1960) criticó esa visión de la probabilidad porque le parecía absurdo pensar que solo existe una proposición secundaria cierta y que debe ser conocida por la mente humana. Sostuvo que la proposición secundaria cambia según las creencias y experiencias de cada individuo, y concluyó que –si se aceptan algunos principios de consistencia– la incertidumbre siempre puede ser medida como

[2] "Para que podamos tener una creencia racional en una proposición p [...] es necesario que se cumpla una de las dos condiciones siguientes: (i) que conozcamos p directamente; o (ii) que conozcamos un conjunto de proposiciones h, y que también conozcamos una proposición secundaria q que afirme una relación cierta entre p y h" (ibíd., p. 17).

[3] Además de conocer la relación entre premisa y conclusión, una probabilidad numérica existe según Keynes cuando todas las alternativas sean exhaustivas y mutuamente excluyentes. Bajo este principio, si no existe información que implique lo contrario, la misma probabilidad se debe asignar a todas las alternativas posibles. En el Capítulo 15 del *Tratado,* Keynes utiliza este principio para argumentar que éste constituye uno de los pocos casos sobre los cuales una probabilidad numérica se puede definir. Para definir la probabilidad numérica sobre, por ejemplo, $a|b$, es necesario que existan otras probabilidades $a_1|b_1$, $a_2|b_2$ ..., $a_q|b_q$, ..., $a_r|b_r$, tal que para cualquier par i,j entre 1 y r, $a_i|b_i = a_j|b_j$. En este sentido, podemos definir $a|b = \Sigma^{q}_{(h=1)}\, a_h|b_h$ y $1 = \Sigma^{r}_{(h=1)}\, a_h|b_h$, tal que $a|b = \frac{q}{r}$. Es decir, la existencia de un grado de creencia racional es necesaria pero no suficiente para determinar una probabilidad numérica.

issn 0124-5996/e-issn 2346-2450

probabilidad. En particular, señaló que "una revisión detallada de la naturaleza de las creencias parciales revela que las leyes de la probabilidad son las leyes de la consistencia" (1960, p. 182). Esta crítica es de gran importancia porque parece cambiar la visión de Keynes sobre la probabilidad.

> En contra de la visión que propuse, Ramsey argumenta que la probabilidad no se ocupa de relaciones objetivas entre proposiciones sino (en cierto sentido) de grados de creencia, y logra demostrar que el cálculo de probabilidades consiste simplemente en un conjunto de reglas que aseguran que el sistema de grados de creencia que mantenemos es un sistema *consistente*. Por tanto, el cálculo de probabilidades pertenece a la lógica formal. Pero la base de nuestros grados de creencia –o las probabilidades *a priori,* como se suelen llamar– es parte de nuestro equipamiento humano, quizá dado por simple selección natural, análoga a nuestras percepciones y recuerdos más que a la lógica formal. Hasta aquí, cedo ante Ramsey pues creo que tiene razón. Pero en el intento de distinguir entre grados de creencias "racionales" y grados de creencias en general creo que aún no ha tenido mucho éxito. No se llega al fondo del principio de inducción diciendo simplemente que es un hábito mental útil (Keynes, 1933, pp. 338-339).

Aunque una parte del pasaje parece indicar que Keynes modificó su visión de la probabilidad, el estudio de sus obras económicas confirma que no fue así. Su apego a la noción de creencias racionales y su búsqueda de una justificación lógica del principio de inducción confirma que siguió pensando que las probabilidades son no numéricas (Roncaglia, 2009).

El principio de inducción es otro de los pilares de su teoría de la probabilidad, así como los grados de creencia racionales, que existen con independencia del concepto de frecuencia. Keynes (1921, p. 243) rechazó la visión "pura" de la inducción basada en el número de repeticiones de un experimento. La concibió, en cambio, como un *proceso* cognitivo, cuya validez no depende de su confirmación empírica sino de que exista una relación de probabilidad (Carabelli, 1988, p. 67). Es decir, la inducción no es válida por el éxito de un experimento sino porque su lógica es razonable.

Aunque esta visión da luz sobre su posterior debate con Tinbergen, en el que Keynes resaltó algunas dificultades para aplicar el método frecuentista a la inferencia estadística[4], tampoco parece llegar al fon-

[4] Las críticas de Samuelson (1946) y Patinkin (1976), quienes dijeron que Keynes no entendió el método de Tinbergen, son inapropiadas. La discrepancia de Keynes con Tinbergen no era instrumental sino de *concepción*. Los lectores interesados en profundizar este tema pueden consultar Carabelli (1988, cap. 9) y las referencias correspondientes.

do del asunto: que la inducción –a diferencia de la deducción– no se puede justificar por relaciones lógicas. Aunque, como señala Keynes, no es posible establecer argumentos inductivos sin hacer referencia a un conocimiento o a una probabilidad *a priori*, tiene poco sentido pensar que la conclusión de un razonamiento inductivo es una consecuencia lógica necesaria de sus premisas.

PROBABILIDAD E INCERTIDUMBRE. VISIÓN BAYESIANA

La visión de la probabilidad como extensión de la lógica y no como un concepto que depende de frecuencias fue un gran aporte de Keynes: representó la probabilidad como grados de creencia en proposiciones condicionadas al conocimiento existente. Pero no presentó una lógica formal aceptable ni dio campo para una teoría matemática de la probabilidad. Esto no se deriva de los axiomas que utilizó en el *Tratado,* sino de su negación de la existencia de una expresión numérica de la probabilidad. En esta sección veremos, entre otras cosas, que la noción de creencias racionales es innecesaria y que es más razonable –y quizá inevitable– utilizar los criterios de consistencia y de sentido común para determinar la probabilidad.

Este argumento se presenta formalmente para hacer énfasis en los principios y axiomas básicos de la inferencia bayesiana, y así centrar la discusión en sus fundamentos. Aunque esta tarea impone un gran peso a los lectores, es necesaria porque en cierto modo obliga a los críticos de la visión bayesiana a elegir cuál de sus axiomas no tiene sentido. Si no les es posible, no les queda más alternativa que aceptar sus conclusiones, incluida la de que la probabilidad es en general subjetiva y que se expresa numéricamente. Los lectores familiarizados con el enfoque bayesiano pueden omitir esta sección.

Fundamentos lógicos de la inferencia bayesiana

Para sentar bases de la teoría de la probabilidad desde una perspectiva bayesiana podemos empezar, igual que Keynes, hablando de proposiciones. Una proposición es cualquier enunciado del que existe un conocimiento limitado y –en caso de saberlo con certeza– clasificarlo como cierto o falso. Por ejemplo, es incierto que dos huevos de una cubeta están rotos, que la temperatura en Bogotá mañana a las 6 am será de 5º C o que una moneda después de 20 lanzamientos caerá 6 veces cara.

Al conjunto de elementos sobre los cuales definimos las probabilidades lo denotamos $A = \{a, b, c, \ldots\}$. Igual que en la lógica deductiva, en el cálculo de la inferencia probable utilizamos las operaciones básicas del álgebra booleana: la negación (~), la suma lógica (+) y el producto lógico (·). Así, para toda tripleta de proposiciones a, b, c en *A*, se cumplen los resultados básicos:

$$\sim\sim a = a, a \cdot a = a, a + a = a, \tag{1.1}$$

$$a \cdot b = b \cdot a, a + b = b + a \tag{1.2}$$

$$\sim(a \cdot b) = \sim a + \sim b, \sim(a + b) = \sim a \cdot \sim b, \tag{1.3}$$

$$(a \cdot b) \cdot c = a \cdot (b \cdot c) = a \cdot b \cdot c, (a + b) + c = a + (b + c), \tag{1.4}$$

$$(a + b) \cdot c = (a \cdot c) + (b \cdot c), (a \cdot b) + c = (a + c) \cdot (b + c) \tag{1.5}$$

Igual que a Keynes, aquí nos interesa desarrollar un álgebra para establecer una relación entre premisas y conclusiones. Pero en vez de mantener la premisa de creencias racionales, buscamos que las creencias sean razonables, en el sentido de que la relación sea aceptada por cualquier persona con sentido común y consistencia[5].

Para definir el principio de sentido común llamamos a|b a la probabilidad condicional de a dada la premisa b. Para expresar que creemos más en a que en c dada b, decimos que

$$a|b > c|b. \tag{2}$$

Si tenemos información adicional d que nos hace creer más en , tal que

$$a|d > a|c, \tag{3}$$

y la creencia en b dada a se mantiene igual:

$$b|a \cdot d = b|a \cdot c, \tag{4}$$

solo puede ser cierto que

$$a \cdot b|d \geq a \cdot b. \tag{5}$$

Por ello, si la proposición a es más plausible dada d, también debe ser cierto que la negación de a es menos plausible, es decir:

$$a|d \leq a|c \tag{6}$$

Si una persona obedece las reglas (2)-(6), decimos que cumple el principio de sentido común. El principio de consistencia, por su parte, implica que si una conclusión se puede obtener de más de una manera, todas las soluciones deben dar el mismo resultado. Dados los

[5] Puesto que la lógica estudia relaciones normativas, quizá sea mejor referirnos a seres ideales cuando hablemos de sentido común y consistencia. Esta distinción deja claro que cosas como la *paradoja de Ellsberg*, que se suele utilizar como argumento contra la posibilidad de asignar valores numéricos a las probabilidades (ver, p. ej., Feduzi, 2007), confunden el enfoque normativo con el descriptivo.

principios de sentido común y consistencia, presentamos ahora los axiomas para derivar el álgebra de la inferencia probable.

(a) *La probabilidad de que dos proposiciones sean ciertas dada la evidencia disponible se determina por sus probabilidades de manera separada, una con base en la evidencia disponible, y otra con la evidencia disponible y el supuesto adicional de que la primera inferencia es verdadera* (Cox, 1961, p. 4).

El axioma (a) busca una regla consistente para estimar la probabilidad de $a \cdot b$ con información separada de a y b, considerando cierta una proposición c. Es decir, existe una función tal que

$$a|b \cdot c = F[(a|b \cdot c), (b|c)] \tag{7}$$

El principio de consistencia dice que la ecuación (7) se puede extender a cualquier número de proposiciones en A. Por ejemplo, si queremos saber cuán plausible es que las proposiciones a, b y c sean simultáneamente ciertas dada d, y consideramos $b \cdot c$ como una sola proposición, se obtiene:

$$a \cdot b \cdot c|d = F[(b \cdot c|d), (a|b \cdot c \,.\, d)] = F[F[(c|d), (b|c \cdot d)],(a|b \cdot c \cdot d)]. \tag{8}$$

Así mismo, si consideramos que $a \cdot b$ es una sola proposición:

$$a \cdot b \cdot c|d = F[(c|d), (a \cdot b|c \cdot d)] = F[(c|d), F[(b|c \cdot d), (a|b \cdot c \cdot d)]]. \tag{9}$$

Por simplicidad de notación, hacemos $x \equiv c|d$, $y \equiv b|c \cdot d$, $z \equiv a|b \cdot c \cdot d$, $u \equiv F(x, y)$ y $v \equiv F(y, z)$. Del principio de consistencia se sigue que (8) y (9) deben ser iguales, es decir, el problema es encontrar una función $F \neq$ constante, tal que

$$F(u, z) = F(x, v). \tag{10}$$

Teorema 1: $F(x, y) = w^{-1}[w(x)w(y)]$ *es una solución de (10), en la que* $w(x) = exp\,\{\int \frac{dx}{H(x)}\}$ *y* $H(x)$ *es una función arbitraria.*

Prueba: ver Jaynes (2003, p. 27).

Del teorema 1 se sigue a su vez que $w(F(x, y)) = w(x)w(y)$. Así, puesto que $x \equiv c|d$, y $y \equiv b|c \cdot d$, necesariamente se cumple que

$$w(a \cdot b|c) = w(b|c)w(a|b \cdot c = w(b|a \cdot c)w(a|c). \tag{11}$$

La ecuación (11) no es más que la *regla del producto*, y es resultado de los principios de consistencia y de sentido común. Un corolario del teorema 1 es que a es cierta dada c si $w(a|c) = 1$, y falsa si $w(a|c) = 0$. Es decir, las afirmaciones falsas o verdaderas son casos extremos de la regla del producto y cualquier caso entre estos dos límites tendrá ser un valor entre 0 y 1[6]. Pasemos al segundo axioma.

(b) *La probabilidad de una proposición dada la evidencia disponible determina la probabilidad de su negación dada la misma evidencia* (Cox, 1961, p. 3).

Formalmente, este axioma dice que existe una función $S(\cdot)$ tal que

$$\sim a = S(a). \tag{12}$$

Si $x \equiv w(a|b)$, entonces $\sim x \equiv w(\sim a|b)$. La función $S(\cdot)$ también debe cumplir la regla del producto:

$$w(a \cdot b|c) = w(a|b \cdot c)\ w(b|c) = S[w(\sim a|b \cdot c)]\ w(b|c) = w(b|c)\ S\left[\frac{w(\sim a \cdot b|c)}{w(b|c)}\right]. \tag{13}$$

Una relación similar se mantiene si expresamos $a \cdot b|c$ así:

$$w(a \cdot b|c) = w(b|a \cdot c)\ w(a|c) = w(a|c) S\left[\frac{w(\sim b \cdot a|c)}{w(a|c)}\right]. \tag{14}$$

Dado el principio de consistencia, el lado derecho de las ecuaciones (13) y (14) debe ser igual. Es decir, $x \equiv w(a|c)$ y $y \equiv w(b|c)$

$$x\ S\left[\frac{S(y)}{x}\right] y = S\left[\frac{S(x)}{y}\right] \tag{15}$$

Teorema 2: La función $S(z) = (1 - z^r)^{1/r}$ *es una solución a la ecuación (15) y a* $S[S(z)] = z$.

Prueba: ver Cox (1961, p. 21).

El teorema 2 da lugar a la *regla de la suma*: $w^r(a|b) + w^r(\sim a|b) = 1$. Dada la forma de la función $S(z)$, las probabilidades serán siempre continuas en el intervalo 0 y 1.

Sin pérdida de generalidad, podemos suponer que $r = 1$, tal que las probabilidades se definen como $P(\cdot) = w(\cdot)$. Conforme a las reglas del producto y de la suma, las distribuciones de probabilidad deben cumplir las siguientes propiedades:

$$P(a \cdot b|c) = P(a|c)P(b|a \cdot c) = P(b|c)P(a|b \cdot c), \tag{16.1}$$

$$1 = P(a|c) + P(\sim a|c), \tag{16.2}$$

Reorganizando (16.1) y utilizando la regla de la suma:

$$P(a|b \cdot c) = \frac{P(a|c)P(b|a \cdot c)}{P(b|c)}, \tag{16.3}$$

[6] En sus definiciones de probabilidad, Keynes (1921, p. 145) supone estos mismos valores para la certeza y la imposibilidad.

$$P(b|c)=\sum_i P(a_i|c)P(b|a_i \cdot c), \tag{16.4}$$

donde la sumatoria de (16.4) involucra todos los valores posibles de a_i. Las ecuaciones (16.3) y (16.4) dan lugar al *teorema de Bayes*, que igual que las reglas del producto y de la suma, es resultado directo de los principios de sentido común y de consistencia. Si las proposiciones b y c son premisas y a la conclusión, el teorema nos da una manera consistente de medir la incertidumbre de a dadas b y c. Su medición sigue siendo subjetiva, en el sentido de que es relativa a la información del individuo. No obstante, es una relación objetiva, en el sentido de que todo par de personas que cumplan los principios de sentido común y de consistencia deben hacer la misma estimación[7].

En síntesis, en esta sección se *prueba* pero no se *supone* que la incertidumbre se puede medir como probabilidad, y que esta última siempre puede tener un valor numérico entre 0 y 1, como indica Lindley (2000, p. 296). Además, un resultado de los principios de sentido común y de consistencia es que la probabilidad no existe como propiedad natural del mundo, sino como representación del estado de conocimiento del observador.

LA INCERTIDUMBRE EN LA *TEORÍA GENERAL*

Las dos secciones anteriores dan las bases para interpretar la visión de la incertidumbre en el contexto de la *Teoría General* de Keynes donde, en contra de lo que se podría pensar por sus palabras en homenaje a Ramsey, mantuvo la idea de que no es posible definir probabilidades. Esto es muy claro en su análisis de las expectativas de largo plazo, donde afirmó que las "bases del conocimiento" a partir de las que estimamos los rendimientos futuros son muy precarias (1936, p. 149). Puesto que la definición de las probabilidades requiere tener certeza en la proposición secundaria, la precariedad del conocimiento hace que no sea posible definirlas. Por ello, también rechazó la posibilidad de asignar valores numéricos a la probabilidad. Por ejemplo, en su examen de las convenciones a partir de las cuales se forman las expectativas sobre los rendimientos futuros señaló que era difícil que todas las alternativas de una proposición fueran exhaustivas y mutuamente excluyentes y que "tampoco podemos racionalizar nuestro

[7] Aunque dos individuos que respeten los principios de sentido común y de consistencia deben hacer la misma estimación, esta no lleva necesariamente al mismo resultado numérico. Esta distinción resalta que el uso del teorema de Bayes es un arte, en cuanto cada individuo –según sus conocimientos y creencias– debe decidir cómo fija $P(a|c)$ y $P(b|a \cdot c)$. Una vez se fijan estos dos elementos, los pasos restantes son algebraicos.

comportamiento argumentando que para una persona en estado de ignorancia los errores en ambos sentidos son igualmente probables, de modo que existe una expectativa actuarial promedio basada en probabilidades iguales" (ibíd., p. 152). A una conclusión similar llegó cuando afirmó que en situaciones de incertidumbre "no hay una base científica para poder calcular cualquier tipo de probabilidad. Simplemente no sabemos" (Keynes, 1973, p. 114).

Además, hizo una importante distinción entre lo "incierto" y lo "improbable" (1936, p. 148), y remitió al capítulo 6 del *Tratado* como referencia[8], cuyo tema principal es la distinción entre el *peso* y la *probabilidad* de un argumento. Cuando hay más evidencia disponible que aumenta nuestro conocimiento se tiene más peso o una mejor base para sacar conclusiones. Sin embargo, puede suceder que la proposición que se desconoce tenga una baja probabilidad.[9]

El peso de un argumento se puede describir por sus propiedades básicas: 1) la información irrelevante no cambia el conocimiento de una proposición; por ello Keynes define la relevancia y el peso en forma correlacionada. 2) el peso de un argumento es mínimo cuando no hay evidencia, es decir, cuando solo se basa en probabilidades *a priori*. 3) los pesos de un argumento son comparables cuando la conclusión se evalúa con base en diferentes subconjuntos del mismo espacio de información (Keynes, 1921, p. 78). 4) aunque no es una propiedad, Keynes dice que el peso de un argumento no es lo mismo que los errores probables, a los que podemos interpretar como la desviación estándar. Es decir, aunque en la práctica la desviación estándar pueda ser una buena medida del peso de un argumento, en teoría es posible que se muevan en direcciones opuestas (ibíd., p. 82).

Utilizando estos conceptos en el contexto de la *Teoría General* se infiere que cuando Keynes habla de situaciones inciertas –en las que las probabilidades son desconocidas–, el peso del argumento es muy bajo. De modo que si la evidencia disponible es muy precaria para formar una base científica o racional a partir de la cual se pueda calcular algún tipo de probabilidad, la relevancia del conocimiento será baja y, por tanto, la incertidumbre será alta.

[8] Esta distinción tiene una clara relación con la incertidumbre de Knight. Por razones de espacio, no me extiendo sobre este tema. Los lectores interesados pueden consultar el artículo de Feduzi, Runde y Zappia (2013) y sus referencias.

[9] Por ejemplo, la proposición puede ser que una moneda va a caer cara. Si después de varios lanzamientos la moneda cae sello, tendremos una buena base para formar nuestro conocimiento sobre la moneda aunque la probabilidad de que la proposición sea cierta es baja.

Aunque el razonamiento que relaciona la precariedad del conocimiento con el peso de un argumento y la incertidumbre es convincente, es poco razonable aceptar literalmente la afirmación de Keynes de que implica que no existe una base científica para definir probabilidades. Una manera de mostrar que no es cierta es utilizar el concepto de entropía como medida del peso de un argumento[10]. También se puede mostrar –como veremos– que existe una base objetiva para determinar probabilidades con información limitada.

Comencemos señalando que la entropía es una medida de incertidumbre definida así[11]:

$$H(X) = -\sum_i^n p(X_i) \log p(X_i), \tag{17}$$

cuyas propiedades son semejantes a las del peso de un argumento. Supongamos que X es una proposición condicional de la forma $\mathrm{a}|\mathrm{h}$ y X_i es $\mathrm{a}_i|\mathrm{h}$, donde h es la evidencia disponible. Así, es cierto que (Cover y Thomas, 2006, pp. 20-22):

$$H(X) \geq 0, \tag{18.1}$$

$$H(X|Y) \leq H(X), \tag{18.2}$$

$$H(X) \leq \log n. \tag{18.3}$$

La ecuación (18.1) dice que la incertidumbre es siempre no negativa. Es igual a 0 cuando exista certeza, es decir, cuando $p(\mathrm{a}_i|\mathrm{h}) = 1$. Si $Y = \mathrm{h}'$ es información adicional, (18.2) dice que la evidencia adicional reduce la entropía si es relevante y la deja constante si no lo es. La ecuación (18.3) dice que la incertidumbre llega al máximo cuando la distribución es uniforme. Puesto que más información reduce la entropía, (18.3) refleja la segunda propiedad del peso de un argumento. Por último, utilizando el ejemplo de Keynes (1921, p. 82), es posible mostrar que la entropía –a diferencia de los errores probables– puede funcionar como medida del peso de un argumento.

Ejemplo 1

Con la evidencia inicial, un sistema se describe como indica el cuadro 1.

[10] La relación entre el concepto de entropía y el peso de un argumento fue señalada inicialmente por Cox, en un pie de página (1961, p. 105).

[11] Por brevedad, aquí basta decir que la entropía es una medida de incertidumbre. Para una presentación exhaustiva, ver Jaynes (2003), y Cover y Thomas (2006).

Cuadro 1

$p(x)$:	1/3	1/4	1/5	1/6	1/20
x	5	6	7	8	9

Con esta evidencia, la entropía de x es igual a 2,139 y la desviación estándar es 2,89. Luego Keynes supone que llega evidencia adicional que transforma el sistema inicial en:

Cuadro 2

$p(x)$:	7/16	5/16	4/16
x	5	8	9

Con la evidencia adicional, la entropía es igual a 1,54 y la desviación estándar es 4,13. Es decir, la incertidumbre se reduce con información adicional, como dice Keynes.

Las propiedades de la entropía muestran entonces que puede servir como referencia para medir el peso de un argumento y, además, como medida de la relevancia de la información a partir de la cual se forman las expectativas sobre el futuro.

El principio de indiferencia de Keynes y el de máxima entropía

Según Keynes (ibíd., p. 44), la medición numérica de la probabilidad solo es posible cuando existe un conjunto de alternativas exhaustivas, mutuamente excluyentes y *a priori* igualmente probables. Este razonamiento es consecuencia de que solo se pueden sumar unidades equivalentes.

Keynes denomina *principio de indiferencia* a la regla que utiliza para establecer probabilidades equivalentes. Una condición necesaria para aplicar el principio es que si no existe información que implique lo contrario, debe fijarse la misma probabilidad a todas las alternativas[12].

La pregunta que nos interesa responder, a la luz del *principio de indiferencia,* es si hay una base objetiva para que cualquier par de individuos llegue a la misma distribución de probabilidad aunque el conocimiento existente sea precario. Siguiendo a Jaynes (1957), veremos que el principio de máxima entropía da una respuesta[13].

[12] Además de esta condición, Keynes también requiere que se cumpla la condición de irrelevancia, es decir, que $a|hh' \neq a|h$ es cierta solo si h' es irrelevante. Si se cumplen ambas condiciones, el principio de indiferencia indica que la probabilidad de cada alternativa $a_i|h$ es $1/n$ si $\sum_{i=1}^{n} a_i|h = 1$ y $\sum_{i=1}^{q} a_i|h = a|h$.

[13] El principio de máxima entropía se puede considerar objetivo en cuanto determina la distribución más probable y menos sesgada consistente con la

Para ilustrar el principio de máxima entropía podemos suponer que cualquier proposición $x_i = a_i|h$ solo puede tomar valores discretos. La masa de probabilidad sobre cada x_i se define como $p_i = p(x_i)$, tal que la entropía es $H(x) = -\sum_i^n p_i \log pi$. En adición, suponga que hay conocimiento sobre algunos momentos que describen los valores de x, que representamos como $\overline{f(x)}_k = \sum_{i=1}^n p_i f(x_i)_k$ para $k = 1, \ldots, K$, donde K es el número de momentos. El problema que soluciona el principio de máxima entropía es que elige los valores de que son más probables y menos sesgados dado el conocimiento de los valores de cada $\overline{f(x)}_k$, para $k = 1, ..., K$.

Ejemplo 2

Supongamos que los empresarios tienen total incertidumbre sobre los valores posibles de los retornos del capital en un momento, salvo que conocen sus valores mínimo y máximo. Si los retornos toman valores en un especio discreto, tal que los retornos $x \in [x_{min}, x_{max}]$, el problema de optimización consiste en solucionar un lagrangiano de la forma:

$$-\sum_i^n p_i \log p_i + (\lambda_0 - 1)(1 - \sum_{i=1}^n p_i). \quad (19)$$

La solución a (19) es $p_i = \exp(-\lambda_0)$, y puesto que $\sum_{i=1}^n p_i = 1$, entonces $p_i = 1/n$. Esto es lo mismo que implica el principio de indiferencia, pero en este caso la igualdad de probabilidades se justifica porque hay máxima incertidumbre sobre los posibles valores x.

Ejemplo 3

Para extender el análisis, supongamos que los retornos x pueden tener cualquier valor en los números reales. En algún momento, un empresario puede tener información sobre el valor esperado del retorno, μ, y un nivel de confianza σ^2 en sus estimaciones. Así, el empresario puede tener una base sobre la forma de la densidad $p(x)$ como la que soluciona un lagrangiano de la forma:

$$-\int p(x) \log p(x)\, dx + (\lambda_0 - 1)(1-\int p(x)dx) + \lambda_1 (\mu - \int x\, p(x)dx) + \lambda_2 (\sigma^2 - \int x^2\, p(x)dx). \quad (20)$$

La información con la cual se determina la densidad $p(x)$ son los momentos que fija el empresario, de modo que realmente determinamos una densidad $p(x|h)$, donde h es la información previa. La solución a la ecuación (20) es bien conocida (ver, p. ej., Zellner, 1991): una distribución normal $p(x|\mu, \sigma^2) = N(\mu, \sigma^2)$.

información disponible (Jaynes, 1957). Es poco razonable que dos individuos tengan distribuciones de probabilidad distintas si tienen la misma información.

El objetivo de la introducción del principio de máxima entropía, *más que proponer un método general con el cual se deban* determinar las distribuciones de probabilidad en situaciones de información limitada, es mostrar que hay métodos para resolver este tipo de problemas. Y, adem*ás, que incluso en situaciones* de "conocimiento precario" puede haber una base objetiva para determinar probabilidades numéricas[14].

APLICACIONES DEL TEOREMA DE BAYES EN EL CONTEXTO DE LA *TEORÍA GENERAL*

La discusión anterior lleva a concluir que se pueden asignar probabilidades numéricas aunque el conocimiento inicial sea precario. El objetivo de esta sección es establecer una metodología para revisar algunos problemas relacionados con la incertidumbre del libro IV de la *Teoría General*.

APLICACIÓN 1: VALORACIÓN SUBJETIVA DE LOS RETORNOS FUTUROS

En las primeras tres secciones del capítulo 12, donde se habla de las expectativas de largo plazo, Keynes empieza haciendo referencia a la formación de las expectativas de un empresario. Este se guía con base en un conjunto de factores que considera relevantes, aunque no lo sean tanto como otros en los que no confía tanto. Entre los factores que pueden ser relevantes, Keynes (1936, p. 147) destaca el stock de activos de capital, el capital agregado, la demanda de los consumidores, los cambios en sus preferencias futuras, las variaciones de los salarios futuros, etc.

Además de que dependen de las predicciones más probables, las expectativas de largo plazo también dependen de la confianza del empresario en que pueda estar equivocado. Esta, que Keynes llama *estado de confianza*, disminuye cuanto más lejano sea el horizonte sobre de la predicción. Si el conocimiento con el cual se empieza es precario y el horizonte de evaluación de los retornos es suficientemente largo, Keynes no duda en afirmar que poco o nada se podrá decir de los factores que determinan los retornos futuros.

La evaluación del futuro se puede basar, además, en la *convención* de que las cosas no cambiarán sustancialmente con respecto al presente. La confianza que se tenga en esta convención depende, a su vez, de

[14] Aunque se pueda considerar que el principio de máxima entropía es un método general para asignar probabilidades, esto no debilita el argumento de que es posible determinar probabilidades de manera subjetiva. De hecho, a veces es preferible desprenderse del principio de máxima entropía y utilizar juicios de valor para determinar los componentes del teorema de Bayes.

que el sistema siga en calma y se mantenga más o menos estable. De acuerdo con Keynes (1936, pp. 153-154), esa estabilidad será menor cuanto mayor sea la parte de los activos que pertenece a personas que ignoran el negocio en cuestión, cuanto mayores sean las fluctuaciones de las ganancias de las inversiones actuales y cuanto más cambie la psicología de las masas por noticias irrelevantes que nada tienen que ver con los retornos probables.

Este argumento se puede utilizar para definir un modelo de probabilidad del empresario. Desde una perspectiva bayesiana, las variables futuras se interpretan como cualquier otra cantidad acerca de la cual existe incertidumbre: se miden usando una distribución de probabilidad con el teorema de Bayes. Denotamos los valores futuros de los retornos con un vector $\tilde{x} = (x_{T+1}, x_{T+2}, \dots, x_{T+N})$, donde N es el número de periodos sobre el que se hacen las predicciones.

Para continuar llamamos X al conjunto de factores que utiliza el empresario para prever los retornos futuros. Los q primeros son los que el empresario considera relevantes y los $n-q$ factores restantes los que considera irrelevantes. Si θ son los efectos de los factores sobre los retornos en un periodo futuro, el efecto medio de los primeros q factores se puede representar como un valor θ_0 con confianza $\sigma^2 V_0$ y el efecto medio de los factores restantes es a 0 con confianza $\sigma^2 V_{00}$. Suponiendo que este es el único conocimiento que tiene el empresario, el resultado del ejemplo 3 indica que una referencia objetiva para determinar la densidad de θ es $p(\theta) = N(\tilde{\theta}_0, \sigma^2 \tilde{V}_0)$, donde $\tilde{\theta}_0 = (\theta'_0, 0')'$ y $\tilde{V}_0 = (V_0, V_{00})$ es una matriz cuadrada con los elementos de V_0 y V_{00} [15].

Los retornos esperados dependen de la predicción más probable que se pueda hacer con base en los factores utilizados y del estado de confianza en la predicción. Si denotamos la predicción más probable de cada x_{T+j} como $X_{T+j}\,\theta$ ($j = 1, \dots N$) y el estado de confianza en la predicción como σ^2, de acuerdo con el resultado del ejemplo 3 tenemos entonces que la densidad de cada x_{T+i} se puede describir como una distribución normal con media $x_{T+i}\theta$ y varianza σ^2, tal que $p(\tilde{x}|\tilde{X}, \theta, \sigma^2) = \prod_{j=1}^{N} N(X_{T+j}\,\theta, \sigma^2)$. Puesto que el estado de confianza puede no ser conocido, este requiere también que se represente como probabilidad que, por simplicidad suponemos, es $\sigma^2 \sim IG\,(s_0, \upsilon_0)$ [16].

Por último, la convención de que el futuro puede no ser muy diferente del presente se puede representar como $E[X_{T+j}] = X_T$ y $V[X_{T+j}] = j\,\Sigma$,

[15] Si el conocimiento sobre el cual se determinan los valores de θ incluye datos observados d, $p(\theta)$ se remplazaría por d, $p(\theta|d)$, que corresponde a la densidad de probabilidad posterior.

[16] La distribución $IG\,(s_0, \upsilon_0)$ corresponde a una gamma inversa con parámetros s_0 y υ_0, que tiene una media $E(\sigma^2) = s_0/\upsilon_0 - 2)$ y una varianza $V(\sigma^2) = \frac{2}{\upsilon_0 - 4} E(\sigma^2)^2$.

para todo j=1, …, N , donde Σ es la “precariedad” de la creencia en esta convención. Por simplicidad, suponemos que se conoce Σ. Así, los valores de $\tilde{X} = (X_{T+1}, X_{T+2}, \ldots, X_{T+N})$ se pueden determinar con base en las propiedades de esa convención. El problema del empresario se reduce entonces a determinar:

$$p(\tilde{x}|\tilde{X}) = \iint p(\theta|\sigma^2)p(\sigma^2)p(\tilde{x}|\tilde{X}, \theta, \sigma^2)d\theta d\sigma^2. \tag{21}$$

En el formalismo bayesiano, (21) representa la *verosimilitud marginal* si el modelo se construye con independencia de los datos o la *densidad de predicción posterior* si se construye con base en datos observados. Aunque el modelo de la ecuación (21) es una aproximación elemental al argumento de Keynes, tiene una solución analítica (Bauwens, Lubrano y Richard, 2000, p. 61):

$$p(\tilde{x}|\tilde{X}) = t_N (\tilde{X}\ \theta^*, s^*, (I_N + \tilde{X}\ M^*\ \tilde{X}), \upsilon^*). \tag{22}$$

Esta es una distribución t con N grados de libertad, media $\tilde{X}\theta^*$ y varianza $\frac{s^*}{\upsilon^* - 2}$ $(I_N + \tilde{X}\ M^*\ \tilde{X})$, en la que si hay datos previos: $M^* = \tilde{V}_0$, $\theta^* = \theta_0$, $s^* = s_0$ y $\upsilon^* = \upsilon_0$. Si la densidad de predicción se forma con base en datos observados, X: $M^* = (\tilde{V}_0^{-1} + X'X)^{-1}$, $\theta^* = M^* (\tilde{V}_0^{-1}\ \theta_0 + X'\ X\hat{\theta})$, $s^* = s_0 + s + (\theta - \hat{\theta})'\ \tilde{V}_0 + (X'\ X)^{-1}\ (\theta - \hat{\theta})$ y $\upsilon^* = \upsilon_0 + T$, donde $\hat{\theta} = (X'\ X)^{-1}\ X'x$ $s = x'(I_T - X(X'X)^{-1}\ X'x$ y x son los datos observados de los retornos en T periodos.

La densidad de predicción permite ver analíticamente algunos puntos resaltados por Keynes:

- Si nuestra base de conocimiento es muy precaria, la confianza que podemos tener en nuestras predicciones es baja. Esto concuerda con el hecho de que si hay pocas observaciones X, M^*, es mayor y υ^* menor, de modo que aumenta la varianza de $\tilde{x}$.
- Además, nuestro estado de confianza en la predicción más probable es menor cuando se empieza con una peor base de conocimiento. Es decir, se reduce cuando tenemos menos datos o conocimiento para hacer las predicciones.
- Sin importar cuán buena sea nuestra base de conocimiento, es cada más difícil identificar los factores que determinan los retornos de las inversiones en el futuro, pues la varianza aumenta con el horizonte de predicción N y con los valores futuros de $\tilde{X}$. Si, además, tenemos poca certeza en los valores de $\tilde{X}$, es decir, si el parámetro Σ es un valor grande, la varianza de la predicción será aún mayor.

Este ejercicio de valoración subjetiva de los retornos futuros es *una* representación del estado de conocimiento del individuo. Pero cuando estima los retornos futuros, la persona que hace esta operación es consistente con los principios de probabilidad, en cuanto asigna una distribución a lo que es incierto, y luego procesa esa información con el teorema de Bayes.

Pero decir que representa el mundo real carece de sentido y lleva a malas interpretaciones. Si en el ejercicio decimos, por ejemplo, que después de dos periodos los retornos futuros estarán entre $[\tilde{x}_5, \tilde{x}_{95}]$ con una probabilidad del 90%, eso no significa que tenemos un 90% de confianza en que sus valores verdaderos estarán en ese intervalo ni que toda persona crea en esa afirmación con una confianza del 90%. Los intervalos de probabilidad son una representación de *su* modelo, *sus* supuestos y *sus* prejuicios. La fantasía de pensar que se representan aspectos del mundo real es quizá la principal limitante en la interpretación de la probabilidad.

Aplicación 2: decisiones de inversión bajo incertidumbre

El ejercicio anterior carece de un elemento esencial en el análisis de Keynes: que los retornos futuros dependen del monto de la inversión elegida, y viceversa. Esto complica las cosas porque el problema ya no es solo de inferencia, sino también de decisión.

De acuerdo con Keynes (1936, p. 136), el nivel actual de inversión corresponde al punto donde la eficiencia marginal del capital es igual a la tasa de interés en el momento de decidir. El problema es entonces que los valores de los retornos futuros con los que se calcula la eficiencia marginal son función del nivel de inversión actual y quizá también del nivel de inversión futura.

Desde el punto de vista de un observador externo que confía en que las fuerzas del mercado igualan las tasas de retorno futuras y la tasa de interés, i, se puede decir que existe una función de pérdida de la forma $L(i, \tilde{x}) = (r(\tilde{x}) - i)^2$, donde $r(\tilde{x})$ es:

$$r(\tilde{x}) = \frac{\sum_{j=1}^{N} \tilde{x}_j}{p_1}$$

y p_I lo que Keynes llama *precio de oferta* de un bien de capital. La intuición de la función de pérdida es que el observador externo favorece resultados en los que los retornos futuros del capital son iguales a la tasa de interés, quizá como indicación del buen funcionamiento del mercado[17].

Puesto que $L(i, \tilde{x})$ es una variable aleatoria porque no conocemos los retornos futuros $\tilde{x}$, nos interesa saber el valor esperado de la función

[17] La elección de la función de pérdida es un problema secundario. En gran medida, se elige la función cuadrática por su conveniencia analítica.

de pérdida con respecto a los posibles valores de $\tilde{x}$. Se debe tener en cuenta que ahora $\tilde{x}$ depende no solo de la matriz $\tilde{X}$ sino también del nivel de inversión actual y futuro, que en forma compacta denotamos $\tilde{\delta} = (\delta T, \delta_{T+1}, \ldots, \delta_{T+N-1})$.

Dada la dependencia de $\tilde{x}$ del nivel de inversión, el problema consiste en elegir los valores de $\tilde{\delta}$ que minimicen:

$$E\,(L(i, \tilde{x})) = \int (\mathrm{r}(\tilde{x}) - i)^2\, p(\tilde{x}\,|\,\tilde{X}, \tilde{\delta})\, d\tilde{x}. \tag{23}$$

Como ya analizamos la relación entre $\tilde{X}$ y $\tilde{x}$, podemos suponer sin pérdida de generalidad que $p(\tilde{x}\,|\,\tilde{X}, \tilde{\delta}\ \theta, \beta, \sigma^2) = N(\tilde{\delta}\beta, \sigma^2)$. Y para no usar métodos de Monte Carlo, que $p(\beta, \sigma^2) \propto \sigma^{-1}$, de modo que (Zellner, 1971, p. 321)

$$p(\tilde{x}\,|\,\tilde{\delta}) = \iint \mathrm{p}(\tilde{x}\,|\,\tilde{\delta}, \beta, \sigma^2) p(\beta, \sigma^2) d\beta d\sigma^2 = \frac{\Gamma(\frac{\upsilon+1}{2})}{\Gamma(\frac{1}{2})\,\Gamma(\frac{\upsilon}{2})\,\upsilon^{\frac{1}{2}}}\ \frac{\mathrm{g}^{\frac{1}{2}}}{\upsilon^{\frac{1}{2}}}\left[1 + \frac{\mathrm{g}}{\upsilon}\ (\tilde{x} - \tilde{\delta}\hat{\beta})^2\right]^{-\frac{\upsilon+1}{2}} \tag{24}$$

Es decir, suponiendo que hay T observaciones previas, la densidad de predicción tiene una distribución t, donde $\upsilon = T\text{-}1$, $\hat{\beta} = (\delta'\delta)^{-1}\,\delta' x$, $\mathrm{g} = [s^2(1 + \tilde{\delta}^2(\delta'\delta)\text{-}1)]^{-1}$, $\upsilon s^2 = (x - \hat{\beta}\,\delta)'(x - \hat{\beta}\,\delta)$ y $\delta = (\delta_0, \delta_1, \ldots, \delta_{T\text{-}1})$.

Si el horizonte de predicción es $N = 1$, el problema en (23) se convierte en minimizar:

$$E(L(\mathrm{i}, \tilde{x})) = E(\mathrm{r}(\tilde{x}) - E\,(\mathrm{r}(\tilde{x})) - (i - E\,(\mathrm{r}(\tilde{x}))))^2 = V(\mathrm{r}(\tilde{x})) + (i\text{-}E\,(\mathrm{r}(\tilde{x})))^2, \tag{25}$$

donde $E\,(\mathrm{r}(\tilde{x}))$ es la eficiencia marginal del capital. Es decir, en elegir la inversión con mínima varianza de los retornos futuros y la menor distancia entre la eficiencia marginal del capital y la tasa de interés. Si $\upsilon > 2$, el nivel óptimo de inversión es[18]:

$$\delta^*_T\ \delta^*_T = \frac{p_I i}{\hat{\beta}}\left(\frac{1}{1+\tau}\right), \text{ donde } \tau = \frac{\upsilon s^2 p_I}{(\upsilon\text{-}2)\delta'\delta\hat{\beta}^2} \tag{26}$$

Keynes (1936, p. 136) supone que un aumento de la inversión disminuye la eficiencia marginal del capital, en parte porque reduce los

[18] Dada (24), la varianza de $\mathrm{r}(\tilde{x})$ se puede calcular como $\frac{V(\tilde{x})}{p_I} = \frac{\upsilon s^2}{p_I(\upsilon\text{-}2)}\,(1 + \frac{\delta^2_T}{\delta'\delta})$, y la eficiencia marginal del capital se puede expresar como $E(r(\tilde{x})) = \delta_T \cdot \frac{\hat{\beta}}{p_I}$. Utilizando estas dos expresiones, la ecuación (25) se convierte en $\frac{\upsilon s^2}{p_I(\upsilon\text{-}2)}\,(1 + \frac{\delta^2_T}{\delta'\delta}) + (i\ \delta_T\,\frac{\hat{\beta}}{p_I})^2$. Es necesario señalar que al eliminar los efectos de X y $\tilde{X}$ en los retornos futuros, la inversión óptima es negativa dado el supuesto de Keynes de rendimientos marginales decrecientes, es decir, dado que $\hat{\beta} < 0$. En cambio, si $\tilde{X}\theta = \alpha > 0$ y se tienen en cuenta las observaciones de X, se tiene un intercepto positivo por el cual $\delta^*_T > 0$. Pero esto complica el análisis más de lo que el ejercicio merece.

retornos futuros esperados y en parte porque incrementa el precio de oferta[19]. En nuestro modelo esto se traduce en que $\beta < 2$, es decir, un aumento de δ_T está asociado a una reducción de $\tilde{x}$. A pesar de que el modelo es muy estilizado, el nivel de inversión óptima tiene características consistentes con lo que se espera en teoría. Por ejemplo, utilizando (26):

$$\frac{d\delta^*_T}{di} = \frac{p_I}{\hat{\beta}}\left(\frac{1}{1+\tau}\right) < 0, \tag{27.1}$$

$$\frac{d\delta^*_T}{dp_I} = \frac{i}{\hat{\beta}}\left(\frac{1}{1+\tau}\right)^2 < 0. \tag{27.2}$$

Es decir, como anticipa Keynes (1936, p. 137), un incremento de la tasa de interés y del precio de oferta reduce el nivel de inversión.

Este análisis se complica notablemente cuando $N > 1$, porque las decisiones futuras dependen de las decisiones presentes, y viceversa. Zellner (1991, pp. 340-343) presenta una solución a este problema utilizando el principio de *optimalidad de Bellman* para determinar la senda de valores óptimos de la variable de control. Pero este es un problema de técnica que no modifica sustancialmente el resultado que se obtiene cuando $N = 1$.

Los lectores habrán notado que los problemas de decisión desde un punto de vista bayesiano tienen en cuenta *todo* lo que es incierto y relevante en casos específicos. Desde esta *óptica*, el problema no es que exista incertidumbre sobre el futuro sino que tal incertidumbre no se modela como probabilidad[20].

Aplicación 3: especulación sobre los retornos futuros

En la aplicación 1 vimos que, según la información que tenga un empresario, se puede determinar la densidad de probabilidad de los retornos futuros. Lo que nos interesa ahora es cómo extender este ejercicio cuando la predicción de los retornos no solo se basa en la información de ese empresario sino que depende además de la psicología de las masas.

[19] En este ejercicio podríamos haber supuesto que p_i es función positiva de δ_T, tal que un aumento de la inversión incrementa el precio de oferta. Pero esto complicaría las matemáticas y requeriría métodos numéricos para obtener una solución.

[20] En la economía colombiana, un ejemplo notable de violación de este principio es la determinación de la regla fiscal, ya que se hacen supuestos sobre el futuro sin tener en cuenta la incertidumbre que implican.

El inversionista profesional, como lo llama Keynes, no se empeña en elaborar el mejor pronóstico de los rendimientos futuros, sino en predecirlos teniendo como base la opinión de las masas (1936, p. 155).

La percepción pública de los rendimientos futuros de un conjunto de activos se puede representar con una secuencia de números $t^{(n)} = (t_1, ..., T_n)$. Los valores $t^{(n)}$ indican la valoración puntual de las masas sobre los rendimientos futuros. La información de que dispone el inversionista sobre la opinión de las masas consiste en datos adicionales que le ayudan a predecir el valor de los retornos futuros; así, la densidad de predicción de $\tilde{x}$ se puede representar como:

$$p(\tilde{x}|t^{(n)}, \tilde{X}) \propto p(t^{(n)}|\tilde{x})p(\tilde{x}|\tilde{X}), \tag{28}$$

donde X son los factores que utiliza el inversionista en su predicción de los retornos futuros, $p(\tilde{x}|\tilde{X})$ es la densidad de predicción de (22) y $p(t^{(n)}|\tilde{x})$ refleja el juicio de valor que le da a la opinión de las masas al determinar el valor de $\tilde{x}$.

Con base en los datos $t^{(n)}$, el inversionista se puede guiar considerando ciertos momentos de $t^{(n)}$, como la mediana (m) y algunos fractiles inferiores (l) y superiores (u), para inferir la dispersión de las opiniones y sus posibles sesgos.

Para establecer la forma mediante la cual el inversionista profesional juzga la opinión de las masas debemos proponer una densidad de probabilidad. Por simplicidad y por su generalidad, seguimos la presentación de Lindley (1987) y utilizamos una densidad logística asimétrica de la forma:

$$p(t|\tilde{x}, s, k) = \frac{\kappa \exp\left(-\frac{\kappa(t-\lambda)}{\sigma}\right)}{\sigma\left[1 + \exp\left(-\frac{t-\lambda}{\sigma}\right)\right]^{\kappa+1}} \tag{29}$$

Donde $\lambda = \alpha + \beta\,\tilde{x}$, $\sigma = \gamma s$ y $\kappa = \delta k$. El valor de s es la dispersión de la opinión pública acerca de los valores de $\tilde{x}$, y k su grado de asimetría. La distribución es sesgada a la derecha si $\kappa < 1$, sesgada a la izquierda si $\kappa > 1$ y simétrica si $\kappa = 1$. Los valores de s y k se obtienen de los momentos m, l y u resultantes de los datos $t^{(n)}$[21].

[21] El primer paso es estimar los valores de l, m y u, utilizando los valores de $t^{(n)}$ y asignando una probabilidad $\in$. Dados esos valores, y sabiendo que para el fractil l de la densidad logística asimétrica estándar se cumple que $l = = -\log(\in^{-\frac{1}{k}} - 1)$, se tiene que para la densidad de (29) se cumple que $\frac{(l-\lambda)}{s} = -\log(\in^{-\frac{1}{k}} - 1)$, $\frac{(m-\lambda)}{s} = -\log(2^{-\frac{1}{k}} - 1)$ y $\frac{(u-\lambda)}{s} = -\log((1-\in)^{-\frac{1}{k}} - 1)$ (Lindley, 1987, p. 718). Así, fijando $\in$ en 0,05 y definiendo el grado de asimetría como $k = (u - m)/(m - l)$, las ecuaciones anteriores permiten inferir los valores de s y k utilizando el conocimiento de l, m y u.

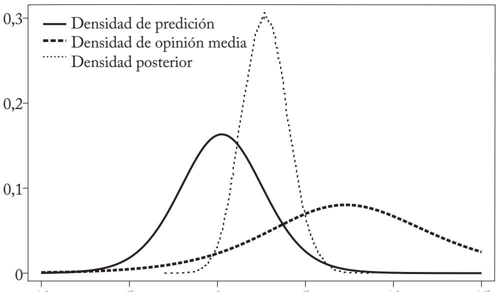
Los parámetros α y β representan la forma en que el inversionista profesional evalúa la opinión de las masas. Si $\alpha > 0$ y $\beta = 1$, pensará que, en promedio, las masas sobrevaloran los retornos. Si $\beta \neq 1$, pensará que el sesgo de dicha opinión es un porcentaje de $\tilde{x}$. El parámetro γ representa la confianza del inversionista en la opinión de las masas. Si $\gamma < 1$, exagerará el peso de la opinión pública en su efecto sobre los retornos futuros. Si $\gamma > 1$, preferirá su propio criterio frente al del público. Por último, el parámetro δ representa la precepción del inversionista de la valoración pública de los efectos de los valores extremos de $\tilde{x}$. Si $\delta > 1$, pensará que el público subestima los valores extremos de $\tilde{x}$ y los corrige con un valor más alto de δ.

Fijando exógenamente los valores de α, β, γ, y δ y determinando los valores de s y k con los datos $t^{(n)}$, la densidad posterior de los retornos futuros actualizada por la opinión de las masas se expresa como:

$$p(\tilde{x}|t^{(n)}, \tilde{X}, s, k) \propto p(t^{(n)}|\tilde{x}, s, k)p(\tilde{x}|\tilde{X}), \tag{30}$$

donde $p(t^{(n)}|\tilde{x}, s, k) = \prod_{i=1}^{n} p(t_i|\tilde{x}, s, k)$se representa tal como en (29), y $p(\tilde{x}|\tilde{X})$ es la densidad de predicción del inversionista. Por simplicidad, y para hacer énfasis en el análisis y no en el método, suponemos que $p(\tilde{x}|\tilde{X})$ es igual a la densidad de (24), que corresponde a una distribución t.

Gráfica 1
Densidades de probabilidad de los retornos futuros



Nota: La densidad de predicción es $p(\tilde{x}|\tilde{X})$, la densidad de opinión media es la verosimilitud medida en el promedio de $t^{(n)}$ fijando $\alpha = -0{,}1$, $\beta = 1{,}2$, $\gamma = 1{,}2$, y $\delta = 0{,}5$. La densidad posterior corresponde a $p(\tilde{x}|t^{(n)}, \tilde{X})$. En este ejercicio, $t^{(n)} = (-1; 1; 2; 2{,}3; 5; 12; 8; 7; 14)$, $\varepsilon = 0{,}05$, $\tilde{X} = 1$. La densidad de predicción supone que $\nu = 14$, $\hat{\beta} = 1{,}59$ $\nu s^2 = 80{,}4$ y $g = 0{,}17$. La densidad posterior se obtiene utilizando el algoritmo de Metropolis.

La gráfica 1 muestra intuitivamente cómo cambia la densidad de predicción del inversionista profesional cuando tiene en cuenta la opinión de las masas. El principal resultado es que, dependiendo cuán relevante sea para el inversionista la opinión pública acerca de los retornos futuros, así cambiará la densidad posterior con respecto a la densidad de predicción inicial. En general, cuanto mayor sea el dominio de la especulación, mayor será el dominio de la densidad $p(t^{(n)}|\tilde{x}, s, k)$ sobre $p(\tilde{x}|\tilde{X})$, pues sin importar cuán buenas seas las predicciones iniciales del inversionista, los retornos futuros serán determinados en su mayor parte por la opinión de las masas.

El inversionista profesional, además de ver que la opinión de las masas cambia la densidad de predicción de los retornos, puede estar interesado en predecir la opinión *promedio* de las personas en el futuro. En esto consiste la metáfora del juego de la belleza utilizada por Keynes (1936, p. 156) para describir la actividad del especulador. Utilizando el modelo de (30) como referencia, ese propósito se logra resolviendo la siguiente integral:

$$p(\tilde{t}\,|\tilde{X}, s, k) = \int p(\tilde{t}\,|\tilde{x}, s, k)p(\tilde{x}|\tilde{X})d\tilde{x}, \qquad (31)$$

donde $\tilde{t}$ son los valores futuros de la opinión pública acerca de los retornos futuros.

Este análisis de la metáfora del juego de la belleza es de gran importancia, entre otras cosas porque indica que las expectativas se forman como propiedades emergentes. Además, proporciona herramientas que complementan los supuestos de expectativas racionales y adaptativas comunes en la literatura. Ese análisis queda pendiente para futuras investigaciones.

Para concluir esta sección se debe reconocer que las aplicaciones mencionadas, más que resolver los problemas de la *Teoría General*, intentan establecer un método para analizarlos. El punto de partida no debe ser el rechazo de modelos que representen la incertidumbre, sino ver cómo se elaboran modelos de probabilidad que tengan sentido a la luz de cada problema específico.

CONCLUSIONES

Aunque el *Tratado* tuvo gran influencia en el contenido filosófico de la teoría de la probabilidad, el rechazo de la existencia de probabilidades numéricas limitó su impacto en la teoría matemática y ha sido objeto de malas interpretaciones de la posibilidad de representar la incertidumbre.

Este artículo muestra que la incertidumbre se puede representar como probabilidad, y que el debate sobre la pertinencia de esta representación se debe centrar en la elección de los supuestos y los modelos utilizados. La probabilidad, como argumenta Keynes, no está asociada a verdades empíricas. Desprendernos de esta ilusión es quizá el mensaje común de los enfoques keynesiano y bayesiano.

REFERENCIAS BIBLIOGRÁFICAS

Bateman, B. (1987). Keynes's changing conception of probability. *Economics and Philosophy, 3*(1), 97-119.

Bauwens, L., Lubrano, M. y Richard, J.-F. (2000). *Bayesian inference in dynamic econometric models.* Oxford: Oxford University Press.

Carabelli, A. (1988). *On Keynes's method.* Nueva York: Palgrave Macmillan.

Cover, T. y Thomas, J. (2006). *Elements of information theory.* Hoboken: Wiley & Sons.

Cox, R. (1961). *The algebra of probable inference.* Baltimore: The John Hopkins Press.

Feduzi, A. (2007). On the relationship between Keynes's conception of evidential weight and the Ellsberg Paradox. *Journal of Economic Psychology, 28*(5), 545-565.

Feduzi, A., Runde, J. y Zappia, C. (2013). De Finetti on Uncertainty. *Cambridge Journal of Economics, 38*(1), 1-21.

Jaynes, E. (1957). Information theory and statistical mechanics. *The Physical Review, 106*(4), 620-630.

Jaynes, E. (2003). *Probability theory: The logic of science.* Cambridge: Cambridge University Press.

Keynes, J. M. (1921). *A treatise on probability.* En E. Johnson y D. Moggridge (eds.), *The collected writings of John Maynard Keynes*, v. VIII. Cambridge: Cambridge University Press for the Royal Society.

Keynes, J. M. (1933). *Essays in biography* En E. Johnson y D. Moggridge (eds.), *The collected writings of John Maynard Keynes*, v. X. Cambridge: Cambridge University Press for the Royal Society.

Keynes, J. M. (1936). *The general theory.* En E. Johnson y D. Moggridge (eds.), *The collected writings of John Maynard Keynes* (vol. VII). Cambridge: Cambridge University Press for the Royal Society.

Keynes, J. M. (1973). *The general theory and after: Part II. Defence and development*. En E. Johnson y D. Moggridge (eds.), *The collected writings of John Maynard Keynes*, v. XIV. Cambridge: Cambridge University Press for the Royal Society.

Lindley, D. (1987). Using expert advice on a skew judgmental distribution. *Operations Research*, *35*(5), 716-721.

Lindley, D. (2000). The philosophy of statistics. *Journal of the Royal Society*. Series D, *49*(3), 293-337.

O'Donnel, R. (1989). *Keynes: Philosophy, economics and politics*. Nueva York: Palgrave Macmillan.

Patinkin, D. (1976). Keynes and econometrics: On the interaction between the macroeconomic revolutions of the interwar period. *Econometrica*, *44*(6), 1091-1123.

Ramsey, F. (1960). Truth and probability [1926]. En R. Braithwaite, *The foundations of mathematics and other logical essays*. Londres: Littlefield, Adams & Co.

Roncaglia, A. (2009). Keynes and probability: An assessment. *The European Journal of the History of Economic Thought*, *16*(3), 489-510.

Samuelson, P. (1946). Lord Keynes and the *General Theory*. *Econometrica*, *14*(3), 187-200.

Zellner, A. (1971). *An introduction to Bayesian inference in econometrics*. Nueva York: Wiley.

Zellner, A. (1991). Bayesian methods and entropy in economics and econometrics. En W. Grandy e I. Shick, *Maximum entropy and Bayesian methods. Fundamental theories of physics*. Dordrecht: Springer.