REAL ACADEMIA DE CIENCIAS
EXACTAS, FÍSICAS Y NATURALES DE ESPAÑA

LAS MATEMÁTICAS DEL ANÁLISIS DE DATOS IMPRECISOS

DISCURSO LEÍDO EN EL ACTO DE SU RECEPCIÓN
COMO ACADÉMICA DE NÚMERO POR LA

EXCMA. SRA. DÑA. MARÍA ÁNGELES GIL ÁLVAREZ

Y CONTESTACIÓN DEL

EXCMO. SR. D. FRANCISCO JAVIER GIRÓN GONZÁLEZ-TORRE

EL DÍA 25 DE ENERO DE 2023

MADRID
Domicilio de la Academia: Valverde, 22
www.rac.es

REAL ACADEMIA DE CIENCIAS
EXACTAS, FÍSICAS Y NATURALES DE ESPAÑA

LAS MATEMÁTICAS DEL ANÁLISIS DE DATOS IMPRECISOS

DISCURSO LEÍDO EN EL ACTO DE SU RECEPCIÓN
COMO ACADÉMICA DE NÚMERO POR LA

EXCMA. SRA. DÑA. MARÍA ÁNGELES GIL ÁLVAREZ

Y CONTESTACIÓN DEL

EXCMO. SR. D. FRANCISCO JAVIER GIRÓN GONZÁLEZ-TORRE

EL DÍA 25 DE ENERO DE 2023

MADRID
Domicilio de la Academia: Valverde, 22
www.rac.es

Depósito legal: M-29735-2022
ISSN: 0214-9540
ISBN: 978-84-87125-78-2

Índice general

Introducción

Excmo. Sr. Presidente de la Academia

Excmas. Sras. Académicas y Excmos. Sres. Académicos

Autoridades, compañeros y amigos, Sras. y Sres.

Querida familia

¡Gracias por vuestra presencia y compañía!

Primeros agradecimientos

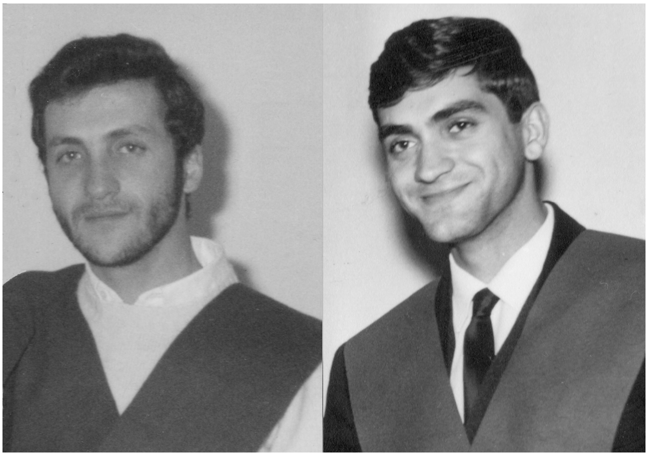
Este discurso de ingreso comienza con mi agradecimiento más profundo y sentido hacia los académicos que han firmado la propuesta: los Profesores Javier Girón, Enrique Castillo y Juan Luis Vázquez. Javier, que exceptuando a mi hermano Pedro Gil fue el primer estadístico al que conocí, ha sido un apoyo constante en muchos momentos de mi vida profesional y de mi vida personal. Alguien que, junto a Lina, sé que siempre están ahí si les necesito. Enrique y Juan Luis, ambos Doctores *Honoris Causa* de mi Universidad de Oviedo, han respaldado la candidatura y me han animado de diversas formas y en distintas ocasiones. De los tres he recibido sugerencias muy valiosas para este día de hoy y para el futuro en la Academia. ¡Gracias de corazón!

Mi gratitud es extensiva a los Profesores académicos David Ríos y Daniel Peña, que me brindaron una ayuda que se ha hecho efectiva repetidamente. También al Presidente y a la Secretaria de la Academia, los Profesores Jesús María Sanz Serna y Ana María Crespo de las Casas, por hacer tan sencillo y cordial el proceso previo a este ingreso.

Sobre la preparación del discurso

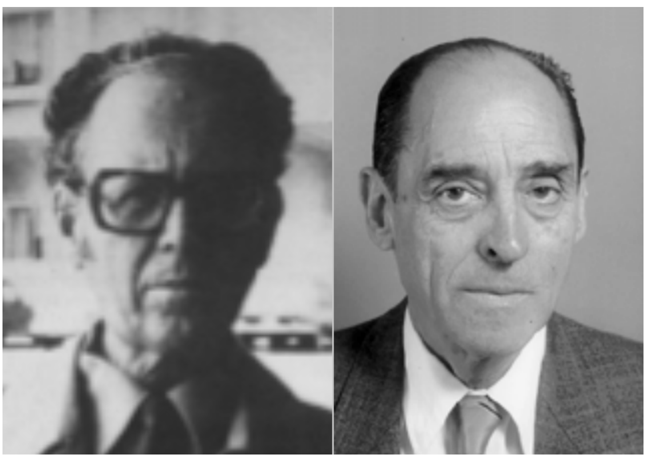
A lo largo de la preparación de este discurso me he sentido acompañada por cuatro estadísticos muy importantes para mí, que desafortunadamente ya no están.

Mis dos estadísticos del alma: mis hermanos Pedro y Jesus.



Pedro (izquierda) y Jesus (derecha) Gil Álvarez

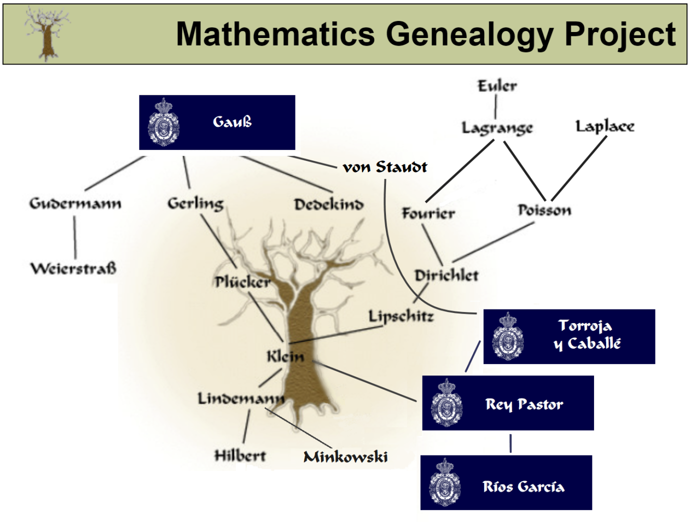
Y dos académicos inolvidables, de cuya mano visité por primera vez la Real Academia de Ciencias Exactas, Físicas y Naturales en 1985: los Profesores Sixto Ríos García y Francisco Azorín Poch.



Sixto Ríos (izquierda) y Francisco Azorín (derecha)

Con mis hermanos me encontré a los pocos minutos de llegar al mundo.

Al Profesor Sixto Ríos pude conocerle (¡al fin!) en un congreso de Estadística e Investigación Operativa celebrado en primavera de 1979, poco antes de que presidiera el tribunal que juzgaría mi tesis doctoral, dirigida por mi hermano Pedro. Científicamente hablando, al ser Pedro discípulo del Profesor Ríos, este se convirtió en mi 'abuelo'. De manera que, gracias a él, disfruto de un pedigrí científico de lujo indiscutible, como puede constatarse en el árbol construido a partir del *Mathematics Genealogy Project.* Continuamos viéndole en reuniones y contando con su presencia en varios actos de la Universidad de Oviedo, siendo seguramente el más inolvidable el de su investidura como Doctor *Honoris Causa* por la Facultad de Ciencias de la misma en el año 2000.



Árbol genealógico científico del Profesor Sixto Ríos García
Fuente: *Mathematics Genealogy Project*

Con el Profesor Azorín, mi primer encuentro se produjo en el verano de 1983, en una reunión sobre conjuntos *fuzzy* (en castellano, borrosos o difusos) y algunos problemas de la teoría de la información. Dos años más tarde tuvimos el honor de contar con él como presidente de la primera tesis doctoral que dirigí (codirigí con Pedro). A pesar de que las circunstancias de la época no propiciaban suficientemente mantener contactos asiduos sobre los muchos temas de interés comunes, y de que se 'marchó' demasiado pronto, aún tuvimos la oportunidad de escucharle en charlas y congresos, invitarle a presidir más tribunales de tesis doctorales (tres de las cuatro primeras que dirigí) y disfrutar del aprendizaje enriquecedor que implicaba cualquier conversación con Azorín.

La planificación de este discurso ha seguido una estrategia próxima a la que consideró David Ríos al tomar posesión de la plaza de académico en la RAC con la Medalla nº 49. Mientras él adoptó como base la lectura de una muestra de los discursos de ingreso en esta Academia (concretamente, la configurada por los veinticinco previos al suyo), yo he recurrido a la de una muestra algo más pequeña y matemática y me he deleitado releyendo con fruición algunos de esos discursos, prestando atención preferencial al del Profesor Sixto Ríos con contestación del Profesor Julio Rey Pastor (1961) y al del Profesor Francisco Azorín con contestación del Profesor Sixto Ríos (1981). Es decir, al muestreo inicial se le ha aplicado un posterior remuestreo, no aleatorio sino selectivo. El último discurso aludido ha sido una fuente de inspiración constante e inestimable y ha supuesto un auténtico regalo.

Analizando esa pequeña muestra, se advierte que comparto con Francisco Azorín Poch, David Ríos Insua y mi compañero de universidad durante tantos años Carlos López Otín, entre otros, el hecho de haberse acordado mi nombramiento como Académica Numeraria en una plaza de nueva creación (en mi caso, la correspondiente a la Medalla nº 55).

Las medallas académicas que previamente portaron Rey Pastor (Medalla nº 20), Sixto Ríos (Medalla nº 12) y Azorín (Medalla nº 38), las llevan actualmente con orgullo manifiesto Jesús Sanz Serna, Enrique Castillo y Javier Girón, respectivamente. Si bien el ser la número 55 de nueva creación me exculpa de glosar a mis 'predecesores' en su ocupación, quiero referirme a Rey Pastor, Sixto Ríos y Azorín plagiando a este último en su discurso de ingreso: *«Sea su memoria un permanente estímulo en nuestra diaria dedicación»*. Aunque sea una quimera, ¡cuánto me gustaría atesorar una pequeña parte de sus conocimientos y que me asistieran algo de la inimitable claridad de ideas, la vena poética y la simpatía natural de Rey Pastor, de la enorme empatía y el respeto por las opiniones de los demás de su discípulo Ríos y del dominio del lenguaje y una modestia a prueba de cualquier galardón o reconocimiento de Azorín!

He podido constatar también que en todos los discursos que he leído, lo que sospecho podría extrapolarse a la población de todos los discursos de ingreso en la RAC, se encierra cierto ‘sentimiento de sorpresa’ junto con algo cercano a una ‘petición de disculpa’ por haber sido nombrados académicos de número dados sus ‘escasos/menguados/...’ méritos. Desde el instante en que supe de mi nombramiento, decidí que iba a evitar calificar mis merecimientos. Si acaso, solo debo excusarme ante el Profesor Javier Girón. Cuando contestó al discurso de ingreso de Daniel Peña hace unos meses, la tarea era *a priori* sencilla: le habría bastado con describir simplemente una parte reducida de sus logros para conseguir una contestación verdaderamente sólida. En compensación, en esta ocasión se le demanda explotar al máximo su creatividad.

Aunque no haya querido dar vueltas sobre ‘qué he hecho yo para merecer esto’, sí quiero manifestar mi agradecimiento más profundo por (recurriendo de nuevo al discurso de Azorín) *«... la benevolencia con que tan generosamente habéis considerado mis limitados méritos»*. Mentiría si dijera que habéis cumplido uno de mis sueños más queridos, porque jamás soñé con ello. Es un honor tan inmenso que nunca he tenido ni la imaginación, ni la fiebre suficientemente alta, como para que alguna vez se me haya pasado por la cabeza. Mi agradecimiento, además de a la benevolencia, se extiende al cariño que he sentido.

Sobre la elección del contenido científico del discurso

La revisión de discursos de ingreso para elaborar el presente, se complementó con la de dos lecciones inaugurales de cursos académicos en la Universidad de Oviedo: la del curso 1913–1914, dictada por Julio Rey Pastor, y la del curso 1996–1997, impartida por Pedro Gil.

El ensayo en el Boletín Informativo de la Fundación Juan March titulado *“Julio Rey Pastor, matemático”*, escrito por Sixto Ríos (1986), y el artículo *“Recuerdo de Julio Rey Pastor”*, del académico D. Manuel López Pellicer (2015) en la Sección General de la Revista de esta Academia, dan buena cuenta de la trayectoria profesional de a quien recuerdan y permiten, además, ser más conscientes de su categoría y valía personales. Saber que me unen a él lazos de sangre científicos y que por un breve tiempo formó parte de mi Universidad, me hace sentir a Rey Pastor de un modo especialmente entrañable.

Con veintidós años había defendido su tesis doctoral sobre *“Correspondencia de Figuras Elementales”* bajo la supervisión del académico D. Eduardo Torroja y Caballé y con veintitrés, en

1911, obtuvo la cátedra de Análisis Matemático de la Universidad de Oviedo.

La fugacidad de su paso por la institución asturiana obedeció en gran medida a su ansia declarada de adquirir más conocimientos matemáticos desde las fuentes pertinentes, lo que le llevó, por mediación de ayudas de la Junta para Ampliación de Estudios, a trabajar con el Profesor Felix Klein en la Universidad Georgia Augusta de Gotinga en la Baja Sajonia. Gracias a ello, en 1912 amplió y perfeccionó su formación postdoctoral con su *"Teoría Geométrica de la Polaridad"*.

El académico D. Julio Rey Pastor,
que tomó posesión en noviembre de 1920 (Medalla nº 20)
Fuente: Archivo Espasa Calpe

Poco antes de recibir el encargo de elaborar la lección inaugural del curso 1913–1914 en la Universidad de Oviedo, había opositado a la cátedra de Análisis de la Universidad Central de Madrid (actual Complutense). Por ello y por la extensión de su estancia en Alemania, pese a que pudo escribir la lección, esta fue leída por el Profesor Rogelio Masip (matemático que fue padre y abuelo de alcaldes de la capital del Principado de Asturias).

Tomo prestadas sus palabras de introducción a la lección, que suscribo plenamente, para justificar la elección del tema de este discurso. El préstamo es muy adecuado, ya que en Rey Pastor la gracia era consustancial, inteligente y difícil de emular. Decía así: *«En ocasión análoga a esta, casi todos habéis comenzado dándonos cuenta de vuestras dudas y vacilaciones para la elección del tema que habíais de desarrollar. Dichosos vosotros que podéis dudar; porque vuestra cultura, tan extensa como profunda, os ofrece mil asuntos y problemas diversos, que os atraen a la vez y con igual fuerza. Este preliminar obligado falta en mi discurso. Yo no podía dudar porque no tenía dónde elegir…».*

El título del presente discurso, *"Las Matemáticas del Análisis de Datos Imprecisos"*, está inspirado en el de la lección inaugural dictada por Pedro Gil en la Universidad de Oviedo ochenta y tres años después de la de Rey Pastor. Esta se tituló *"Las Matemáticas de lo Incierto"*, haciendo referencia a un tema más general que incluía las matemáticas del azar (que identificaba con la Probabilidad y la Estadística), las de la imprecisión (que asociaba a la Teoría de los Conjuntos Borrosos o *fuzzy*) y las de la comunicación (en lo que afectaba a la Teoría de la Información).

Las Matemáticas del Análisis de Datos Imprecisos combinan las matemáticas del azar y las de la imprecisión de manera que las primeras conciernen a la 'generación aleatoria' de los datos experimentales y las segundas a la 'naturaleza' de tales datos.

No es la primera vez que aleatoriedad e imprecisión coexisten en un discurso de ingreso en la Academia, si bien el enfoque es muy diferente al anterior, puesto que han transcurrido más de cuarenta años y la mayoría de los fundamentos y de los métodos que van a describirse se han establecido y desarrollado a lo largo de ese tiempo.

En diciembre de 1981, el Profesor Francisco Azorín Poch tomó posesión como académico de número de esta Real Academia con la medalla de nueva creación número 38. Como ya se ha comentado, su discurso *"Conjuntos Borrosos, Estadística y Probabilidad"* fue contestado por el Profesor Sixto Ríos García.

En ese discurso, Azorín presentaba una excelente revisión de los conceptos y resultados básicos de la Teoría de Conjuntos Borrosos, introducidos por Lotfi A. Zadeh unos años antes, corroborando o anticipando su aplicabilidad, entonces más potencial que probada, en conexión con diversos problemas de la Estadística. Zadeh había señalado, a principios de los sesenta del siglo XX, la inexistencia de un marco matemático que pudiera hacer frente a la complejidad de muchos sistemas en los que no era posible establecer mediciones/valoraciones precisas y Azorín consideraba la nueva teoría como *«... un tratamiento más completo de la realidad para extender y profundizar su comprensión...»*.

Incidía el Profesor Azorín sobre la necesidad de una cobertura bien formalizada de esta teoría, apuntando que *«... las fuentes de vaguedad, de imprecisión, de incertidumbre, estadística y no estadística, probabilística y no probabilística, hacen deseable extender el concepto de conjunto nítido, pero cuidando de no caer en los inconvenientes que acompañaron al desarrollo de la probabilidad y de la estadística desde su introducción, y que serían la falta de rigor, coherencia, pertinencia y laconismo,...»* En este sentido, consideraba que la Teoría de Conjuntos Borrosos daba respuesta a ese deseo, puesto que *«... Los distintos aspectos de vaguedad o borrosidad antes mencionados, han pasado por una larga etapa, desde la simple constatación de la existencia de zonas intermedias o grises, pasando por intuiciones esporádicas, aplicaciones empíricas y algunas anticipaciones de importancia, hasta el trabajo de Lofti Zadeh, titulado "Fuzzy Sets" publicado en 1964 en Memo. ERL, núm. 64–44 de la Universidad de California en Berkeley y en 1965 en la revista Information and Control.»*

Y, en lo concerniente a su interacción con la probabilidad y la estadística, escribía que: *«Es pues necesario, establecer un tratamiento riguroso y nítido de la vaguedad, conectar la teoría y la práctica de los conjuntos borrosos con otros resultados y aplicaciones y, en particular, con los que se refieren al tratamiento establecido de la incertidumbre y del error, por medio de la probabilidad y de la estadística.»*

En su contestación, el Profesor Sixto Ríos comentaba que se había interesado recientemente por los conjuntos borrosos en relación con su investigación sobre decisiones multicriterio, aclarando que *«... el punto de partida de las investigaciones de Zadeh fue construir una teoría de conjuntos difusos que nos aproximara al razonamiento en la vida corriente, rechazando la idea de que el razonamiento podría ser especificado y refinado para ajustarse a la lógica formal exacta clásica, y adoptando, en cambio, el punto de vista de que la potencia del razonamiento humano reside precisamente en la habilidad para tratar directamente con conceptos inexactos o difusos.»* Pronosticaba también Ríos que, tras las aplicaciones al control de sistemas que acababan de ver la luz, *«... ello hace esperar que otros éxitos serán pronto realizados en los campos de las ciencias sociales, políticas, económicas, en que, por naturaleza, las lingüísticas y los razonamientos aproximados son más naturales que la cuantificación* *numérica de las variables y los razonamientos exactos,...»* lo que iba en total consonancia con las aplicaciones que Zadeh tuvo siempre en mente.

Pedro Gil fue conocedor del trabajo de Zadeh sobre conjuntos borrosos/difusos en el curso de la preparación de los ejercicios de la Agregaduría en Investigación Operativa para la Universidad de Oviedo, que se celebraría en 1976. En concreto, el sexto ejercicio consistía en una prueba escrita y de defensa oral de un tema a elección del tribunal, que presidía el Profesor Ríos, y el trabajo publicado en *Information and Control* en 1965 fue uno de los temas elegibles a propuesta de Ríos.

Cuando los primeros discípulos de Pedro en la Universidad de Oviedo comenzamos a realizar las tesis doctorales bajo su dirección, no recuerdo que los 'conjuntos *fuzzy*' estuvieran entre los temas sobre los que trabajar. Sí recuerdo bien que hacia 1982 nos habló bastante acerca de ellos y de su interés previsible en clara connivencia con la probabilidad y la estadística. Muy posiblemente habría llegado a sus oídos el contenido del discurso pronunciado por Azorín y aquello pudo representar, de algún modo, el detonante de la investigación a la que se refieren varias de las secciones de este discurso de ingreso.

Nuestro propósito fue abordar la estadística con datos imprecisos, entendiendo la imprecisión en la forma considerada por el propio Zadeh, dejando al margen otras connotaciones del término, en muchos casos de carácter filosófico. Al principio optamos por datos provenientes de una variable aleatoria con valores reales en los que la imprecisión surgía en la observación o percepción de tales valores y las conclusiones estadísticas extraídas se referían a la distribución de la variable original subyacente. El modelo empleado permitía definir un espacio de probabilidad discreto, de forma un tanto artificial, a partir de algunas nociones y suposiciones establecidas expresamente (sistema ortogonal de valores *fuzzy* y la probabilidad de tales valores según Zadeh), pero sin venir inducidas directamente dentro del contexto probabilístico.

A mediados de los ochenta se publicaron los primeros trabajos sobre los elementos aleatorios con valores *fuzzy*, que constituyen un modelo matemático sólidamente establecido dentro del marco probabilístico para los mecanismos que generan aleatoriamente datos intrínsecamente imprecisos, que en el mundo real provienen mayoritariamente de valoraciones humanas. Sobre la base de tal modelo y de resultados probabilísticos para dichos elementos aleatorios, se han ido desarrollando en los últimos años varios métodos de análisis inferencial con datos *fuzzy*.

En este discurso se describen sucintamente, en primer lugar, los fundamentos para dicho análisis: los elementos aleatorios generalizados, los conjuntos *fuzzy* y una métrica operativa entre ellos, los conjuntos (números en el caso unidimensional) *fuzzy* aleatorios y resultados e implicaciones más relevantes.

Posteriormente, se resumen varios de los procedimientos inferenciales con datos *fuzzy* disponibles hasta el momento. Además, se exponen dos de las primeras aplicaciones de esa metodología del análisis de datos *fuzzy*: una de índole teórica, con implicaciones para el análisis inferencial sobre la distribución de una o varias variables aleatorias con valores numéricos; otra de índole práctica, relativa al empleo de una escala de valoración *fuzzy* de asignación libre como base para el diseño de las respuestas en los ítems de muchos cuestionarios.

Los elementos aleatorios *à la* Fréchet

La naturaleza, la ciencia y la tecnología
ofrecen numerosos ejemplos de elementos aleatorios
que no son ni números, ni series, ni vectores, ni funciones...

Pero, ¿es posible decir algo útil sobre elementos aleatorios
cuya naturaleza no conocemos (o más exactamente
cuando no tenemos en cuenta su naturaleza, sea cual sea)?

(Maurice R. Fréchet)

Las matemáticas del análisis de datos provenientes de la medición o de la observación de una magnitud asociada a un experimento aleatorio comienzan con el establecimiento de los modelos correspondientes a dicho experimento y a tal magnitud.

Si bien en 1925 el matemático francés Paul Lévy escribió el que se considera como el primer libro de probabilidad que recurría a ideas de la teoría de la medida, la monografía que el soviético Andrei Kolmogórov publicó en 1933 mostró de forma natural que todas las ideas principales de la probabilidad podían formularse en términos de la teoría de la medida. Esta había tenido un desarrollo notable en los primeros años del siglo XX, gracias, entre otros, a los matemáticos franceses Camille Jordan, Émile Borel y Henri Lebesgue.

Como consecuencia directa, la probabilidad pasó de ser vista como una simple colección de cálculos a valorarse como una teoría matemática en la que los elementos básicos eran los dos modelos antedichos.

El que concierne al experimento aleatorio consiste en una terna, el *espacio de probabilidad*, compuesta por: el conjunto de posibles resultados experimentales, la clase de los sucesos de interés observables y la probabilidad asignada a la ocurrencia de esos sucesos.

Con el objeto de garantizar a un tiempo la aplicabilidad del modelo y su operatividad matemática, se imponen ciertos requisitos formales (la condición estructural de σ-álgebra para la segunda componente de la terna y las condiciones axiomáticas de no negatividad, normalización y aditividad numerable para la tercera). Estos requisitos no son habitualmente limitantes y permiten identificar los sucesos con subconjuntos de resultados, operaciones entre sucesos con operaciones entre conjuntos y deducir un buen número de propiedades de interés práctico para cualquier procedimiento de asignación probabilística.

Los axiomas de Kolmogórov dieron respuesta así a uno de los problemas planteados por David Hilbert (el Problema 6, sobre el tratamiento matemático de los axiomas de la Física, uno de cuyos objetivos era abordar mediante axiomas la teoría de las probabilidades) que fue expuesto en el *Congreso Internacional de Matemáticos* celebrado en la Universidad de la Sorbona de París en 1900.

Los números aleatorios y primeras generalizaciones

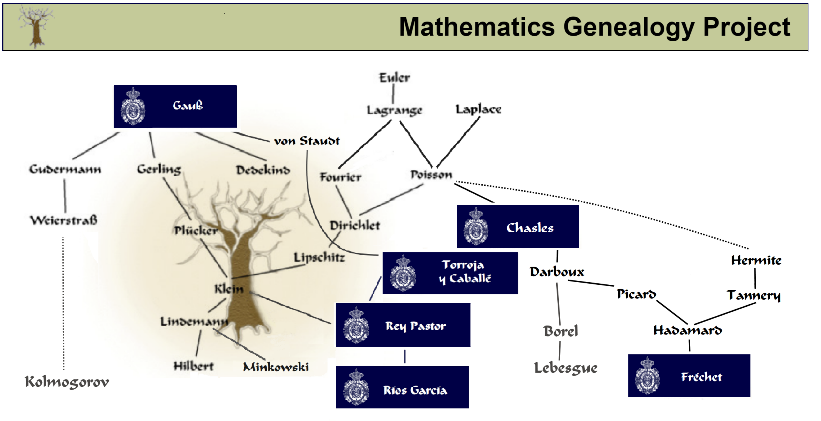
El modelo para las magnitudes cuantitativas, es decir, que atribuyen a cada resultado experimental un valor numérico, son las llamadas <u>*variables aleatorias*</u> unidimensionales (o elementos aleatorios con valores reales), que se definen como funciones del espacio de resultados en el espacio de los números reales. También por razones de operatividad matemática se impone a estas funciones reales una condición: la medibilidad Borel.

Aunque se considera a Pafnuti Tchebycheff como el primero en razonar sistemáticamente en términos de variables aleatorias, de sus medias y de sus momentos, la medibilidad exigida por Kolmogórov (exigencia poco restrictiva en la práctica) asegura que tengan sentido las probabilidades de que la variable tome valores en casi cualquiera de los conjuntos reales que podamos imaginar (los borelianos). Tales probabilidades determinan la denominada ley o distribución de probabilidad de esa variable aleatoria a la que pueden asociarse diversas medidas resumen o valores representativos de interés.

La teoría de la probabilidad fue incorporando sucesivamente al estudio de los números aleatorios (como Fréchet denominaba a las variables aleatorias con valores en el espacio euclídeo unidimensional) el de otros elementos aleatorios de dimensión más alta o más generales, como: el de los puntos aleatorios en el plano o en el espacio (o variables aleatorias con valores en espacios euclídeos de dimensión 2 o 3, respectivamente), el de las series aleatorias, el de los vectores aleatorios y el de las funciones numéricas aleatorias de variables numéricas no aleatorias (como los procesos estocásticos). Merecen mención expresa al respecto las contribuciones de matemáticos soviéticos (como Aleksandr Khinchin y Kolmogórov, entre otros), de matemáticos franceses (como Maurice Fréchet y Paul Lévy y su discípulo, de origen alemán, Wolfgang Döblin), del sueco Harald Cramér, del croata William Feller y del estadounidense Joseph Doob.

Elementos aleatorios de cualquier naturaleza

Hacia finales de los cuarenta del siglo pasado, el matemático francés **Maurice R. Fréchet**, académico extranjero de esta Academia desde 1967 y actualmente académico histórico de la misma, quiso dirigir por un tiempo su fascinación por las teorías y conceptos más generales y abstractos al estudio de los elementos aleatorios.



Árbol genealógico científico del Profesor Maurice Fréchet
Fuente: *Mathematics Genealogy Project*

Considerado como uno de los fundadores del análisis moderno, en su tesis doctoral defendida en 1906 bajo la supervisión de Jacques Hadamard se introdujo la noción de espacio métrico abstracto y fueron muy destacables sus contribuciones al análisis funcional y a la topología general.

Tras la Primera Guerra Mundial, fue contratado como profesor en la Facultad de Ciencias de la Universidad de la recién liberada Estrasburgo, donde empezó a dar clases de probabilidad y estadística; sus primeras investigaciones en el campo datan de esa época. En 1928 se creó el Instituto Henri Poincaré, como parte de la Facultad de Ciencias de la Sorbona. El cálculo de probabilidades ya tenía su gran exponente en dicha universidad en la persona del Profesor Émile Borel, quien quiso que Fréchet se sumara a la docencia en la misma.

Aunque son muy conocidas sus investigaciones sobre modelos de distribuciones, probabilidades en sistemas de sucesos compatibles y dependientes, cópulas o la acotación inferior de la varianza de los estimadores, etc., una de las líneas en las que el propio Fréchet depositó más entusiasmo fue la de su teoría de los elementos aleatorios de cualquier naturaleza. A pesar de tratarse de una de las últimas líneas abiertas en su carrera investigadora, y de que a veces se señala que su época científica más fructífera fue la relativa al análisis, las bases de datos principales sobre citas confirman inequívocamente que su trabajo más citado es aquel en que se introduce dicha línea.

"Les éléments aléatoires de nature quelconque dans un espace distancié" fue publicado en 1948, cuando Fréchet cumplía setenta años, en los prestigiosos *Annales de l'Institut Henri Poincaré*. En otro artículo del mismo año en *Mathematics Magazine*, se refería al tema del anterior como a un nuevo capítulo en la teoría de la probabilidad del que se abrían muchos problemas futuros y que había dado lugar a varios cursos que se habían expuesto en la Sorbona en 1945 y 1946.

De hecho, en uno de los trabajos del primer volumen de la revista *Trabajos de Estadística* (fundada por Sixto Ríos García) aparece resumido un ciclo de conferencias sobre esos elementos aleatorios que Fréchet impartió, en febrero de 1950, en el Departamento de Estadística del CSIC.

Como motivación de la introducción de estos nuevos elementos aleatorios, Fréchet explicaba que:

«El Cálculo de Probabilidades ha sido implícita o explícitamente, hasta una época reciente, el estudio de los números aleatorios y de puntos aleatorios en un espacio de dimensión uno, dos o tres (probabilidades geométricas).»

«Poco después, se han querido extender a menudo los resultados obtenidos a las series aleatorias, a los vectores aleatorios y a las funciones reales aleatorias de variables reales no aleatorias.»

«... Se han estudiado los números aleatorios obtenidos al elegir al azar una ciudad de cierto país y anotar un número relativo a esa ciudad, como el número de habitantes, el número de casas, etc. Pero también podrían considerarse elementos asociados a una ciudad elegida al azar que no pueden describirse mediante alguna de las nociones matemáticas usuales: número, función, curva, etc. Por ejemplo, la moralidad de su población, su tendencia política, la impresión de belleza que transmite, etc. Es una categoría nueva de elementos aleatorios.»

Fréchet sugería que, en lugar de tratar cada tipo de elemento aleatorio separada y sucesivamente, sería interesante establecer una teoría de los elementos aleatorios 'abstractos' con el objeto de unificar en la medida de lo posible todas las ideas y rasgos comunes ignorando la naturaleza de sus valores. Sabía que, a simple vista, su objetivo podría parecer demasiado audaz, pero que en realidad y en condiciones bastante generales su consecución no tenía nada de misterioso.

Como la nueva teoría debía incluir como caso particular las de los elementos aleatorios ya consolidados, tendrían que extenderse las definiciones clásicas y tratar de generalizar enunciados y demostraciones de los resultados fundamentales. No obstante, había que ser conscientes de que algunos conceptos podrían carecer de sentido (por ejemplo, intentar extender la función de distribución de un número aleatorio sería una fantasía si el conjunto de posibles valores del elemento aleatorio abstracto no dispusiera de un orden total universalmente aceptable). En ese caso habría que obviar su generalización o reemplazarla, si fuera viable, por alguna alternativa que sí tuviera sentido.

Para alcanzar su propósito, Fréchet estimó suficiente suponer que los elementos aleatorios considerados elegían sus valores dentro de un espacio métrico. La existencia de una distancia sobre el espacio de los posibles valores del elemento aleatorio permitía sin dificultades, entre otros:

- formalizar la medibilidad Borel de dicho elemento y, en consecuencia, la ley o distribución de probabilidad (inducida) del mismo;
- definir medidas de la dispersión (tipo cuadrática) de esa distribución;
- definir el valor medio de la distribución, minimizando dicha dispersión;
- definir varios tipos de convergencias estocásticas.

De este modo, los ***elementos aleatorios à la Fréchet***, funciones medibles Borel del espacio de probabilidad inicial en un espacio medible asociado al espacio métrico, constituyen un modelo matemático apropiado para las magnitudes que generan aleatoriamente datos con valores de cualquier naturaleza que estén dentro de un espacio dotado de una distancia (por ejemplo, los datos imprecisos modelizados mediante conjuntos *fuzzy* y alguna métrica oportuna, que se verán más tarde).

Entre los elementos aleatorios que se desarrollaron, estudiaron y aplicaron exhaustivamente años después, cabe destacar los conjuntos aleatorios compactos y convexos, en los que el espacio métrico está compuesto por la clase de los conjuntos compactos, convexos no vacíos de un espacio euclídeo y, usualmente, la distancia de Hausdorff correspondiente (ver, por ejemplo, Kendall, 1974, Matheron, 1975, Ripley, 1976 y Molchanov, 2005). Previamente, Herbert E. Robbins (1944, 1945) había dado una formulación matemática sólida de los conjuntos aleatorios, investigando sus relaciones con las probabilidades geométricas.

Relaciones de Maurice Fréchet con España

Maurice Fréchet mantuvo una correspondencia científica asidua con matemáticos de todo el mundo, especialmente soviéticos y nacidos o residentes en los EEUU. Pero, también son reseñables sus intercambios epistolares con matemáticos y estadísticos españoles y portugueses, así como sus visitas a nuestro país.

Hay constancia, entre otras, de la presentación de una conferencia sobre su trabajo *"Sur une définition du nombre de dimensions"* en la Real Academia de Ciencias de Madrid en 1927, de la exposición de una serie de charlas en el Instituto Jorge Juan de Matemáticas del CSIC en 1942 y de la impartición de cursos y conferencias tanto en la Facultad de Ciencias de la Universidad Central de Madrid como en el CSIC en 1950.

También aparecen diversas publicaciones suyas en revistas científicas españolas, como la *Revista de la Real Academia de Ciencias de Madrid* o *Trabajos de Estadística.*

Debe destacarse la influencia de Fréchet, junto con la de Georges Darmois[1], en aspectos cruciales de la enseñanza de la

[1]Darmois sucedió como Director del Departamento de Cálculo de Probabilidades y Física Matemática de la Sorbona a Maurice Fréchet quien, a su vez, había sucedido a Émile Borel. Los tres presidieron, en diferentes años, la Sociedad Estadística de París.

Arriba: Conferencia del Profesor Maurice Fréchet
en el Instituto Jorge Juan del CSIC (1942)
Debajo: Inauguración de los Cursos de Estadística
en la Facultad de Ciencias de la Universidad de Madrid (1950)
Maurice Fréchet y Sixto Ríos García en el centro de la foto
(izquierda y derecha, respectivamente)

Estadística en España y de la creación en 1952 de la Escuela de Estadística, precursora de la actual Facultad de Estudios Estadísticos de la Complutense.

Los cursos que acaban de referirse, la correspondencia mantenida con Sixto Ríos (gran impulsor de tal creación) y los desplazamientos realizados en ambos sentidos para discutir sobre ese respecto, supusieron un apoyo indiscutible.

Los valores de conjunto *fuzzy* como modelo para los datos imprecisos

Contar lo que se puede contar,
medir lo que se puede medir
y hacer medible lo que no se puede medir

(Atribuida a Galileo Galilei,
entre otros por Hermann Weyl, 1947)

Paradójicamente, una de las contribuciones
principales de la lógica fuzzy, ...,
es su elevado potencial para 'precisar' lo que es impreciso

(Lotfi A. Zadeh)

Buena parte de los ejemplos con los que Fréchet ilustraba la necesidad de ampliar los espacios de valores de los elementos aleatorios abstractos parecían anticipar la noción que, unos años después, llevaría al Profesor **Lotfi A. Zadeh** a introducir los conjuntos *fuzzy*.

Aunque Zadeh no era matemático, sino ingeniero eléctrico, su afecto por la probabilidad y la estadística quedaban de manifiesto en uno de sus últimos trabajos (2015), en el que revisaba su pasado investigador diciendo:

«Los primeros años en mi carrera académica coincidieron con el nacimiento de la era de las computadoras y la información. Fue un período emocionante, estimulado por la competencia entre los Estados Unidos y la Unión Soviética. En Columbia, mi investigación se centró en la teoría de sistemas y los sistemas de información. La teoría de la probabilidad ocupó una posición central en mi actividad. Mi primer artículo, publicado en 1949 en el Journal of Applied Physics, se tituló "Probability criterion for the design of servomechanisms". Mi segundo artículo, publicado también en el Journal of Applied Physics en 1950, se tituló "An extension of Wiener's theory of prediction".»

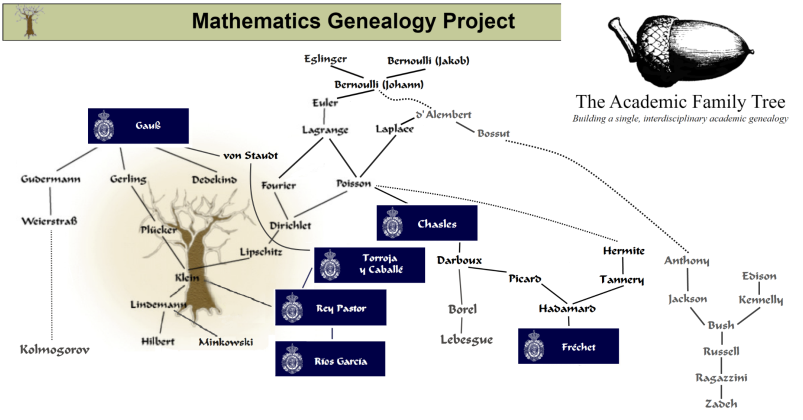
Y, señalaba también:

«Tuve una estrecha relación con el Departamento de Estadística Matemática y con su director, Herbert Robbins, un matemático brillante que se convirtió en mi mejor amigo...».

Esa fue una amistad que, años más tarde, en la Universidad de Berkeley, hizo extensiva a muchos otros estadísticos, como Jerzy Neyman o David Blackwell.

En realidad, muchos de los títulos de las primeras publicaciones de Zadeh involucran términos indudablemente estadísticos como 'operadores estocásticos', 'funciones de correlación', 'predicción', 'series temporales', etc. Lo mismo se aplica a varias comunicaciones presentadas en congresos de la *American Mathematical Society*.

No es de extrañar, por tanto, que pueda encontrarse a Lotfi A. Zadeh buscando en el *Mathematics Genealogy Project*, si bien para ampliar más su árbol genealógico científico haya que recurrir a otras fuentes como *The Academic Family Tree*, descubriendo así su descendencia, entre otros, de Thomas Alva Edison y, echando la vista hacia varias generaciones atrás, de Johann y Jakob Bernoulli.



Árbol genealógico científico del Profesor Lotfi A. Zadeh
Fuentes: *Mathematics Genealogy Project*,
y *The Academic Family Tree*

Tras graduarse en Ingeniería Eléctrica en la Universidad de Teherán, Zadeh realizó el máster en el Instituto Tecnológico de Massachusetts, donde quedó impactado por las lecciones recibidas del matemático estadounidense Norbert Wiener (que había colaborado con Fréchet). En 1946 pudo iniciar su doctorado en la Universidad de Columbia en Nueva York, bajo la supervisión de

John R. Ragazzini, quien pocos años después tuvo también como estudiantes de doctorado a Rudolf E. Kálmán, Eliahu I. Jury y Gene F. Franklin, cuyas contribuciones en el control de sistemas y, en particular, en los filtros para la detección de señales, han sido muy influyentes.

Además del trabajo seminal de Zadeh y Ragazzini ya referido, *"An extension of Wiener's theory of prediction"*, del que Zadeh había escrito una versión preliminar interna en 1949, otro trabajo conjunto de relevancia fue *"The analysis of sampled-data systems"*, publicado en 1952 en *Transactions of the American Institute of Electrical Engineers* y que llevó al conocido método de 'la transformada Z' para el procesamiento de señales.

Una vez defendida su tesis doctoral, Zadeh fue contratado en Columbia en 1950 como *assistant professor*, promocionándose en 1953 a *associate professor* y en 1957 a *professor*. Si bien durante cierto tiempo se había ido alejando de las matemáticas tradicionales, que consideraba algo insuficientes en el estudio de los sistemas reales, pronto retomó su afición por las mismas. Conociendo su atracción por la lógica desde 1950, cuando predijo que esta, en particular la multivaluada, iría cobrando cada día más interés, su amigo estadístico Herbert E. Robbins y el topólogo Deane Montgomery, que desde 1951 era miembro permanente del Instituto de Estudios Avanzados de Princeton en Nueva Jersey, respaldaron a Zadeh para una estancia en el mismo. Su propuesta fue aceptada, a pesar de que las solicitudes de científicos que no fueran matemáticos, físicos teóricos o historiadores difícilmente recibían respuesta positiva.

En el año 1959, pasó a formar parte como *professor* del Departamento de Ingeniería Eléctrica de la Universidad de California en Berkeley. Cuando en 1963 fue nombrado director del mismo, dedicó grandes esfuerzos a fortalecer el apoyo y la expansión de las ciencias de la computación, emergentes en aquel momento y por las que él apostaba decididamente. En 1968 renunció a su cargo, dejando convertido su departamento en el actual Departamento de Ingeniería Eléctrica y Ciencias de la Computación. Berkeley había comenzado así su avance imparable hacia los puestos de liderazgo que ha seguido ocupando en Ciencias de la Computación.

La teoría de los conjuntos *fuzzy*

Las tecnologías de la información y la comunicación en aquel tiempo habían dado lugar a la construcción y al diseño de sistemas muy complejos, cuyo estudio conllevaba el empleo de métodos exactos de identificación, clasificación, caracterización, evaluación, etc. Ello obligaba a establecer diferenciaciones meticulosas de casos y a adaptar definiciones a las circunstancias reales, para lo cual el lenguaje matemático tradicional resultaba con frecuencia escaso.

Según sus propias palabras en 1980:

«Como teórico de sistemas con orientación matemática, me habían condicionado a creer que las herramientas analíticas basadas en la teoría de conjuntos y la lógica binaria eran todo lo que se necesitaba en la construcción de un marco para un cuerpo de técnicas precisas, rigurosas y eficientes para el análisis de cualquier tipo de sistema, bien sea natural o hecho por el hombre. En el periodo 1961–1963, mientras preparaba un libro sobre teoría de sistemas con Charles A. Desoer, empecé a ser consciente de que los sistemas complejos no podían tratarse eficazmente mediante el uso de enfoques convencionales; en buena parte, porque los lenguajes de descripción basados en las matemáticas clásicas no eran lo bastante expresivos para caracterizar relaciones 'input-output' en ambiente de imprecisión, de incertidumbre y de información incompleta.»

El artículo de Zadeh ***"Fuzzy sets"*** fue publicado en 1965 en la revista *Information and Control*, que desde 1987 pasó a denominarse *Information and Computation.* Hasta finales de 2022 *"Fuzzy sets"* ha recibido más de 51000 citas en *Web of Science* y cerca de 60000 en *SCOPUS*. Es uno de los treinta trabajos más citados de todos los tiempos y campos.

Zadeh no aminoró el ritmo de su actividad investigadora durante su etapa como director del departamento. Al plantearse en 1964 algunas cuestiones elementales relativas a la falta de nitidez de los límites de clases en los sistemas, introdujo el concepto de conjunto *fuzzy* como una extensión 'flexible' de la noción de conjunto clásico.

Un conjunto clásico (o 'bien definido') de un referencial o universo (también bien definido) puede caracterizarse por alguna propiedad que los elementos del referencial cumplen o no, satisfaciéndose la misma para y solo para los elementos del conjunto. Alternativamente, un conjunto clásico se determina de forma única mediante su función indicador, que asume valor 1 para los elementos del referencial que cumplen la propiedad y valor 0 para los restantes.

En contraste,un ***conjunto fuzzy*** (o *ill-defined*, según Zadeh) de un referencial puede caracterizarse por alguna propiedad que los elementos de este no necesariamente satisfacen o no, sino que su cumplimiento admite gradación. Esta gradación es capital, ya que gracias a ella podrán construirse escalas para medir muchas características que no serían medibles con escalas numéricas. La función indicador de los conjuntos clásicos se extiende a los *fuzzy* a través de la llamada 'función de pertenencia' (aplicación del referencial en el intervalo cerrado unidad $[0, 1]$), que a cada elemento del referencial le asigna el grado de cumplimiento de la propiedad asociada que tiene ese elemento. En otras palabras, el grado de pertenencia de un elemento del referencial al conjunto *fuzzy* puede interpretarse intuitivamente como el 'grado de compatibilidad del elemento con la propiedad que caracteriza al conjunto' o como 'el grado de verdad de la afirmación de que tal elemento la satisface'.

Con los conjuntos *fuzzy*, el paso de la lógica binaria tradicional a la multivaluada que Zadeh había estudiado en Princeton, se extiende hasta la lógica *fuzzy* que amplía los posibles valores a un continuo. En su introducción de los conjuntos *fuzzy*, Zadeh extendió también las operaciones entre conjuntos unión, intersección y complementación, así como la relación de inclusión. Estas extensiones y la estructura que determinan sus propiedades fueron descritas con detalle en el discurso de ingreso en la RAC del Profesor Francisco Azorín Poch. No van a recordarse en el presente discurso, ya que para la investigación en el análisis de datos *fuzzy* que va a exponerse no son necesarias.

La lógica *fuzzy* ha supuesto una herramienta muy importante para la modelización matemática de muchos sistemas complejos y ha servido de fundamento en un buen número de desarrollos de la Inteligencia Artificial. Aunque Zadeh había anticipado como aplicaciones más inmediatas de la lógica *fuzzy* las relativas a los campos de la Educación, la Psicología, la Salud y las Humanidades, sus implicaciones más visibles han sido de índole tecnológica (por ejemplo, control de vehículos, cámaras de vídeo o fotográficas, lavadoras, etc.), aprovechando el hecho real de que razonar sobre la base de conocimientos imprecisos, en forma similar al razonamiento humano, resulta a menudo suficiente y requiere emplear muchas menos reglas y restricciones, lo que redunda en reducción de tiempo y de coste.

Los números *fuzzy*

En 1973, Lotfi Zadeh argumentaba como motivación de la necesidad de los conjuntos *fuzzy* que:

«... a medida que aumenta la complejidad de un sistema, nuestra capacidad para hacer afirmaciones precisas sobre su comportamiento que tengan sentido disminuye hasta alcanzar un umbral más allá del cual la precisión y el sentido (o la relevancia) se vuelven casi mutuamente excluyentes. En este orden de ideas, los análisis cuantitativos precisos son apenas aplicables a los problemas sociales, políticos, económicos y de otro tipo del mundo real que involucran valoraciones humanas.»

Un caso particular de conjuntos *fuzzy* cuyo estudio ha recibido bastante atención es aquel en el que el referencial es un espacio euclídeo. Cuando el referencial es el espacio de los números reales, se denominan números *fuzzy* y revisten un interés práctico especial para muchas calificaciones o evaluaciones humanas.

La noción de número *fuzzy* fue introducida por Zadeh en 1975, en el trabajo *"The Concept of a Linguistic Variable and its Application to Approximate Reasoning-I"*, con el objeto de modelizar cantidades numéricas imprecisas. Para hacer la definición más manejable en diversos contextos, fue refinándose con condiciones adicionales apenas restrictivas.

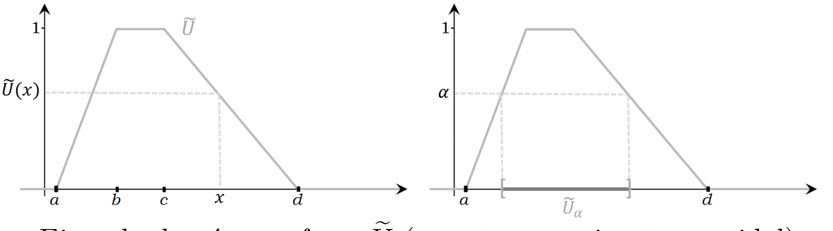
Un ***número fuzzy***, según su PERSPECTIVA VERTICAL, es un valor impreciso formalizado por una aplicación $\widetilde{U} : \mathbb{R} \to [0,1]$ que es semicontinua superiormente, cuasi-cóncava, normal (toma valor 1 para al menos un número real) y con soporte (o conjunto de números reales a los que asocia valor positivo) acotado (en ocasiones esta última condición se obvia). Se interpreta $\widetilde{U}(x)$ como el 'grado de compatibilidad' del número real x con $\widetilde{U}$.

Equivalentemente, un número *fuzzy*, según su PERSPECTIVA HORIZONTAL, es una aplicación $\widetilde{U} : \mathbb{R} \to [0,1]$ tal que para cada $\alpha \in [0,1]$ el conjunto <u>α-nivel</u>, integrado por los números reales con compatibilidad al menos igual a α con $\widetilde{U}$ si $\alpha > 0$ o igual a la clausura topológica del soporte si $\alpha = 0$, es un intervalo compacto no vacío (en ocasiones se obvia la compacidad del 0-nivel).

Un ejemplo de número *fuzzy* para el que son muy sencillas la interpretación, la caracterización, el trazado y la adaptación a los cálculos del análisis de datos que va a revisarse posteriormente, es la de un número *fuzzy* trapezoidal, cuyos 'brazos' son lineales.

Una valoración imprecisa modelizada mediante un número *fuzzy* como el del ejemplo en la gráfica siguiente indicaría que cualquier puntuación en el intervalo $[b, c]$ sería plenamente compatible con tal valoración, y que las puntuaciones en el intervalo

(a, d) serían compatibles con la misma en alguna (mayor o menor) medida.



Ejemplo de número *fuzzy* $\widetilde{U}$ (en este caso, tipo trapezoidal) y de su α nivel, $\widetilde{U}_\alpha$

En el modelizado mediante un número *fuzzy*, la medición se lleva a cabo mediante una escala doblemente continua (vertical, para indicar la 'posición' y horizontal para modular la 'imprecisión'). Además, permite captar diferencias individuales entre quienes establecen las valoraciones, de modo que no se pierda información estadísticamente relevante. Una escala basada en la clase de los números *fuzzy* es más rica y expresiva que cualquier escala basada en el lenguaje natural (inevitablemente finita), en el sentido de que se interpreta y entiende lo que significa cada número *fuzzy* dado, aunque, en general, no pueda etiquetarse ni lingüística ni numéricamente de manera diferencial.

Los números *fuzzy* admiten un buen manejo matemático conceptual y computacional, ya que pueden establecerse distancias y aritmética oportunas al efecto, como se verá más tarde.

Relaciones de Zadeh con España

A lo largo de su vida, el Profesor Lotfi A. Zadeh recibió muchos reconocimientos y galardones. Fue miembro, entre otras, de la Academia Nacional de Ingeniería de EEUU y miembro extranjero de diversas Academias de Ciencias y de Ciencias y Tecnología.

Entre los reconocimientos más frecuentes se encuentra un elevado número de Doctorados *Honoris Causa*. Solo en España, fue investido como tal por las Universidades de Oviedo (en 1995), Granada (en 1996) y la Politécnica de Madrid (en 2007).

Fue mentor de gran número de investigadores españoles (postdoctorales, asociados, etc.) que disfrutaron de estancias como parte de su grupo en la Universidad de California en Berkeley.

Arriba: Doctorado *Honoris Causa* de Lotfi A. Zadeh
por la Universidad de Oviedo en 1995
Debajo: Discurso del Premio Fronteras del Conocimiento
en 2012 en TIC (Fundación BBVA)

Su presencia en congresos internacionales celebrados en ciudades de España fue casi permanente y más aún cuando se creó, a iniciativa del propio Zadeh, el *Centro Europeo de Soft Computing* en Mieres, que cerró sus puertas en 2016 tras once años de trabajo intenso. Como Presidente de su Comité Científico, sus viajes a Asturias fueron asiduos en aquella época.

En el año 2012, recibió el *Premio Fronteras del Conocimiento* de la Fundación BBVA en Tecnologías de la Información y la Comunicación.

Los elementos aleatorios generadores de datos *fuzzy*

La naturaleza, la ciencia y la tecnología
ofrecen numerosos ejemplos de elementos aleatorios
que no son ni números, ni series, ni vectores, ni funciones...

... Podemos también encontrar elementos
que aún no se han descrito matemáticamente

(Maurice R. Fréchet)

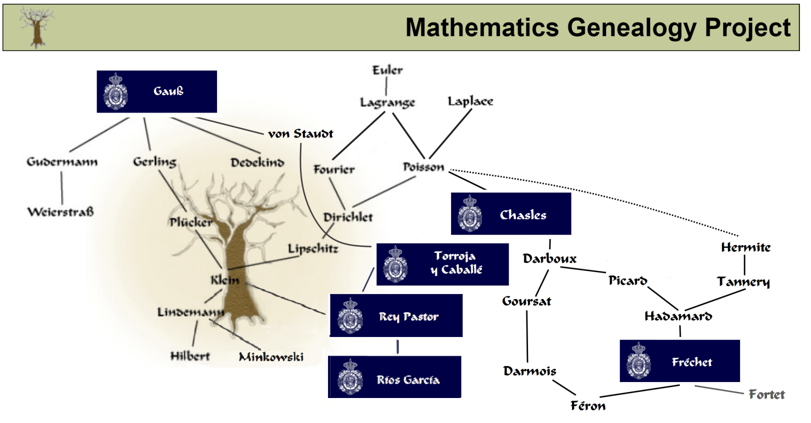
La teoría de la probabilidad y la lógica fuzzy
son complementarias en lugar de competidoras

(Lotfi A. Zadeh)

Por razones temporales, es muy posible que Lotfi A. Zadeh y Maurice Fréchet no tuvieran la oportunidad de discutir si los conjuntos *fuzzy* podían ser o no valores potenciales de elementos aleatorios abstractos. No hay constancia de que tal discusión se mantuviera, pero sí de algunos encuentros entre Zadeh y un 'doctorando' de Fréchet (y de Darmois): Robert L. Féron, al que nos referiremos en breve.

Antes de cumplirse la primera década tras la introducción de los conjuntos *fuzzy*, varios investigadores provenientes de diversos campos científicos y países se fueron interesando por la nueva teoría. Si bien en Francia se considera que la primera publicación de trabajo en revista que involucraba conjuntos *fuzzy* apareció en 1972 firmada por el matemático Edwin Diday (pionero y experto reconocido en el análisis de datos simbólicos), fue Arnold Kaufmann (profesor de Investigación Operativa y Mecánica Aplicada en la Escuela Nacional Superior de Minas de París) quien realizó los mayores esfuerzos para popularizar las nociones básicas de la nueva teoría. En 1973 publicó la primera monografía existente sobre el tema, que fue traducida muy pronto y a la que siguieron otros volúmenes que cubrían operaciones y relaciones entre tales conjuntos, así como algunas de las aplicaciones iniciales a varios campos: reconocimiento de patrones, autómatas y sistemas, decisión multicriterio, lingüística, lógica, etc.

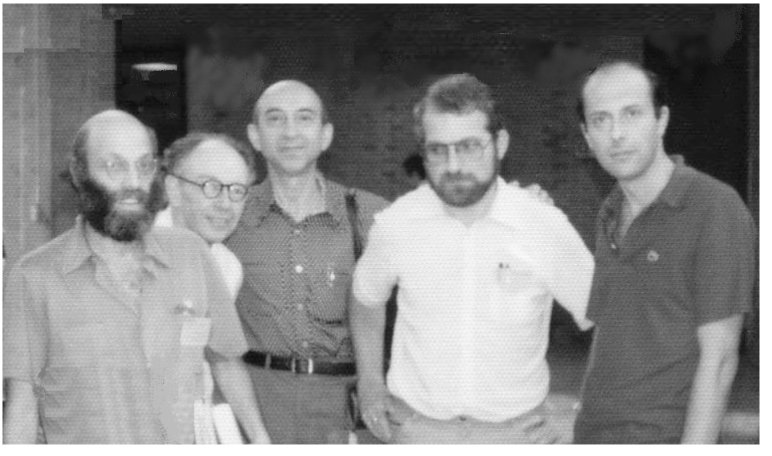
El impacto divulgativo de los libros de Kaufmann en la comunidad investigadora francesa fue visible en poco tiempo. Entre los 'afectados' y en relación directa con este discurso, es obligado destacar a **Robert L. Féron**. Féron fue un estadístico que se doctoró en Matemáticas en 1954 en la Facultad de Ciencias de la Sorbona con su tesis sobre información, regresión y correlación, supervisada por Maurice Fréchet y Georges Darmois, mientras disfrutaba de una estancia de varios años en el CNRS. En 1958 fue contratado como *maître de conférences* en la Facultad de Ciencias de la Universidad de Lyon, donde promocionó a *professeur* hasta su jubilación.



Árbol genealógico científico del Profesor Robert L. Féron
Fuente: *Mathematics Genealogy Project*

Además de los trabajos derivados de su tesis doctoral, Féron y Fréchet investigaron sobre el problema de determinar la relación entre una distribución conjunta y sus distribuciones marginales, que estaba ligado a sus estudios acerca de la correlación. Hacia mediados de los setenta, una vez aparecido el primer volumen de Kaufmann sobre conjuntos *fuzzy*, Féron empezó a interesarse sobre la teoría de los elementos aleatorios generalizados, introducida por Maurice Fréchet, considerando que el espacio de llegada podía ser el de los conjuntos *fuzzy* sobre cierto referencial.

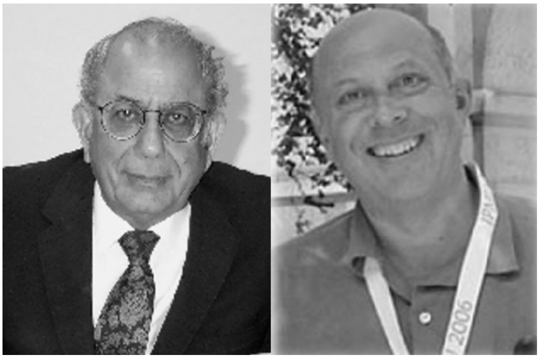
De este modo, en varios de sus trabajos publicados entre 1976 y 1979, estableció el concepto de ***conjunto fuzzy aleatorio***. Su definición era un tanto general, indicando esencialmente que se trataba de una función medible. Y las métricas consideradas sobre el espacio de conjuntos *fuzzy* correspondían a promedios en 'sentido vertical' (es decir, expresaban la métrica como ciertas distancias entre las funciones que determinaban los conjuntos *fuzzy* considerados). Publicó, así, varios trabajos sobre el tema en *Comptes Rendus de L'Académie des Sciences de Paris*, transmitidos por Maurice Fréchet o por Robert Fortet (miembro de número y miembro correspondiente, respectivamente, de dicha Academia) o en las *Publications Econométriques*, revista francesa que Féron fundó y editó entre 1967 y 1986.



De izquierda a derecha: Ronald R. Yager, Robert L. Féron, Lotfi A. Zadeh, Erich P. Klement y Dan A. Ralescu (*Int. Cong. Applied Systems Research and Cybernetics*, 1980)

Los conjuntos *fuzzy* aleatorios *à la* Puri y Ralescu

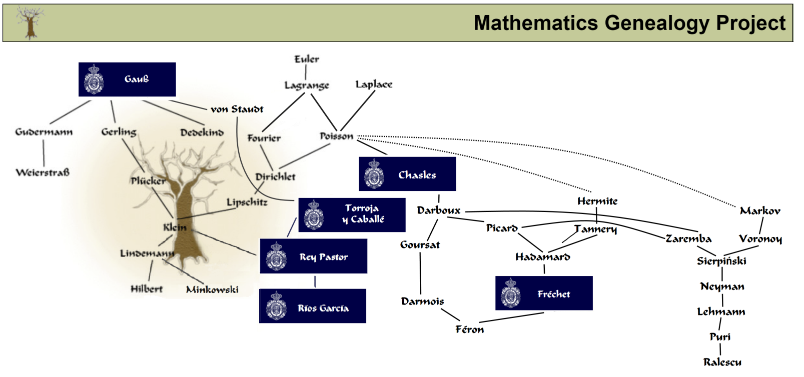
Si bien las primeras ideas para los mecanismos aleatorios que generan datos *fuzzy* provinieron de Féron, el modelo que más se ha utilizado, que está establecido siguiendo todas las pautas de Fréchet y que cuenta con todo el rigor y todos los 'beneplácitos' del marco probabilístico, es el introducido por Madan L. Puri y Dan A. Ralescu a mediados de los ochenta.



Madan L. Puri (izquierda) y Dan A. Ralescu (derecha)

Tras graduarse, obtener el máster y ser nombrado *teaching assistant* en matemáticas en la Universidad de Punjab, en India, en 1957 **Madan L. Puri** se trasladó a EEUU a la Universidad de Colorado en Boulder. Después de un año allí, se mudó a la Universidad de California en Berkeley, donde fue *research assistant* en estadística y se doctoró en 1962 con una tesis dirigida por el Profesor Erich L. Lehmann sobre la eficiencia asintótica de ciertos contrastes de hipótesis. Más tarde fue contratado como *assistant professor* en el Instituto Courant de Ciencias Matemáticas de Nueva York y pasó a ser *professor* en el Departamento de Matemáticas de la Universidad de Indiana en Bloomington en 1968. Su investigación ha sido extensa y muy diversa, mostrando siempre un gran interés por los aspectos más teóricos, aunque sin perder de vista sus implicaciones prácticas. Madan L. Puri ha recibido muchos reconocimientos de Universidades, Fundaciones y Academias Científicas, así como de Sociedades Estadísticas Internacionales.

Dan A. Ralescu completó un máster en Matemáticas en 1972 en la Universidad de Bucarest, Rumanía. Entre 1972 y 1976 ejerció como investigador en el Instituto para la Gestión de Sistemas de Información de Bucarest, donde se familiarizó con la lógica *fuzzy*; en el curso 1976–1977 fue asistente de docencia en la Universidad de Dijon, Francia. En 1979 realizó un segundo máster en Matemáticas y Estadística en Bloomington. En esta misma institución obtuvo el grado de doctor en Estadística en 1980, con una tesis dirigida por Puri sobre la admisibilidad de estimadores en algunos problemas estadísticos. A partir de la defensa de su tesis, se unió a la Universidad de Cincinnati, en Ohio, como *assistant professor*, promocionándose después a *associate professor* y, desde 1985, a *professor*. A lo largo de los años ha realizado diversas estancias como profesor visitante en distintos centros de EEUU, Asia y Europa (entre ellos, la Universidad de Oviedo).



Árbol genealógico científico de los
Profesores Madan L. Puri y Dan A. Ralescu
Fuente: *Mathematics Genealogy Project*

En una revisión de la *Web of Science* pueden encontrarse once trabajos de los que Puri y Ralescu son coautores. Esos trabajos no guardan relación con la tesis doctoral del último. Las revistas que recogen estos artículos se encuentran entre las más prestigiosas en Probabilidad y Matemáticas (*Annals of Probability*, *Journal of Mathematical Analysis and Applications*, *Comptes Rendus de l'Académie des Sciences*, ...).

De los once, cuatro se refieren a resultados límite y de distribución de conjuntos aleatorios; los siete restantes, a funciones con valores *fuzzy* y, especialmente, a los conjuntos *fuzzy* aleatorios. Se publicaron entre 1981 y 1991.

Según la *Web of Science*, a finales de 2022 el número medio de citas recibidas por cada uno de esos once trabajos asciende a cerca de 250, siendo el titulado ***"Fuzzy random variables"*** y publicado en 1986 en el *Journal of Mathematical Analysis and Applications* el que destaca con cerca de 1400 citas. Conviene subrayar que, si bien Puri y Ralescu acuñaron los elementos aleatorios con valores de conjunto *fuzzy* como "variables aleatorias *fuzzy*", parece más idóneo, por coherencia con la noción de conjunto aleatorio, emplear el término que usó Ferón: conjuntos *fuzzy* aleatorios. Y así lo hacemos, a partir de este punto.

Los contenidos del trabajo *"Fuzzy random variables"*, en el que se introducen buena parte de los conceptos y resultados cruciales de los modelos en los que se apoya el análisis de datos *fuzzy* que va a exponerse sucintamente, se describen aquí en el caso unidimensional (es decir, con valores de número *fuzzy*) por dos razones: para facilitar la presentación y la interpretación de las ideas fundamentales y por tratarse del tipo de datos *fuzzy* más natural y frecuente en las situaciones del mundo real. No obstante, su extensión al caso de valores de conjunto *fuzzy* con referencial en un espacio euclídeo de dimensión finita e incluso con estructura más general, es habitualmente inmediata.

Siguiendo la línea de trabajo planteada por Fréchet y particularizada por Féron, Puri y Ralescu introdujeron de forma más concreta los conjuntos *fuzzy* aleatorios como una generalización de los conjuntos aleatorios (y, por lo tanto, de las variables y los vectores aleatorios).

En el caso unidimensional, dado un espacio de probabilidad $(\Omega, \mathcal{A}, P)$, si se denota por $\mathscr{F}(\mathbb{R})$ el conjunto de los números *fuzzy*, Puri y Ralescu definieron un ***conjunto fuzzy aleatorio*** como una aplicación $\mathcal{X} : \Omega \to \mathscr{F}(\mathbb{R})$ que nivel a nivel fuera un intervalo aleatorio compacto (es decir, que para cualquier $\alpha \in [0,1]$, las aplicaciones $\mathcal{X}_\alpha : \Omega \to \mathscr{K}(\mathbb{R})$, con $\mathcal{X}_\alpha(\omega) = \{t \in \mathbb{R} : \mathcal{X}(\omega)(t) \geq \alpha\}$, fueran funciones medibles Borel con respecto de la σ-álgebra de Borel generada por la topología inducida por la métrica de Hausdorff d_H sobre el espacio de los intervalos compactos no vacíos, $\mathscr{K}(\mathbb{R})$).

Esta definición no exigía, en principio, una medibilidad Borel global para el conjunto *fuzzy* aleatorio. Es decir, no se acreditaba que se tratara de un elemento aleatorio *á la* Fréchet. En dos trabajos (uno en 1985 y otro junto con Erich P. Klement en 1986) Puri y Ralescu reemplazaron la condición de medibilidad Borel nivel a nivel por la medibilidad Borel (global) con respecto de la métrica del supremo en $\mathscr{F}(\mathbb{R})$, $d_\infty(\widetilde{U}, \widetilde{V}) = \sup_{\alpha\in[0,1]} d_H(\widetilde{U}_\alpha, \widetilde{V}_\alpha)$. Pero ni se satisfacía la separabilidad del espacio métrico asociado ni las definiciones eran equivalentes.

Estos inconvenientes quedaron soslayados años después por miembros del Grupo SMIRE/SMIRE⚡CODIRE, en ocasiones colaborando con Ralescu, demostrando que:

- ciertas métricas operativas tipo $\mathscr{L}^2$, definidas por ellos y fáciles de interpretar intuitivamente, eran topológicamente equivalentes a las establecidas anteriormente por Klement, Puri y Ralescu, dando lugar a espacios métricos separables,
- y que la medibilidad con respecto de la σ-álgebra de Borel generada por las topologías inducidas por esas nuevas métricas sí eran equivalentes a la condición supuesta en la definición original nivel a nivel de Puri y Ralescu.

Research Group
STATISTICAL METHODS WITH IMPRECISE RANDOM ELEMENTS

SMIRE

Research Group
STATISTICAL METHODS WITH IMPRECISE RANDOM ELEMENTS ⚡ COMPARISON OF DISTRIBUTIONS OF RANDOM ELEMENTS

SMIRE ⚡ CODIRE

Coordination: C/ Federico García Lorca, nº 18, 33007 Oviedo, SPAIN

Grupos de Investigación SMIRE/SMIRE⚡CODIRE
de la Universidad de Oviedo. Todos sus miembros son
descendientes científicos del Profesor Sixto Ríos García

Como consecuencia principal de la medibilidad Borel respecto de tales métricas en $\mathscr{F}(\mathbb{R})$, muchos de los conceptos probabilísticos básicos relativos a los conjuntos *fuzzy* aleatorios se inducen de forma inmediata: la *distribución de un conjunto fuzzy aleatorio*, la *independencia entre conjuntos fuzzy aleatorios*, etc.

Un encaje isométrico para el espacio de números *fuzzy*

Mientras que la noción de distribución (inducida) asociada a un conjunto *fuzzy* aleatorio se define de forma directa por su medibilidad Borel, los valores resumen más relevantes de esa distribución no se obtendrían directamente. E igual ocurriría con la aritmética elemental (más concretamente, la suma y el producto por un escalar real) entre números *fuzzy*.

Sin embargo, gracias a la mediación de un resultado de apoyo, valores resumen como la media, la varianza y la covarianza de conjuntos *fuzzy* aleatorios y la aritmética de números *fuzzy* pueden establecerse sin dificultades ni herramientas externas: sencillamente, lo harán a partir de las mismas medidas resumen y aritmética para elementos aleatorios con valores funcionales.

En este punto, sería razonable formular la siguiente cuestión:

Puesto que los números *fuzzy* son funciones reales de variable real, ¿podrían tratarse los datos con valores de número *fuzzy* como un caso especial de datos funcionales en el contexto del análisis de datos?

La respuesta a esta pregunta va a tener doble sentido:

- Directamente NO, ya que si se aplica la aritmética funcional sobre números *fuzzy*, habitualmente la función resultante ni siquiera corresponde a un número *fuzzy* y, en las escasas situaciones en que así fuera, se perderían el significado y la interpretación de tales operaciones;
- Indirectamente, SÍ. Se puede probar la existencia de una isometría sobre el espacio de los números *fuzzy* que hace corresponder cada número *fuzzy* con una función de cierto espacio de Hilbert separable y las métricas entre los números *fuzzy* coincidirán con las métricas entre las imágenes funcionales correspondientes, deduciéndose como valor adicional que la suma y producto por escalar real con números *fuzzy* es la introducida en 1975 por Lotfi Zadeh.

La idea directriz del resultado de apoyo, base de la respuesta afirmativa a la cuestión planteada, aparecía en el trabajo de Klement, Puri y Ralescu (1986) con la métrica del supremo en $\mathscr{F}(\mathbb{R})$. El resultado sigue siendo válido, tras la adaptación adecuada, para las métricas señaladas para las que la medibilidad Borel para conjuntos *fuzzy* aleatorios equivale a la dada nivel a nivel por Puri y Ralescu. Por ejemplo, si se considera la métrica generalizada D_τ^φ de SMIRE/SMIRE↯CODIRE entre valores en $\mathscr{F}(\mathbb{R})$ tal que $D_\tau^\varphi(\widetilde{U}, \widetilde{V})$ viene dada por

$$\sqrt{\int_{[0,1]} \left[(\operatorname{mid} \widetilde{U}_\alpha - \operatorname{mid} \widetilde{V}_\alpha)^2 + \tau(\operatorname{spr} \widetilde{U}_\alpha - \operatorname{spr} \widetilde{V}_\alpha)^2\right] \, d\,\varphi(\alpha)}$$

(con mid = punto medio/centro, spr = semiamplitud/radio, peso $\tau > 0$ y φ = medida de ponderación normalizada asociada con distribución continua de soporte $[0, 1]$), se trata de una métrica tipo $\mathscr{L}^2$, invariante por traslaciones y por giros a ambos lados y que, como se apuntó antes, determina sobre $\mathscr{F}(\mathbb{R})$ un espacio métrico separable.

La esencia del resultado de apoyo reside en la función soporte de un número *fuzzy*, definida para conjuntos *fuzzy* por Puri y Ralescu en su trabajo de 1985 como una extensión de la función soporte de un conjunto cerrado, convexo y no vacío introducida en 1903 por Minkowski.

En el caso de números *fuzzy*, la *función soporte* sería $s : \mathscr{F}(\mathbb{R}) \to \mathbb{H}_2$, siendo este último el espacio de Hilbert de funciones cuadrado integrable

$$\mathbb{H}_2 = \{f : [0,1] \times \{-1,1\} \to \mathbb{R} \text{ con } f^2 \text{ integrable}$$
$$\text{respecto de } \mathrm{Unif}([0,1]) \otimes \mathrm{Unif}(\{-1,1\})\}$$

y $s(\widetilde{U}) = s_{\widetilde{U}}$, donde $s_{\widetilde{U}}(\alpha, u) = \sup_{v \in \widetilde{U}_\alpha} \langle u, v \rangle = \sup_{v \in \widetilde{U}_\alpha} u \cdot v$, es decir, $s_{\widetilde{U}}(\alpha, -1) = -\inf \widetilde{U}_\alpha$, $s_{\widetilde{U}}(\alpha, 1) = \sup \widetilde{U}_\alpha$.

Se cumple que $s_{\widetilde{U}}$ caracteriza a $\widetilde{U}$ y que la función soporte establece un encaje isométrico del espacio métrico $(\mathscr{F}(\mathbb{R}), D_\tau^\varphi)$ en un <u>cono convexo del espacio de Hilbert</u> $(\mathbb{H}_2, \|\cdot\|_\tau^\varphi)$ con la norma $\mathscr{L}^2$ dada por $\|h_1 - h_2\|_\tau^\varphi = \sqrt{\langle h_1 - h_2, h_1 - h_2 \rangle_\tau^\varphi}$, donde el producto interior (o escalar) es

$$\langle f, g \rangle_\tau^\varphi = \sum_{u \in \{-1,1\}} \int_{[0,1]} \frac{\mathrm{mid}\, f(\alpha, u) \cdot \mathrm{mid}\, g(\alpha, u)}{2}\, d\varphi(\alpha)$$
$$+ \tau \sum_{u \in \{-1,1\}} \int_{[0,1]} \frac{\mathrm{spr}\, f(\alpha, u) \cdot \mathrm{spr}\, g(\alpha, u)}{2}\, d\varphi(\alpha)$$

con

$$\mathrm{mid}\, f(\alpha, u) = \frac{f(\alpha, u) - f(\alpha, -u)}{2},$$
$$\mathrm{spr}\, f(\alpha, u) = \frac{f(\alpha, u) + f(\alpha, -u)}{2}.$$

La isometría garantiza que para $\widetilde{U}, \widetilde{V} \in \mathscr{F}(\mathbb{R})$ se tiene que $D_\tau^\varphi(\widetilde{U}, \widetilde{V}) = \|s_{\widetilde{U}} - s_{\widetilde{V}}\|_\tau^\varphi$.

Sobre la base del encaje isométrico al que acaba de aludirse, los <u>datos de número *fuzzy*</u> pueden identificarse con <u>datos funcionales</u> dentro de un espacio de Hilbert separable de funciones. Pero la identificación no es bidireccional, ya que la imagen por la función soporte no abarca todo el espacio de Hilbert, sino solo un cono convexo dentro de él (es decir, un subconjunto cerrado para todas las combinaciones lineales no negativas).

Como las definiciones de conjunto *fuzzy* aleatorio, esa identificación es extensible a conjuntos *fuzzy* de espacios euclídeos de dimensión superior. Puede señalarse también que, si se eliminara la condición de compacidad del 0-nivel en la definición de los valores *fuzzy*, la imagen del espacio de valores *fuzzy* por la función soporte sería un cono convexo y cerrado.

OBSERVACIÓN: Merece enfatizarse que un conjunto *fuzzy* aleatorio $\mathcal{X}$ podría haberse definido equivalentemente como aquel tal que $s_{\mathcal{X}}$ fuera un elemento aleatorio con valores en el cono convexo $s(\mathscr{F}(\mathbb{R}))$ del espacio de Hilbert de funciones $\mathbb{H}_2$, con la medibilidad Borel basada en la norma $\|\cdot\|_{\tau}^{\varphi}$. En este discurso se ha seguido el orden cronológico según el que se introdujo el concepto, pero se podría haber empezado por el encaje isométrico y deducir las definiciones iniciales como consecuencia.

Implicaciones cruciales del encaje isométrico: aritmética de números *fuzzy* y medidas resumen de conjuntos *fuzzy* aleatorios

La trascendencia del encaje isométrico es múltiple. Por un lado, Zadeh había definido una aritmética entre conjuntos *fuzzy* basada en el principio de extensión propuesto por él (1975). Por otro lado, Puri y Ralescu presentaron en 1986 la media tipo Aumann e investigadores de SMIRE/SMIRE⚡CODIRE formularon en 2000 una varianza tipo Fréchet. Sin embargo, esas definiciones no tienen por qué introducirse expresamente para el caso *fuzzy*; tanto la aritmética propuesta por Zadeh como las medidas resumen antedichas podrían haberse deducido a través de la función soporte a partir de las definiciones correspondientes a la aritmética funcional y a las medidas resumen para elementos aleatorios con valores funcionales. Además, podría también deducirse la noción de covarianza entre dos conjuntos *fuzzy* aleatorios.

De este modo:

- ▶ <u>Aritmética entre números fuzzy:</u> si dados dos números *fuzzy* $\widetilde{U}, \widetilde{V} \in \mathscr{F}(\mathbb{R})$ y un escalar real $\lambda \in \mathbb{R}$ se define la <u>*suma*</u> de los dos primeros como el número *fuzzy* $\widetilde{U} + \widetilde{V}$ tal que
$$s_{\widetilde{U}+\widetilde{V}} = s_{\widetilde{U}} + s_{\widetilde{V}}$$
y el <u>*producto por el escalar*</u> como el número *fuzzy* tal que
$$s_{\lambda\cdot\widetilde{U}} = \lambda \cdot s_{\widetilde{U}},$$
se llega a que ambas operaciones equivalen a las definidas por Zadeh:
$$(\widetilde{U} + \widetilde{V})(t) = \sup_{(y,z)\,:\,y+z=t} \min\left\{\widetilde{U}(y), \widetilde{V}(z)\right\}$$

$$(\lambda \cdot \widetilde{U})(t) = \sup_{y\,:\,\lambda y = t} \widetilde{U}(y) = \begin{cases} \widetilde{U}\left(\frac{t}{\lambda}\right) & \text{si } \lambda \neq 0, \\ \mathbb{1}_{\{0\}}(t) & \text{si } \lambda = 0. \end{cases}$$

Según Hung T. Nguyen (1978), doctorando de Joseph Kampé de Fériet, académico histórico extranjero de esta Academia, esta aritmética entre números *fuzzy* coincide a su vez nivel a nivel con la aritmética de conjuntos usual, es decir, que para todo $\alpha \in [0,1]$:

$$\begin{aligned}(\widetilde{U} + \widetilde{V})_\alpha &= \text{suma de Minkowski de } \widetilde{U}_\alpha \text{ y } \widetilde{V}_\alpha \\ &= \{x + y \,:\, x \in \widetilde{U}_\alpha,\, y \in \widetilde{V}_\alpha\}, \\ (\lambda \cdot \widetilde{U})_\alpha &= \lambda \cdot \widetilde{U}_\alpha = \{\lambda \cdot x \,:\, x \in \widetilde{U}_\alpha\}.\end{aligned}$$

- Media de un conjunto fuzzy aleatorio: si dado un conjunto *fuzzy* aleatorio $\mathcal{X}$ asociado al espacio de probabilidad $(\Omega, \mathcal{A}, P)$, se define la *media de* $\mathcal{X}$ como el número *fuzzy*, si existe, $\widetilde{E}(\mathcal{X})$ tal que
$$s_{\widetilde{E}(\mathcal{X})} = E(s_{\mathcal{X}}),$$
donde $E(s_{\mathcal{X}})$ representa la integral de Bochner del elemento aleatorio $s_{\mathcal{X}}$ con valores en el espacio de Hilbert $\mathbb{H}_2$, entonces, esa noción equivale a la media tipo Aumann introducida por Puri y Ralescu (1986), que es el único número *fuzzy* $\widetilde{E}(\mathcal{X}) \in \mathscr{F}(\mathbb{R})$ tal que para cada $\alpha \in [0,1]$:
$$(\widetilde{E}(\mathcal{X}))_\alpha = E^A(\mathcal{X}_\alpha) = \{E(X) \,:\, X \overset{c.s.[P]}{\in} \mathcal{X}_\alpha\},$$
es decir, el α-nivel de $\widetilde{E}(\mathcal{X})$ es la integral de Aumann (llamada también de Kudō-Aumann) del conjunto aleatorio $\mathcal{X}_\alpha$, dada por las medias de todas las selecciones X de ese conjunto aleatorio.
$\widetilde{E}(\mathcal{X})$ podría haberse definido también como el valor medio de Fréchet (i.e., el que minimiza la dispersión cuadrática) para D_τ^φ y otras métricas $\mathscr{L}^2$.
El operador $\widetilde{E}(\cdot)$ conserva las principales propiedades de la media de variables aleatorias (equivarianza por transformaciones lineales, aditividad, etc.), es coherente con la aritmética de Zadeh para conjuntos *fuzzy* aleatorios 'discretos' y satisface Leyes Fuertes de los Grandes Números para prácticamente todas las métricas que pueden establecerse sobre $\mathscr{F}(\mathbb{R})$.

- <u>VARIANZA *À LA* FRÉCHET DE UN CONJUNTO FUZZY ALEATORIO:</u> prefijados τ y φ, si dado un conjunto *fuzzy* aleatorio $\mathcal{X}$ asociado al espacio de probabilidad $(\Omega, \mathcal{A}, P)$, se define la (τ, φ)*-varianza à la Fréchet de* $\mathcal{X}$ como el número real, si existe, dado por

  $$\sigma^2_{\tau,\varphi}(\mathcal{X}) = E\left([\|s_{\mathcal{X}} - E(s_{\mathcal{X}})\|^{\varphi}_{\tau}]^2\right),$$

  entonces, en el caso de conjunto aleatorio con valores de número *fuzzy*, la isometría asegura que equivale a la varianza definida por SMIRE/SMIRE↯CODIRE (2000) como el número real dado por

  $$\sigma^2_{\tau,\varphi}(\mathcal{X}) = E\Big([D^{\varphi}_{\tau}(\mathcal{X}, \widetilde{E}(\mathcal{X}))]^2\Big).$$

  El operador $\sigma^2_{\tau,\varphi}(\cdot)$ conserva las principales propiedades de la varianza de variables aleatorias (anulación si y solo si concierne a conjuntos *fuzzy* aleatorios con distribución degenerada, invarianza por traslaciones, aditividad bajo independencia, etc.).

- <u>COVARIANZA DE DOS CONJUNTOS FUZZY ALEATORIOS:</u> prefijados τ y φ, si $\mathcal{X}$ e $\mathcal{Y}$ son dos conjuntos *fuzzy* aleatorios asociados al espacio de probabilidad $(\Omega, \mathcal{A}, P)$, se define la (τ, φ)*-covarianza de* $\mathcal{X}$ *e* $\mathcal{Y}$ como el número real, si existe, dado por

  $$\sigma_{\tau,\varphi}(\mathcal{X}, \mathcal{Y}) = E\left(\langle s_{\mathcal{X}} - E(s_{\mathcal{X}}), s_{\mathcal{Y}} - E(s_{\mathcal{Y}})\rangle^{\varphi}_{\tau}\right).$$

  La existencia del producto interior en el espacio de Hilbert hace posible que esta definición tenga sentido.

  El operador $\sigma_{\tau,\varphi}(\cdot,\cdot)$ conserva varias de las principales propiedades de la covarianza de variables aleatorias (por ejemplo, la anulación para conjuntos *fuzzy* aleatorios independientes), pero no todas.

Una metodología para el análisis de datos *fuzzy*

... Estamos a la espera de paquetes de programas
fácilmente adaptables, que permitan a los investigadores
usar los datos fuzzy y, después,
aplicar a esos datos técnicas estadísticas oportunas
y otros análisis con el fin de contrastar hipótesis
y asegurar que el significado se ha captado apropiadamente

(Beryl Hesketh y colaboradores)

Siguiendo las ideas del análisis de datos numéricos, el objetivo del análisis de datos *fuzzy* es:

- **resumir** la información contenida en los mismos, mediante ciertas medidas basadas en la muestra de datos disponibles,
- y **extraer/extrapolar conclusiones** acerca de cómo se distribuye el conjunto *fuzzy* aleatorio que genera tales datos, sobre la base de la información resumida.

De las implicaciones derivadas del encaje isométrico del espacio de datos *fuzzy* potenciales en un cono convexo de un espacio de Hilbert separable, se concluye que pueden conservarse/adaptarse la mayoría de la ideas y nociones del análisis estadístico de datos numéricos, así como extenderse/desarrollarse procedimientos del análisis de datos numéricos para analizar datos *fuzzy*.

Como la distribución de probabilidad y la independencia de conjuntos *fuzzy* aleatorios se inducen de forma inmediata, sin necesidad de introducir expresamente su definición, también cobran sentido directo herramientas matemáticas como el concepto de muestra aleatoria simple.

Por la misma razón, en el problema de estimación 'paramétrica' (entendiendo por 'parámetros' las medidas resumen poblacionales), conceptos como el de estimador insesgado siguen siendo aplicables, tanto si el parámetro es numérico como si toma valores *fuzzy*. Nótese que el de estimador asintóticamente insesgado también lo es, si bien en el caso de parámetros *fuzzy* habría que involucrar alguna métrica para formalizar el límite correspondiente de la sucesión de estimaciones *fuzzy*. Otra propiedad de los estimadores fácilmente extensible sería la de consistencia (débil y fuerte), teniendo en cuenta la participación de alguna métrica cuando el parámetro a estimar sea *fuzzy*.

De modo análogo, en el contraste de hipótesis estadísticas sobre la distribución (o aspectos de la misma) de un conjunto *fuzzy* aleatorio, conceptos como el p-valor, tamaño, potencia del test, etc. se conservan de forma evidente.

A pesar de que, conceptualmente, la extensión de la mayoría de las nociones del análisis clásico de datos reales (o vectoriales) al de datos *fuzzy* no conlleva apenas dificultades, la extensión de muchos métodos del análisis de datos numéricos se enfrenta a varios inconvenientes que no pueden salvarse trivialmente. Así, en comparación con el análisis de datos numéricos, hay que destacar que:

- ▼ la aritmética en el espacio de valores *fuzzy* no determina sobre el mismo una estructura lineal, sino cónica; de hecho, como ocurre en el caso intervalar o de valores de conjunto clásico, no existe un ‘operador diferencia’ que respete todas las propiedades de la diferencia de valores reales y esté bien definida para cada par de valores;
- ▼ no existe un orden total universalmente aceptable entre valores *fuzzy*;
- ▼ no se han establecido, aún, modelos realistas y suficientemente generales que sean apropiados para las distribuciones de conjuntos *fuzzy* aleatorios;
- ▼ debe tenerse precaución en asegurar que los resultados que se empleen para trabajar con datos *fuzzy* no se muevan fuera del cono convexo $s(\mathscr{F}(\mathbb{R}))$; por ejemplo, aunque existen algunos teoremas del límite central para conjuntos *fuzzy* aleatorios, el elemento aleatorio límite es un gaussiano, pero rara vez es un conjunto *fuzzy* aleatorio y no pueden aplicarse directamente con fines inferenciales.

Herramientas y directrices habituales de la metodología del análisis de datos *fuzzy*

Los métodos desarrollados hasta el momento para el análisis de datos *fuzzy* se basan habitualmente:

- en la particularización de procedimientos del análisis de datos funcionales, siempre que en el proceso no haya que desplazarse fuera del cono convexo $s(\mathscr{F}(\mathbb{R}))$;
- en técnicas *ad hoc*, con frecuencia fundamentadas en la consideración de poblaciones finitas o en la teoría de grandes muestras junto con aproximaciones *bootstrap* de las mismas;
- o en aproximaciones mediante estudios de simulación, especialmente cuando no es posible deducir resultados generales, pero sí conclusiones mayoritarias.

■ En los MÉTODOS OBTENIDOS POR PARTICULARIZACIÓN DE LOS DEL ANÁLISIS DE DATOS FUNCIONALES:

La idea directriz es pasar a través de la función soporte s desde el espacio de los valores *fuzzy* $\mathscr{F}(\mathbb{R})$ al cono convexo de funciones cuadrado integrables $s(\mathscr{F}(\mathbb{R}))$, y aplicar a los datos *fuzzy* transformados por s la técnica funcional. El resultado del proceso será válido si ninguno de los pasos intermedios o final transcurre en $\mathbb{H}_2 \setminus s(\mathscr{F}(\mathbb{R}))$.

Como ejemplos de esta particularización se encuentran, por un lado, los procedimientos cuya aplicación posterior no reviste mayor complejidad, aunque muy posiblemente exigen condiciones de partida que deben verificarse meticulosamente, pero es sencillo comprobar que las soluciones forman parte del cono convexo (esto ocurre, entre otras, con la media tipo Aumann o con varias de las estimaciones robustas de localización).

Por otro lado, hay procedimientos cuya particularización no podría garantizarse que discurriera invariablemente dentro del cono convexo. Así ocurriría si, por ejemplo, se quisieran realizar contrastes sobre la igualdad de la media tipo Aumann de uno, dos o varios conjuntos *fuzzy* aleatorios. La carencia de modelos para las distribuciones de los conjuntos *fuzzy* aleatorios y de un teorema del límite central en el que el proceso gaussiano límite sea un conjunto *fuzzy* aleatorio, dificulta *a priori* el desarrollo de métodos para tales contrastes. Sin embargo, la aplicación de una aproximación *bootstrap* del teorema del límite central para elementos aleatorios con valores en espacios de Hilbert separables va a permitir dar una solución muy pertinente.

A finales de los setenta, **Bradley Efron** desarrolló el *bootstrap*. Se trata de un método computacionalmente intensivo para aproximar la distribución muestral de un estadístico por remuestreo y sin asumir hipótesis sobre la generación de los datos. Un estadístico es una función medible de la muestra aleatoria simple (aquella a partir de la que se genera la muestra de datos) y sus valores representan una contrapartida muestral de algún parámetro o rasgo de la distribución del elemento aleatorio en la población. En el *bootstrap*, el remuestreo se lleva a cabo de modo que las muestras de observaciones no se extraen de la población, sino de la muestra disponible de partida.

En la búsqueda de una manera de determinar la idoneidad de un resultado inferencial sin recurrir a repetir la ejecución experimental y la medición vinculada, Efron concibió la idea consistente en tomar datos aleatoriamente de la única muestra disponible y analizarlos; repetir ese mismo proceso un gran número de veces y aproximar el margen de errores a partir de ese remuestreo aleatorio reiterado. En otras palabras, se imita la variación asociada a remuestrear a partir de una población mediante la variación asociada a remuestrear a partir de la muestra original. Para su aplicación es clave el uso de ordenadores, ya que cuanto mayor sea el número de muestras generadas mejor podrá afinarse la precisión del resultado.

Saludo entre Sir David R. Cox (izquierda) y Bradley Efron (derecha) tras el discurso del Premio Fronteras del Conocimiento (recibido conjuntamente) en 2016 en Ciencias Básicas (Fundación BBVA)

En cuanto al término *bootstrap*, Efron se inspiró en los cuentos del Barón de Münchhausen; en uno de ellos, el Barón se salva de ahogarse tirando de las cinchas traseras de sus propias botas (*bootstrap*), lo que supone un buen símil para una técnica que ‘tira de los datos’ disponibles sin recabar más.

El artículo de Efron ***"Bootstrap methods: Another look at the Jackknife"*** fue publicado en 1979 en la revista *The Annals of Statistics*. Es uno de los trabajos estadísticos más citados en la literatura y su repercusión es muy alta en muchos campos, sobre todo en estudios biomédicos.

Bradley Efron se doctoró en Estadística, bajo la supervisión de Rupert G. Miller Jr., en la Universidad de Stanford, donde ha desarrollado toda su carrera académica e investigadora, desempeñándose en la actualidad como profesor emérito. Ha recibido reconocimientos de prestigio muy elevado. Entre ellos puede mencionarse el *Premio Fronteras del Conocimiento* en Ciencias Básicas de la Fundación BBVA en 2016, compartido con el Profesor David R. Cox (quien, por su parte, fue una ayuda decisiva en la tesis doctoral del Profesor Azorín Poch, durante la estancia de este en el Laboratorio de Estadística de la Universidad de Cambridge).

De las muchas aplicaciones del *bootstrap*, merece atención especial en el contexto de este discurso la relativa a algunos problemas relacionados con los *procesos empíricos generales*. La teoría de los procesos empíricos, si bien se había iniciado hacia los años veinte del siglo pasado, tuvo un renacimiento y una evolución muy rápida a partir de los estudios de Vapnik y Chervonenkis (1971) y de Dudley (1978) y de los avances en teoría de la probabilidad y en los resultados límite para elementos aleatorios con valores en espacios de Banach. En una revisión sobre el tema, Giné (1996) resumía el objetivo del mismo diciendo que: *«La teoría de los procesos empíricos aborda la cuestión básica de hasta qué punto las frecuencias (o la media muestral) aproximan las probabilidades (o el valor esperado)»*.

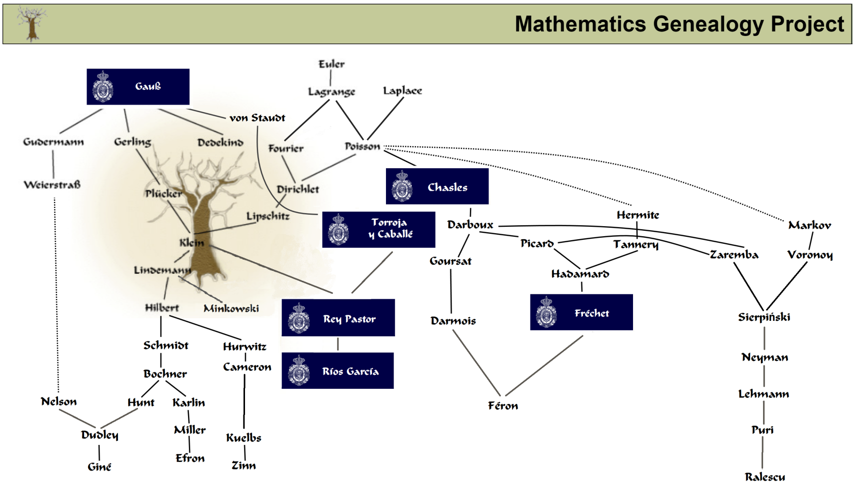
Evarist Giné (izquierda) y Joel Zinn (derecha)

Entre las contribuciones más relevantes, pueden destacarse los trabajos desarrollados por Evarist Giné y Joel Zinn (junto con varios de sus colaboradores), buena parte de los cuales se llevaron a cabo durante los años que ambos coincidieron en el Departamento de Matemáticas de la Universidad de Texas A&M.

Tras licenciarse en Matemáticas en la Universidad de Barcelona, **Evarist Giné** realizó su tesis doctoral en el Instituto Tecnológico de Massachusetts, bajo la supervisión de Richard M. Dudley. Después de varios cursos en el Instituto Venezolano de Investigaciones Científicas en Caracas, en los que inició una de sus líneas principales sobre el estudio de teoremas límite en espacios de Banach de dimensión infinita, ejerció un año como profesor visitante en la Universidad de California en Berkeley, donde tuvo la oportunidad de coincidir con los estadísticos de la conocida como la 'Era Dorada de Berkeley'. A continuación de una breve vuelta a Venezuela, un corto tiempo en la Universidad Autónoma de Barcelona (en el BOE 1983–1856 aparece la Orden de 19/11/1982 por la que se le nombra profesor agregado de "Estadística matemática y Cálculo de probabilidades - Teoría de la Probabilidad" en su Facultad de Ciencias) y otro en la Universidad Estatal de Louisiana, fue contratado en la Universidad de Texas A&M promocionando a *professor*. Se considera que en esa época empezó buena parte de su investigación más original e influyente, con diferentes colaboraciones, pero especialmente con Joel Zinn. Más tarde, estuvo dos años en la Universidad de la Ciudad de Nueva York (CUNY) y en 1990 pasó a ser *professor* en la Universidad de Connecticut, donde permaneció hasta el final.

Sus conexiones con investigadores españoles fueron muy habituales. Además de dirigir en Texas A&M la tesis del Profesor Juan Romo, actualmente en la Universidad Carlos III de Madrid, mantuvo varias colaboraciones, entre otros, con estadísticos de la Universidad de Valladolid (como los Profesores Eustasio del Barrio y Carlos Matrán) y de la Autónoma de Barcelona (como el Profesor Frederic Utzet). Fue también Editor Asociado de la revista *TEST* de la Sociedad de Estadística e Investigación Operativa durante veinte años. A lo largo de toda su carrera científica, recibió un número considerable de nombramientos, distinciones y premios.

En cuanto a **Joel Zinn**, tras graduarse en Matemáticas en el Queens College de CUNY, se trasladó a Madison, Wisconsin, para realizar el doctorado en Matemáticas con la supervisión de Jim Kuelbs. Tras pasar por las Universidades de Minnesota, Massachusetts y Estatal de Michigan, con treinta y cinco años fue contratado en la Universidad de Texas A&M como *associate professor*. Allí permaneció hasta el final de su carrera como profesor emérito.



Árbol genealógico científico de los
Profesores Bradley Efron, Evarist Giné y Joel Zinn
Fuente: *Mathematics Genealogy Project*

Compartiendo con Giné su querencia por la investigación fundamental que se publicaba en las principales revistas, fueron coautores de muchas publicaciones (la *Web of Science* recoge 21 de ellas). Sus trabajos tenían una formalización muy rigurosa y una puesta al día excepcional en los avances de la teoría de la probabilidad, algunos de los cuales dieron visibilidad a trabajos inéditos de Lucien Le Cam. Sus resultados fueron muy oportunos para respaldar el papel del *bootstrap* en problemas complejos.

A partir de principios de los ochenta, Giné y Zinn establecieron, mayoritariamente de forma conjunta, varios resultados límite para elementos aleatorios con valores en espacios de Banach separables o para procesos empíricos. En el trabajo ***"Bootstrapping general empirical measures"***, publicado en *The Annals of Probability* en 1990, ambos dieron condiciones para que se satisficiera casi seguro la aproximación *bootstrap* del teorema del límite central para procesos empíricos indexados por cierta clase de funciones; demostraron la consistencia casi seguro del *naive boostrap* para la media muestral en un espacio de tales procesos, siendo válida dicha consistencia para elementos aleatorios independientes con valores en espacios de Hilbert separables.

Recientemente, González-Rodríguez y Colubi (2017) han derivado la consistencia para otras aproximaciones *bootstrap* de las medias muestrales (por ejemplo, *wild bootstrap*, *bootstrap* ponderado, etc.).

Siguiendo las pautas marcadas por Giné y Zinn (1990), investigadores de SMIRE/SMIRE↯CODIRE introdujeron en 2012 un test ANOVA de igualdad de medias para elementos aleatorios con valores en un espacio de Hilbert separable sobre la base de resultados límite para esos elementos y de algunos teoremas clásicos. Si bien el resultado sería un procedimiento de contraste asintótico consistente, los procesos gaussianos involucrados en la distribución límite suelen desconocerse, por lo que sería idóneo recurrir a una aproximación *bootstrap* del contraste asintótico ANOVA que va a ser consistente y únicamente intervienen en él valores muestrales. El método resultante podría aplicarse directamente para elementos aleatorios con valores en $\mathbb{H}_2$ con la métrica dada por la norma $\|\cdot\|_{\tau}^{\varphi}$.

Si se tratara de particularizar el test ANOVA asintótico para contrastar la igualdad de medias de conjuntos *fuzzy* aleatorios, la particularización no sería efectiva en la práctica, ya que el límite comprometería procesos gaussianos cuyos valores no irían asociados con conjuntos *fuzzy* aleatorios. Sin embargo, la aproximación *bootstrap* del test ANOVA sí puede particularizarse para datos *fuzzy* sin inconvenientes prácticos, ya que únicamente intervendrían en los cálculos medias muestrales (cuyos valores permanecerían dentro del cono de los valores *fuzzy*) y de distancias entre valores *fuzzy*, que tomarían valores reales. Siguiendo el mismo esquema de razonamiento se establecieron los casos uni- y bimuestral.

■ En los MÉTODOS *AD HOC* PARA ANALIZAR DATOS *FUZZY*:
Las vías usuales son:

– la búsqueda de soluciones exactas (medidas, estimaciones, o contrastes), siempre que el planteamiento lo haga posible;
– la consideración de poblaciones finitas o de conjuntos *fuzzy* aleatorios discretos, tanto para el desarrollo de estudios exactos (por ejemplo, estimaciones de parámetros según diferentes tipos de muestreo) como para algunos de carácter asintótico, apoyándose en la teoría de grandes muestras;
– el empleo de procedimientos asintóticos para conjuntos *fuzzy* aleatorios generales y, si es posible, aproximación *bootstrap* y análisis de la consistencia correspondiente.

■ El *quid* de los MÉTODOS BASADOS EN SIMULACIONES:
Consiste en recurrir a técnicas de simulación de datos *fuzzy* para, o bien corroborar/ilustrar resultados teóricos ya probados, o bien detectar tendencias mayoritarias cuando no hay conclusiones generales. Los mecanismos de generación de los datos *fuzzy* simulados suelen estar inspirados en ciertos ajustes a ejemplos del mundo real en los que SMIRE/SMIRE⚡CODIRE ha participado a lo largo de estos años.

Estudios para analizar datos *fuzzy* según la metodología con conjuntos *fuzzy* aleatorios

Los estudios y métodos para analizar datos *fuzzy* según el enfoque y la metodología basados en los conjuntos *fuzzy* aleatorios han sido iniciados y desarrollados hasta el momento esencialmente por SMIRE/SMIRE⚡CODIRE y se resumen esquemáticamente como sigue:

⧫ Métodos inferenciales sobre la media tipo Aumann

- ◇ estimación 'puntual' y 'por intervalo' de la media (*fuzzy*) poblacional;
- ◇ test 'bilaterales' (i.e., sobre la igualdad) de medias (*fuzzy*) poblacionales de una muestra y de k muestras independientes o ligadas.

⧫ Métodos inferenciales sobre la varianza tipo Fréchet

- ◇ estimación puntual y por intervalo de la varianza (real) poblacional;
- ◇ test bi- y unilaterales de varianzas poblacionales de una muestra y de igualdad de varianzas de k muestras independientes.

⧫ Estadística robusta con datos *fuzzy*

- ◇ cuantificación/estimación de la localización (tendencia central) de conjuntos de datos *fuzzy* y análisis de su robustez frente a cambios de datos o presencia de valores atípicos;
- ◇ cuantificación/estimación de la escala (dispersión) de conjuntos de datos *fuzzy* y análisis de su robustez frente a cambios de datos o presencia de valores atípicos.

♦ Análisis de sensibilidad respecto de la forma de los datos *fuzzy*

- ◇ estudios empíricos descriptivos e inferenciales, con diferentes medidas de localización, y contraste de igualdad de medias para distintas descripciones *fuzzy* de datos imprecisos simulados (que compartan el 0- y el 1-nivel y tengan una interpretación 'cercana');
- ◇ estudios empíricos descriptivos e inferenciales, con diferentes medidas de escala, y contraste de igualdad de varianzas para las descripciones *fuzzy* de datos imprecisos antedichas.

OBSERVACIÓN: Se concluye que los resultados estadísticos no están afectados significativamente por la forma adoptada, lo que respalda el uso de números *fuzzy* trapezoidales, facilitando los cálculos, las interpretaciones y, también, la implementación, la cumplimentación y la recogida de datos *online*.

♦ Análisis comparativo entre las escalas de valoración/medida para datos imprecisos

- ◇ estudios empíricos descriptivos e inferenciales, con diferentes medidas de localización y de escala para distintas escalas de valoración de los datos imprecisos (escalas tipo Likert, visuales analógicas, intervalares, etc.);
- ◇ contraste de igualdad de medias para las distintas escalas de valoración de datos imprecisos simulados.

OBSERVACIÓN: Las conclusiones estadísticas están a menudo afectadas significativamente por la escala de valoración/medida adoptada, lo que refuerza el interés del estudio comparativo para las aplicaciones principales.

♦ Otros estudios estadísticos con datos *fuzzy*

- ◇ estudios sobre la cuantificación, estimación y contrastes de la desigualdad de un conjunto *fuzzy* aleatorio;
- ◇ estudios de análisis de regresión (lineal y no paramétrica) con datos *fuzzy*;
- ◇ estudio del análisis de decisión estadística bayesiano con pérdidas *fuzzy*;
- ◇ clasificación de datos *fuzzy*;
- ◇ etc.

El desarrollo de estos estudios ha ido acompañado casi sistemáticamente del desarrollo de **algoritmos** y de ***software*** (mayoritariamente paquetes en R) que implementan gran parte de los cálculos asociados a la metodología.

La fortaleza de esta metodología para datos *fuzzy* reside en el hecho de que se apoya en fundamentos matemáticos sólidos (sobre todo del marco probabilístico): espacios de Hilbert separables, encaje isométrico, medibilidad Borel, etc.

Pero, como ocurría con los datos numéricos o multivariantes, la habitual complejidad matemática tras los resultados teóricos apenas redunda en complejidad de aplicación/computacional, gracias al progreso e incremento del *software* que se está realizando. Si los usuarios potenciales se percataran de esto último, se irían haciendo más populares el empleo de datos *fuzzy* para describir datos imprecisos (especialmente provenientes de 'puntuaciones humanas', como anticipaban Zadeh, Azorín y Ríos) y la metodología para su análisis estadístico.

Y se espera que vayan resolviéndose gradualmente los muchos problemas abiertos y sigan surgiendo otros nuevos.

Algunas aplicaciones de la metodología

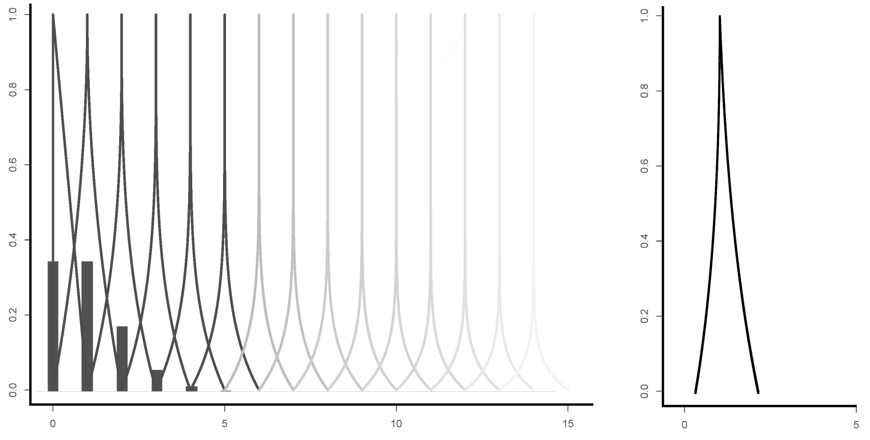
Entre las aplicaciones de la metodología que se acaba de exponer hay dos de índole muy diferente que revisten un interés singular. Para ambas quedan bastantes aspectos e implicaciones por examinar.

⧫ Caracterización de la distribución de una variable aleatoria mediante la media tipo Aumann de transformaciones *fuzzy* especiales y aplicaciones inmediatas

Cuando se considera una variable aleatoria real, hay dos funcionales principales que caracterizan su distribución con independencia de que se trate de una distribución discreta, continua o mixta: la función de distribución y la función característica.

En SMIRE/SMIRE↯CODIRE se propuso en 2006 una familia de funciones de variable real con valores de número *fuzzy* que, al componerse con una variable aleatoria cualquiera, dan lugar a conjuntos *fuzzy* aleatorios cuya media tipo Aumann caracteriza la distribución de la variable aleatoria. Es decir, existen funciones $\gamma : \mathbb{R} \to \mathscr{F}(\mathbb{R})$ tales que, si X e Y son dos variables aleatorias asociadas al mismo espacio de probabilidad, se cumple que:

$$X, Y \text{ idénticamente distribuidas} \Leftrightarrow \widetilde{E}(\gamma(X)) = \widetilde{E}(\gamma(Y)).$$



Ejemplo de γ-transformación *fuzzy* caracterizadora de una Poisson $\mathscr{P}(1)$ (izquierda) y su media tipo Aumann (derecha)

Al igual que la función de distribución, el funcional caracterizador definido a través de la media tipo Aumann de la variable real transformada mediante una función γ del tipo indicado, es una función de variable real con valores en el intervalo $[0, 1]$. Pero, además, es un funcional basado en un valor medio, lo que le confiere un valor estadístico añadido.

Como consecuencia directa, la combinación de una métrica entre números *fuzzy* y la representación *fuzzy* caracterizadora da lugar a una distancia entre distribuciones de variables aleatorias. Por ello, acudiendo a los test para las medias tipo Aumann pueden construirse test para el contraste de bondad de ajuste, el contraste de igualdad de distribuciones o el contraste de simetría, entre otros procedimientos.

♦ Empleo de escalas de valoración *fuzzy* libre en cuestionarios con ítems de respuesta única

La medición asociada con valoraciones humanas (actitudes, opiniones, percepciones, sentimientos, etc.) ha consistido tradicionalmente en elegir o marcar números, etiquetas lingüísticas/cualitativas o símbolos para una amplia gama de variables. Aunque las mediciones de estas valoraciones se han llevado a cabo a través de diferentes vías e instrumentos, han recurrido principalmente al empleo de cuestionarios.

El formato de respuesta de cada pregunta del cuestionario juega un papel crucial en el diseño del mismo, de manera que debe ser un paso explícito de ese diseño a añadir a los de la definición de los ‘constructos’, de la población objetivo, de la selección de los ítems, etc. El formato de respuesta más común es el basado en una escala de valoración que, habitualmente, define las gradaciones de un continuo como el acuerdo, la satisfacción, la intensidad, etc. Los encuestados evalúan las preguntas y los elementos, eligiendo/marcando la 'categoría' que 'mejor representa' su puntuación o respuesta.

La imprecisión intrínseca a las valoraciones humanas y la continuidad usual de sus gradaciones avalan la necesidad de escalas de valoración que capten la imprecisión y las diferencias individuales. Con este propósito, el científico computacional Tim Hesketh junto con los psicólogos Richard Pryor y Beryl Hesketh, de la Universidad de Nueva Gales del Sur, introdujeron en 1988 las escalas de valoración *fuzzy* libre (*fuzzy rating scales*) como una nueva escala psicométrica.

Desarrollaron aplicaciones a algunos problemas del mundo real, casi todos ellos en relación con estudios sobre el empleo, la profesión y el género. Pero, los análisis estadísticos de las respuestas a los cuestionarios no se correspondieron con un análisis que, siguiendo el principio holístico de que el todo no es igual a la suma de las partes, contemplara cada dato *fuzzy* como un todo. Por el contrario, realizaron análisis separados para cada uno de los extremos de los 0- y 1-niveles. A ello probablemente obedece que escribieran la frase que se ha citado en la introducción de la metodología que acaba de presentarse, acerca de la demanda de técnicas estadísticas oportunas para analizar datos *fuzzy*, varias de las cuales ya están disponibles.

En SMIRE/SMIREϟCODIRE de la Universidad de Oviedo y en el Grupo SMABSS del Principado de Asturias (integrado por los componentes de SMIREϟCODIRE junto con algunos profesores de Psicología e Informática de nuestra universidad), se han elaborado y administrado distintos cuestionarios en los que todos o casi todos los ítems involucraban simultáneamente una escala de valoración *fuzzy* libre y una escala tipo Likert (de 4, 5, 6 y 7 puntos). Las poblaciones objetivo han sido muy diversas: desde alumnos de grado o máster de Matemáticas/Ciencias de la Computación a niños de cuarto curso de Primaria. Si bien la cumplimentación de un cuestionario con escala *fuzzy* libre suele requerir cierta guía o cierto entrenamiento previo, en la práctica no debería ser una tarea compleja como nos demostraron de modo incontestable los niños de nueve años.

El análisis de las respuestas a estos cuestionarios ha permitido:

– corroborar empíricamente que las conclusiones pueden ser muy diferentes según la escala de medida empleada;

– inspirar patrones para las simulaciones desarrolladas;

– comparar mediante esos ejemplos y las simulaciones que, según el índice de Cronbach, la 'consistencia interna' de los constructos es mayoritariamente más alta cuando las respuestas se basan en la escala de valoración *fuzzy* libre que cuando lo hacen en escalas tipo Likert (codificadas numéricamente o mediante etiquetas lingüísticas *fuzzy*) o, incluso, en escalas visuales analógicas; en consecuencia, el empleo de las *fuzzy rating scales* se revela recomendable cuando las conclusiones derivadas de la administración de los cuestionarios puedan tener impacto trascendental (en políticas sociales, sanitarias, etc.).

Algunas consideraciones adicionales

Desde el día en que el Profesor Francisco Azorín Poch pronunció su discurso de ingreso en esta Academia hasta nuestros días, las ideas de reconciliación y coexistencia de las teorías de probabilidad y estadística con la de conjuntos *fuzzy*, y el desarrollo de modelos y métodos híbridos, han prosperado mucho.

Conviene diferenciar en este punto el <u>análisis *fuzzy* de datos</u> del <u>análisis de datos *fuzzy*</u>. Mientras que el primero se ocupa de analizar datos clásicos utilizando métodos basados o inspirados en la teoría de conjuntos *fuzzy* (por ejemplo, clasificación *fuzzy* de datos, análisis de regresión *fuzzy*, etc.), el segundo trata de analizar datos intrínsecamente *fuzzy*. El editorial del número especial *"Fuzzy data analysis and classification. Special issue in memoriam of Professor Lotfi A. Zadeh, father of fuzzy logic"* en la revista *Advances in Data Analysis and Classification* (D'Urso y Gil, 2017) hace un recorrido bastante exhaustivo de los procedimientos existentes según ambos enfoques.

Aunque en la literatura pueden encontrarse bastantes contribuciones probabilísticas sobre los conjuntos *fuzzy* aleatorios, son aún pocos los investigadores que se han preocupado por las de carácter estadístico. Pueden señalarse algunos trabajos notables del Profesor Wolfgang Näther (descendiente científico de Felix Klein) pero sus colaboradores más cercanos, tras obtener el doctorado, orientaron su quehacer laboral al trabajo en la empresa.

También se mantienen colaboraciones frecuentes sobre el tema específico de este discurso con investigadores de, entre otras, la Sapienza Università di Roma, la Technische Universität Wien y la Katholieke Universiteit Leuven. En torno al mismo y a otros estrechamente relacionados, pueden destacarse, además, las actividades siguientes:

- la consecución de la Acción COST (Cooperación Europea en Ciencia y Tecnología) *IC0702 - Combining Soft Computing Techniques and Statistical Methods to Improve Data Analysis Solutions*, que entre 2008 y 2012 se gestionó desde el Centro Europeo de Soft Computing y en el que nuestra participación fue crucial en todas las etapas, liderando el grupo de trabajo más activo y numeroso sobre *'Statistics with imperfect and incomplete data'*;
- la serie de congresos bienales internacionales *'Soft Methods in Probability and Statistics'* que, desde 2002, se organizan con el concurso permanente de miembros de SMIRE/ SMIRE⚡CODIRE en los Comités Asesor y Científico; cada congreso va acompañado de un libro editado por Springer Nature con una selección de trabajos presentados en él (recogidos individualmente después en la *Web of Science*) y, en los últimos años, de un número especial de la revista *International Journal of Approximate Reasoning* con una selección más depurada de versiones más largas de algunos de los trabajos, e incluso abierta a otras contribuciones relevantes;
- la edición de varios números especiales de revistas como *Information Sciences*, *Computational Statistics and Data Analysis*, *Fuzzy Sets and Systems*, *Metron* y *Advances in Data Analysis and Classification.*

Comentarios finales y más agradecimientos

Es el mejor de los buenos
quien sabe que en esta vida
todo es cuestión de medida:
un poco más, algo menos

(Antonio Machado)

Por la última sección voy entrando y llega el momento de confesar que, si bien la docencia me atraía desde corta edad, por la investigación tardé tiempo en sentir lo mismo. Muy posiblemente, al principio la entendí como una obligación para desarrollar una vida universitaria completa y para retornar algo de lo que nos había dado Pedro Gil, que con treinta y tres años había dirigido ya cuatro tesis y asumido desde los veintiocho la carga de la dirección de un departamento en el que todos éramos muy jóvenes y necesitábamos de su guía para saber qué se esperaba de un profesor de universidad. No contemplé la investigación nunca como una Ítaca a la que llegar, aunque he disfrutado tremendamente del viaje y espero poder seguir haciéndolo.

Gracias a los académicos por permitirme formar parte de esta 'cofradía', como la denominaba Rey Pastor en su contestación al discurso de ingreso de Sixto Ríos. Junto al honor que representa ser cofrade, es una oportunidad única e impagable para admirar y aprender. Además, lo he pasado tan bien preparando este discurso...

Gracias a todos mis compañeros del Departamento de Estadística, Investigación Operativa y Didáctica de la Matemática, de la Facultad de Ciencias de la Universidad de Oviedo y de la joven Academia Asturiana de Ciencia e Ingeniería (AACI). Y a los del área de conocimiento de Estadística e Investigación Operativa de toda España (son muchos los que podrían estar en mi lugar).

Un agradecimiento especial a todos los miembros de SMIRE (a los que están y a los que ya no están —estos últimos, afortunadamente, por razones de jubilación o por haber formado ellos su propio grupo—), de SMIREϟCODIRE y de SMABSS y a los colaboradores externos. De cada uno de ellos he aprendido mucho, como en su día aprendí de tantos buenos profesores y compañeros en todos los niveles.

Siento un orgullo enorme de mi gran familia científica. Las personas más influyentes en nuestra investigación, o bien descienden de Leonhard Euler/los Bernoulli o lo hacen de Carl Gauss o, incluso, de todos ellos, como nosotros mismos.

Pero estoy, incluso, más ufana de mi familia biológica. En la vida hay personas que, de alguna manera, nos vienen dadas: padres, hermanos y, en cierta medida, hasta los hijos (las hijas, en mi caso); y, aún no habiendo podido elegir de modo plenamente libre, la vida se ha portado extraordinariamente bien conmigo. Eso sí, sin ser una experta en la toma de decisiones, puedo presumir de que la única vez que he realizado una elección personal completamente libre, sin que me viniera dada y con una información *a priori* un tanto 'indeterminada', la decisión tomada fue óptima (se mire con el criterio que se mire): ¡gracias, José Manuel, por quererme, cuidarme, inspirarme y ser siempre el mejor regalo posible de generosidad, bondad y honestidad para nuestras hijas, para mí y para los que te tenemos cerca!

Para terminar y tratando de rebajar la tensión emocional del momento, dejadme recordar de la conocida película "AMANECE, QUE NO ES POCO", con guión y dirección de José Luis Cuerda, algunas de las rogativas que, a modo de plegarias, se decían en las "Letanías de los que están en los cielos" de una suerte de 'rosario del crepúsculo' previo a las elecciones a alcalde en el pueblo. Tras mencionar cada uno de los distintos tipos de ángeles según los escritos bíblicos (tronos, dominaciones, potestades, principados, etc.), los vecinos hacían sus rogativas y, pese a que todas se adaptan muy bien al espíritu de la RAC, hay dos que, como dirían en la propia película, *«vienen aquí pintiparadas»*:

«¡Dadnos, santos del cielo, la capacidad de relativizar!»

«¡Dadnos, santos del cielo,
una visión global bastante aproximada!»

¡GRACIAS! ¡MUCHAS GRACIAS!

Referencias y bibliografía principal

Araujo, A., Giné, E.: *The Central Limit Theorem for Real and Banach Valued Random Variables.* J. Wiley & Sons, New York (1980)

Aumann, R.J.: Integrals of set-valued functions. *J. Math. Anal. Appl.* **12** (1), 1–12 (1965)

Barbut, M.: Une episode insolite des relations scientifiques franco-ibériques: le sejour au Portugal et en Espagne de Maurice Fréchet, en janvier et février 1942. En: Santos del Cerro, J., García Secades, M. (eds.) *Historia de la Probabilidad y la Estadística (III)*, pp. 209–219. Delta Pub. Univ., Elche (2006)

Bingham, N.H.: Studies in the history of Probability and Statistics XLVI. Measure into Probability: From Lebesgue to Kolmogórov. *Biometrika* **87** (1), 145–56 (2000)

Diday, E.: Optimisation en classification automatique et reconnaissance des formes. *RAIRO Oper. Res.* **6** (3), 61–95 (1972)

Dudley, R.M.: Central limit theorems for empirical measures. *Ann. Probab.* **6**, 899–929 (1978)

Efron, B.: Bootstrap methods: Another look at the Jackknife. *Ann. Statist.* **7** (1), 1–26 (1979)

Efron, B.: Nonparametric estimates of standard error: The jackknife, the bootstrap and other methods. *Biometrika* **68** (3), 589–599 (1981)

Efron, B.: More efficient bootstrap computations. *J. Am. Stat. Assoc.* **85** (409), 79–89 (1990)

Efron, B.: The Bootstrap and Modern Statistics. *J. Am. Stat. Assoc.* **95** (452), 1293–1296 (2000)

Efron, B.: Second Thoughts on the Bootstrap. *Stat. Sci.* **18** (2), 135–140 (2003)

Efron, B.: A 250-year argument: Belief, behavior, and the bootstrap. *Bull. Amer. Math. Soc.* **50** (1), 129–146 (2013)

Efron, B., Tibshirani, R.: Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Stat. Sci.* **1** (1), 54–75 (1986)

Féron, R.: Ensembles aléatoires flous. *C.R. Acad. Sc. Paris, Sér. A* **282**, 903–906 (1976)

Féron, R.: Economie d'échange aléatoire floue. *C.R. Acad. Sc. Paris, Sér. A* **282**, 1379–1382 (1976)

Féron, R.: Ensembles flous, ensembles aléatoires flous, et économie aléatoire floue. *Publications Econométriques* **IX** (1), 25–64 (1976)

Féron, R.: Ensembles aléatoires flous dont la fonction d'appartennance prend ses valeurs dans un treillis distributif fermé. *Publications Econométriques* **XII** (1), 81–118 (1979); Bibliographical addendum **XII** (2), 63–67 (1979)

Féron, R.: Sur les notions de distance et d'écart dans une structure floue et leurs applications aux ensembles aléatoires flous. Cas où le référentiel n'est pas métrique. *C.R. Acad. Sc. Paris, Sér. A* **289**, 35–38 (1979)

Féron, R, Féron, M.: Fuzzy specifications and random fuzzy events considered as basic tools for statistical prediction. *Fuzzy Sets Syst.* **28** (3), 285–293 (1988)

Fréchet, M.: Sur quelques points du calcul fonctionnel. *Rend. Circ. Matem. Palermo* **22**, 1–72 (1906) [Traducción: Sobre una definición del número de dimensiones. *Rev. R. Acad. Cienc. Exact. Fís. Nat.* **108** (No. Extr. 1), 115–122 (2006)]

Fréchet, M.: Les éléments aléatoires de nature quelconque dans un espace distancié. *Ann. Inst. H. Poincaré* **10** (4), 215–310 (1948)

Fréchet, M.: Conferencias sobre los elementos aleatorios de naturaleza cualquiera. *Trab. Estad.* **1** (2), 157–181 (1950)

Fréchet, M.: La Estadística. Sus fines, su aplicaciones, su enseñanza. *Trab. Estad.* **1** (2), 183–197 (1950)

Fréchet, M.: Une nouvelle théorie: celle des éléments aléatoires abstraits. *Rev. Phil. de la France et de l'Étranger* **147**, 145–158 (1957)

Giné, E.: Empirical processes and applications: an overview. *Bernoulli* **2** (1), 1–28 (1996)

Giné, E., Hahn, M.G., Zinn, J.: Limit theorems for random sets. An application of probability in Banach space results. En: Beck, A., Jacobs, K. (eds.) *Probability in Banach Spaces, IV* (Series: Lecture Notes in Mathematics, Vol. 990), pp. 112–135. Springer, Heidelberg (1983)

Giné, E., Zinn, J.: Some limit theorems for empirical processes. *Ann. Probab.* **12** (4), 929–989 (1984)

Giné, E., Zinn, J.: Necessary conditions for the bootstrap of the mean. *Ann. Statist.* **7** (2), 684–691 (1989)

Giné, E., Zinn, J.: Bootstrapping general empirical measures. *Ann. Probab.* **18** (2), 851–869 (1990)

González-Manteiga, W., Prada-Sánchez, J.M., Romo, J.J.: The Bootstrap - A review. *Comput. Stat.* **9**, 165-205 (1994)

Hesketh, B. Griffin, B., Loh, V.: A future-oriented retirement transition adjustment framework. *J. Vocat. Behav.* **79** (2), 303–314 (2011)

Hesketh, T., Hesketh, B.: Computerized fuzzy ratings: The concept of a fuzzy class. *Behav. Res. Methods Instr. Comput.* **26** (3), 272–277 (1994)

Hesketh, B., McLachlan, K., Gardner, D.: Work adjustment theory: An empirical test using a fuzzy rating scale. *J. Vocat. Behav.* **40** (3), 318–337 (1992)

Hesketh, T., Pryor, R., Hesketh, B.: An application of a computerized fuzzy graphic rating scale to the psychological measurement of individual differences. *Int. J. Man-Mach. Stud.* **29** (1), 21–35 (1988)

Hilbert, D.: Mathematische Probleme. Vortrag, gehalten auf dem internationalen Mathematike-Congress zu Paris 1900. *Gött. Nachr.*, 253–297 (1900) [Traducción: Mathematical Problems. *Bull. Amer. Math. Soc.* **8** 437–479 (1902)]

Kaufmann, A.: *Introduction à la Théorie des Sous-Ensembles Flous à l'usage des ingénieurs, Vol. 1. Eléments Théoriques de Base.* Masson, Paris (1973) [Traducción: *Introduction to the Theory of Fuzzy Subsets, Vol. 1.* Academic Press, New York (1975)]

Kaufmann, A.: *Introduction à la Théorie des Sous-Ensembles Flous, Vol. 2. Applications à la Linguistique, à la Logique, et à la Sémantique.* Masson, Paris (1975)

Kaufmann, A.: *Introduction à la Théorie des Sous-Ensembles Flous, Vol. 3. Applications à la Classification et à la Reconnaissance des Formes, aux Automates et aux Systèmes, et au Choix des Critères.* Masson, Paris (1975)

Kaufmann, A.: *Introduction à la Théorie des Sous-Ensembles Flous, Vol. 4. Compléments et Nouvelles Applications.* Masson, Paris (1977)

Kendall, D.G.: Foundations of a theory of random sets. En: Harding, E.F., Kendall, D.G. (eds.) *Stochastic Geometry.* J. Wiley & Sons, New York (1974)

Klement, E.P., Puri, M.L., Ralescu, D.A.: Limit theorems for fuzzy random variables. *Proc. R. Soc. London A* **407**, 171–182 (1986)

Kolmogórov, A.N.: *Grundbegriffe der Wahrscheinlichkeitsrechnung.* Springer, Heidelberg (1933)

Lévy, P.: *Calcul des Probabilités.* Gauthier-Villars, Paris (1925)

López Pellicer, M.: Recuerdo de Julio Rey Pastor. *Rev. R. Acad. Cienc. Exact. Fís. Nat.* **108** (1-2), 55–72 (2015)

Matheron, G.: *Random Sets and Integral Geometry.* J. Wiley & Sons, New York (1975)

Minkowski, H.: Volumen und Oberfläche. *Math. Ann.* **57** (4), 447–495, (1903)

Molchanov, I.: *Theory of Random Sets* (Series: Probability Theory and Stochastic Modelling, Vol. 87). Springer, London (2005)

Nguyen, H.T.: A note on the extension principle for fuzzy sets. *J. Math. Anal. Appl.* **64** (2), 369–380 (1978)

Nualart, D.: Kolmogórov and Probability Theory. *Arbor* **CLXXVIII** (704), 607–619 (2004)

Puri M.L., Ralescu, D.A.: The concept of normality for fuzzy random variables. *Ann. Probab.* **13** (4), 1373—1379 (1985)

Puri M.L., Ralescu, D.A.: Fuzzy random variables. *J. Math. Anal. Appl.* **114** (2), 409–422 (1986)

Puri M.L., Ralescu, D.A.: Convergence theorem for fuzzy martingales. *J. Math. Anal. Appl.* **160** (1), 107–122 (1991)

Ríos García, S.: Julio Rey Pastor, matemático. *Bol. Inform. Fund. Juan March* **158**, 2–14 (1986)

Ripley, B.D.: The foundations of stochastic geometry. *Ann. Probab.* **4**, 995–998 (1976)

Robbins, H.E.: On the measure of a random set. *Ann. Math. Statist.* **14**, 70–74 (1944)

Robbins, H.E.: On the measure of a random set, II. *Ann. Math. Statist.* **15**, 342–347 (1945)

Vapnik, V.N., Chervonenkis, A. Ya.: On the uniform convergence of relative frequencies of events to their probabilities. *Theor. Prob. Appl.* **16** (2), 264–280 (1971)

Zadeh, L.A.: Fuzzy sets. *Inform. Contr.* **8** (3), 338–353 (1965)

Zadeh, L.A.: The concept of a linguistic variable and its application to approximate reasoning. Part 1. *Inf. Sci.* **8**, 199–249; Part 2. *Inf. Sci.* **8**, 301–353; Part 3. *Inf. Sci.* **9**, 43–80 (1975)

Zadeh, L.A.: Discussion: Probability theory and fuzzy logic are complementary rather than competitive. *Technometrics* **37** (3), 271–276 (1995)

Zadeh, L.A.: Fuzzy logic - a personal perspective. *Fuzzy Sets Syst.* **281**, 4–20 (2015)

► Selección de discursos de ingreso en la RAC

Académico: Julio Rey Pastor
Título del discurso: *"Investigaciones sobre el problema del ultracontinuo"*
Contestación: Augusto Krahe
Toma de posesión: 14 de noviembre de 1920

Académico: Pedro M. González Quijano
Título del discurso: *"Azar y determinismo"*
Contestación: Leonardo Torres Quevedo
Toma de posesión: 2 de diciembre de 1925

Académico: Pedro Puig Adam
Título del discurso: *"Matemática y Cibernética"*
Contestación: Antonio Torroja y Miret
Toma de posesión: 5 de marzo de 1952

Académico: Ricardo San Juan Llosa
Título del discurso: *"La abstracción matemática"*
Contestación: Julio Rey Pastor
Toma de posesión: 22 de febrero de 1956

Académico: Sixto Ríos García
Título del discurso: *"Procesos de decisión"*
Contestación: Julio Rey Pastor
Toma de posesión: 21 de junio de 1961

Académico: Darío Maravall Casesnoves
Título del discurso: *"La economía y la sociología como motores de la investigación matemática"*
Contestación: Sixto Ríos García
Toma de posesión: 8 de mayo de 1968

Académico: Manuel Valdivia Ureña
Título del discurso: *"Recientes aspectos del análisis funcional"*
Contestación: Germán Ancochea Quevedo
Toma de posesión: 20 de abril de 1977

Académico: Francisco Azorín Poch
Título del discurso: *"Conjuntos borrosos, estadística y probabilidad"*
Contestación: Sixto Ríos García
Toma de posesión: 2 de diciembre de 1981

Académico: Miguel de Guzmán Ozámiz
Título del discurso: *"Impactos del Análisis Armónico"*
Contestación: Alberto Dou Mas de Xesàs
Toma de posesión: 23 de marzo de 1983

Académico: Francisco Javier Girón González-Torre
Título del discurso: *"Conceptos y técnicas de la estadística bayesiana: Comentarios sobre su estado actual"*
Contestación: Sixto Ríos García
Toma de posesión: 13 de marzo de 1991

Académico: Jesús Ildefonso Díaz Díaz
Título del discurso: *"El mundo de la ciencia y las matemáticas del mundo"*
Contestación: Alberto Dou Mas de Xesàs
Toma de posesión: 19 de noviembre de 1997

Académico: Manuel López Pellicer
Título del discurso: *"En torno al casi centenario Análisis Funcional"*
Contestación: Manuel Valdivia Ureña
Toma de posesión: 29 de abril de 1998

Académico: Carlos López Otín
Título del discurso: *"De genomas y degradomas. Crónica de la exploración molecular de los sistemas proteolíticos humano"*
Contestación: Margarita Salas Falgueras
Toma de posesión: 25 de octubre de 2006

Académico: Jesús María Sanz Serna
Título del discurso: *"Integración geométrica"*
Contestación: Amable Liñán Martínez
Toma de posesión: 28 de noviembre de 2007

Académico: José Bonet Solves
Título del discurso: *"El impacto del análisis funcional en algunos problemas del análisis"*
Contestación: Manuel Valdivia Ureña
Toma de posesión: 23 de abril de 2008

Académico: David Ríos Insua
Título del discurso: *"TIC3: Matemáticas, política y tecnologías de la información y de las comunicaciones"*
Contestación: Francisco Javier Girón González-Torre
Toma de posesión: 27 de febrero de 2008

Académico: Enrique Castillo Ron
Título del discurso: *"Una vida dedicada a la matemática y sus aplicaciones"*
Contestación: Francisco Javier Girón González-Torre
Toma de posesión: 23 de marzo de 2011

Académico: Juan Luis Vázquez Suárez
Título del discurso: *"Senderos de la ciencia. Del operador laplaciano a los procesos difusivos no lineales"*
Contestación: Jesús Ildefonso Díaz Díaz
Toma de posesión: 26 de marzo de 2014

Académico: Manuel de León Rodríguez
Título del discurso: *"Una historia breve de la Mecánica Geométrica"*
Contestación: Pedro Luis García Pérez
Toma de posesión: 29 de noviembre de 2017

Académico: Luis Vega González
Título del discurso: *"El Análisis de Fourier y las ecuaciones diferenciales, nuevos retos"*
Contestación: Juan Luis Vázquez Suárez
Toma de posesión: 28 de abril de 2021

Académico: Daniel Peña Sánchez de Rivera
Título del discurso: *"Observación y cálculo en Estadística de Datos Masivos"*
Contestación: Francisco Javier Girón González-Torre
Toma de posesión: 4 de mayo de 2022

► Selección de lecciones inaugurales de cursos académicos en la Universidad de Oviedo

Conferenciante: Julio Rey Pastor, Catedrático de Análisis Matemático
Título del discurso: *"Los matemáticos españoles del siglo XVI"*
Curso académico: 1913-1914

Conferenciante: Pedro Gil Álvarez, Catedrático de Estadística e Investigación Operativa
Título del discurso: *"Las matemáticas de lo incierto"*
Curso académico: 1996-1997

▶ Selección de trabajos de SMIRE/SMIRE⚡CODIRE

Alonso de la Fuente, M., Terán, P.: Harmonizing two approaches to fuzzy random variables. *Fuzzy Optim. Dec. Making* **19**, 177–189 (2020)

Bertoluzza, C., Gil, M.Á., Ralescu, D.A. (eds.): *Statistical Modeling, Analysis and Management of Fuzzy Data* (Series: Fuzziness and Soft Computing, Vol. 87). Physica, Heidelberg (2002)

Blanco-Fernández, A., Casals, R.M., Colubi, A., Corral, N., García-Bárzana, M., Gil, M.Á., González-Rodríguez, G., López, M.T., Lubiano, M.A., Montenegro, M., Ramos-Guajardo, A.B., de la Rosa de Sáa, S., Sinova, B.: A distance-based statistical analysis of fuzzy number-valued data. *Int. J. Approx. Reas.* **55**, 1487–1501 (2014)

Blanco-Fernández, A., Casals, R.M., Colubi, A., Corral, N., García-Bárzana, M., Gil, M.Á., González-Rodríguez, G., López, M.T., Lubiano, M.A., Montenegro, M., Ramos-Guajardo, A.B., de la Rosa de Sáa, S., Sinova, B.: Rejoinder on "A distance-based statistical analysis of fuzzy number-valued data". *Int. J. Approx. Reas.* **55**, 1601–1605 (2014)

Borgelt, C., Gil, M.Á., Sousa, J.M.C., Verleysen, M. (eds.): *Towards Advanced Data Analysis by Combining Soft Computing and Statistics* (Series: Studies in Fuzziness and Soft Computing, Vol. 285). Springer, Heidelberg (2013)

Borgelt, C., González-Rodríguez, G., Trutschnig, W., Lubiano, M.A., Gil, M.Á., Grzegorzewski, P., Hryniewicz, O. (eds.): *Combining Soft Computing and Statistical Methods in Data Analysis* (Series: Advances in Intelligent and Soft Computing, Vol. 77). Springer, Heidelberg (2010)

Cascos Fernández, I., López-Díaz, M., Gil, M.Á.: Convergence criteria for f-inequality set-valued indices. *Statistics* **38**, 59–66 (2004)

Castaño-Pérez, A.M., Lubiano, M.A., García-Izquierdo, A.L.: Gendered beliefs in STEM undergraduates: A comparative analysis of fuzzy rating *versus* Likert scales. *Sustainability* **12** (15), 6227 (2020)

Colubi, A., Coppi, R., D'Urso, P., Gil, M.Á.: Statistics with fuzzy random variables. *Metron* **LXV** (3), 277–303 (2007)

Colubi, A, Domínguez-Menchero, J.S., López-Díaz, M., Ralescu, D.A.: On the formalization of fuzzy random variables. *Inf. Sci.* **133** (1/2), 3–6 (2001)

Colubi, A, Domínguez-Menchero, J.S., López-Díaz, M., Ralescu, D.A.: A $D_{[0,1]}$-representation of random upper semicontinuous functions. *Proc. Amer. Math. Soc.* **130** (11), 3237–3242 (2002)

Colubi, A., Fernández-García, C., Gil, M.Á.: Simulation of random fuzzy variables: an empirical approach to statistical/probabilistic studies with fuzzy experimental data. *IEEE Trans. Fuzzy Syst.* **10** (3), 384–390 (2002)

Colubi, A., González Rodríguez, G., Gil, M.Á., Trutschnig, W.: Nonparametric criteria for supervised classification of fuzzy data. *Int. J. Approx. Reas.* **52** (9), 1272–1282 (2011)

Colubi, A., López-Díaz, M., Domínguez-Menchero, J.S., Gil, M.Á.: A generalized Strong Law of Large Numbers. *Prob. Theor. Rel. Fiel.* **114**, 401–417 (1999)

Coppi, R., Gil, M.Á., Kiers, H.A.L.: The fuzzy approach to statistical analysis. *Comput. Stat. Data Anal.* **51** (1), 1–14 (2006)

De la Rosa de Sáa, S., Carleos, C., López, M.T., Montenegro, M.: A case study-based analysis of the influence of the fuzzy data shape in quantifying their Fréchet's variance. En: Gil, E., Gil, E., Gil, J., Gil, M.Á. (eds.) *The Mathematics of the Uncertain: A Tribute to Pedro Gil* (Series: Studies in Systems, Decision and Control, Vol. 142), pp. 709—719. Springer, Cham (2018)

De la Rosa de Sáa, S., Gil, M.Á., González-Rodríguez, G., López, M.T., Lubiano, M.A.: Fuzzy rating scale-based questionnaires and their statistical analysis. *IEEE Trans. Fuzzy Syst.* **23** (1), 111–126 (2015)

De la Rosa de Sáa, S., Lubiano, M.A., Sinova, B., Filzmoser, P.: Robust scale estimators for fuzzy data. *Adv. Data Anal. Class.* **11** (4), 731–758 (2017)

De la Rosa de Sáa, S., Lubiano, M.A., Sinova, B., Filzmoser, P., Gil, M.Á: Location-free robust scale estimates for fuzzy data. *IEEE Trans. Fuzzy Syst.* **29** (6), 1682–1694 (2021)

Destercke, S., Denoeux, T., Gil, M.Á., Grzegorzewski, P., Hryniewicz, O. (eds.): *Uncertainty Modelling in Data Science. Advances in Intelligent Systems and Computing* (Series: Advances in Intelligent Systems and Computing, Vol. 832), Springer, Cham (2019)

Dubois, D., Lubiano, M.A., Prade, H., Gil, M.Á., Grzegorzewski, P., Hryniewicz, O. (eds.): *Soft Methods for Handling Variability and Imprecision* (Series: Advances in Soft Computing, Vol. 48). Springer, Heidelberg (2008)

D'Urso, P., Gil, M.Á.: Fuzzy Statistical Analysis: methods and applications. *METRON* **71** (3), 197–199 (2013)

D'Urso, P., Gil, M.Á.: Fuzzy data analysis and classification. *Adv. Data Anal. Class.* **11** (4), 645–657 (2017)

Ferraro, M.B., Giordani, P., Vantaggi, B., Gagolewski, M., Gil, M.Á., Grzegorzewski, P., Hryniewicz, O. (eds.): *Soft Methods for Data Science* (Series: Advances in Intelligent Systems and Computing, Vol. 456), Springer, Cham (2017)

García, D., Lubiano, M.A., Alonso, M.C.: Estimating the expected value of fuzzy random variables in the stratified random sampling from finite populations. *Inf. Sci.* **138** (1/4), 165–184 (2001)

García-Escudero, L.A., Gordaliza, A., Mayo, A., Lubiano, M.A., Gil, M.Á., Grzegoezewski, P., Hryniewicz, O. (eds.): *Building Bridges between Soft and Statistical Methodologies for Data Science* (Series: Advances in Intelligent Systems and Computing, Vol. 1433). Springer, Cham (2023)

García-Izquierdo, A.L., Ramos-Villagrasa, P.J., Lubiano, M.A.: Developing biodata for public manager selection purposes: A comparison between fuzzy logic and traditional methods. *J. Work Organ. Psych.* **36** (3), 231–242 (2020)

Gil, E., Gil, E., Gil, J., Gil, M.Á. (eds.): *The Mathematics of the Uncertain: A Tribute to Pedro Gil* (Series: Studies in Systems, Decision and Control, Vol. 142). Springer, Cham (2018)

Gil, M.Á.: A note on the connection between fuzzy numbers and random intervals. *Stat. Prob. Lett.* **13** (4), 311–319 (1992)

Gil, M.Á.: Fuzzy random variables. *Inf. Sci.* **133** (1/2), 1–2 (2001)

Gil, M.Á.: Statistique et Analyse des Données. En: Bouchon-Meunier, B., Marsala, Ch. (eds.) *Logique floue, principes, aide à la décision (Traité IC2, série informatique et systèmes d'information), Vol. 1,* pp. 205–243. Hermes Sci. Pub.-Lavoisier, Cachan (2003)

Gil, M.Á., Colubi, A., Terán, P.: Random fuzzy sets: why, when, how. *Bol. Est. Inv. Oper.* **30** (1), 5–29 (2014)

Gil, M.Á., González-Rodríguez, G.: Fuzzy *vs* Likert scales in Statistics. En: Trillas, E., Bonissone, P.P, Magdalena, L., Kacprzyk, J. (eds.) *Combining Experimentation and Theory. A Hommage to Abe Mamdani* (Series: Studies in Fuzziness and Soft Computing, Vol. 271), pp. 407–420. Springer, Heidelberg (2012)

Gil, M.Á., González-Rodríguez, G., Colubi, A., Montenegro, M.: Testing linear independence in linear models with interval-valued data. *Comput. Stat. Data Anal.* **51** (6), 3002–3015 (2007)

Gil, M.Á., González-Rodríguez, G., Kruse, R.: Statistics with Imperfect Data. *Inf. Sci.* **245**, 1–3 (2013)

Gil, M.Á., Hryniewicz, O. Statistics with imprecise data. En: Meyers, R.A. (ed.) *Encyclopedia of Complexity and Systems Science, Part 19*, pp. 8679–8690. Springer, Heidelberg (2009)

Gil, M.Á., Jain, P.: Comparison of experiments in decision problems with fuzzy utilities. *IEEE Trans. Syst., Man Cybern.* **22** (4), 662–670 (1992)

Gil, M.Á., López, M.T., Lubiano, M.A., Montenegro, M.: Regression and correlation analyses of a linear relation between random intervals. *Test* **10** (1), 183–201 (2001)

Gil, M.Á., López-Díaz, M.: Fundamentals and Bayesian analyses of decision problems with fuzzy-valued utilities. *Int. J. Approx. Reas.* **15** (3), 203–224 (1996)

Gil, M.Á., López-Díaz, M., López-García, H.: The fuzzy hyperbolic inequality index associated with fuzzy random variables. *Eur. J. Oper. Res.* **110** (2), 377–391 (1998)

Gil, M.Á., López-Díaz, M., Ralescu, D.A.: Fuzzy Sets and Probability/ Statistics Theories. *Fuzzy Sets Syst.* **157** (19), 2545–2545 (2006)

Gil, M.Á., López-Díaz, M., Ralescu, D.A.: Overview on the development of fuzzy random variables. *Fuzzy Sets Syst.* **157** (19), 2546–2557 (2006)

Gil, M.Á., López-Díaz, M., Rodríguez-Muñiz, L.J.: An improvement of a comparison of experiments in statistical decision problems with fuzzy utilities. *IEEE Trans. Syst., Man Cybern.* **28** (6), 856—864 (1998)

Gil, M.Á., Lubiano, M.A., de la Rosa de Sáa, S., Sinova, B.: Analyzing data from a fuzzy rating scale-based questionnaire. A case study. *Psicothema* **27** (2), 182–191 (2015)

Gil, M.Á., Lubiano, M.A., Montenegro, M., López, M.T.: Least squares fitting of an affine function and strength of association for interval-valued data. *Metrika* **56** (2), 97–111 (2002)

Gil, M.Á., Montenegro, M., González-Rodríguez, G., Colubi, A., Casals, M.R.: Bootstrap approach to the multi-sample test of means with imprecise data. *Comput. Stat. Data Anal.* **51** (1), 148–162 (2006)

González-Rodríguez, G., Blanco, A., Colubi, A., Lubiano, M.A.: Estimation of a simple linear regression model for fuzzy random variables. *Fuzzy Sets Syst.* **160** (3), 357–370 (2009)

González-Rodríguez, G., Blanco, A., Corral, N., Colubi, A.: Least squares estimation of linear regression models for convex compact random sets. *Adv. Data Anal. Clas.* **1** (1), 67–81 (2007)

González-Rodríguez, G., Colubi, A., D'Urso, P., Montenegro, M.: Multi-sample test-based clustering for fuzzy random variables. *Int. J. Approx. Reas.* **50** (5), 721–731 (2009)

González-Rodríguez, G., Colubi, A., Gil, M.Á.: A fuzzy representation of random variables: an operational tool in exploratory analysis and hypothesis testing. *Comput. Stat. Data Anal.* **51** (1), 163–176 (2006)

González-Rodríguez, G., Colubi, A., Gil, M.Á.: Fuzzy data treated as functional data. A one-way ANOVA test approach. *Comput. Stat. Data Anal.* **56** (4), 943–955 (2012)

González-Rodríguez, G., Colubi, A., Gil, M.Á., Lubiano, M.A.: A new way of quantifying the symmetry of a random variable: estimation and hypothesis testing. *J. Stat. Plan. Infer.* **142** (12), 3061–3072 (2012)

González-Rodríguez, G., Montenegro, M., Colubi, A., Gil, M.Á.: Bootstrap techniques and fuzzy random variables: synergy in hypothesis testing with fuzzy data. *Fuzzy Sets Syst.* **157** (19), 2608–2613 (2006)

Grzegorzewski, P., Gagolewski, M., Hryniewicz, O., Gil, M.Á. (eds.): *Strengthening Links Between Data Analysis and Soft Computing* (Series: Advances in Intelligent Systems and Computing, Vol. 315). Springer, Cham (2015)

Grzegorzewski, P., Hryniewicz, O., Gil, M.Á. (eds.): *Soft Methods in Probability, Statistics and Data Analysis* (Series: Advances in Soft Computing, Vol. 16). Physica, Heidelberg (2002)

Kruse, R., Berthold, M.R., Moewes, C., Gil, M.Á., Grzegorzewski, P., Hryniewicz, O. (eds.): *Sinergies of Soft Computing and Statistics for Intelligent Data Analysis* (Series: Advances in Intelligent Systems and Computing, Vol. 190). Springer, Heidelberg (2013)

Kruse, R., Gebhardt, J., Gil, M.Á.: Fuzzy Statistics. En: Webster, J. (ed.) *Wiley Encyclopedia of Electrical and Electronics Engineering, Vol. 8,* pp. 181–196. John Wiley & Sons, Inc., New York (1999)

Lawry, J., Miranda, E., Bugarín, A., Li, S., Gil, M.Á., Grzegorzewski, P., Hryniewicz, O. (eds.): *Soft Methods for Integrated Uncertainty Modelling* (Series: Advances in Soft Computing, Vol. 37). Springer, Heidelberg (2006)

López-Díaz, M., Gil, M.Á.: Constructive definitions of fuzzy random variables. *Stat. Prob. Lett.* **36** (2), 135–143 (1997)

López-Díaz, M., Gil, M.Á.: Reversing the order of integration in iterated expectations of fuzzy random variables, and statistical applications. *J. Stat. Plan. Infer.* **74** (1), 11–29 (1998)

López-Díaz, M., Gil, M.Á.: Approximating integrably bounded fuzzy random variables in terms of the 'generalized' Hausdorff metric. *Inf. Sci.* **104** (2), 279–291 (1998)

López-Díaz, M., Gil, M.Á.: An extension of Fubini's Theorem for fuzzy random variables. *Inf. Sci.* **115** (1/4), 29–41 (1999)

López-Díaz, M., Gil, M.Á., Grzegorzewski, P., Hryniewicz, O., Lawry, J. (eds.): *Soft Methodology and Random Information Systems* (Series: Advances in Soft Computing, Vol. 26). Springer-Verlag, Heidelberg (2004)

López-Díaz, M., Ralescu, D.A.: Tools for fuzzy random variables: Embeddings and measurabilities. *Comput. Stat. Data Anal.* **51** (1), 109–114 (2006)

López-Díaz, M., Rodríguez-Muñiz, L.J.: Influence diagrams with super value nodes involving imprecise information. *Eur. J. Oper. Res.* **179** (1), 203–219 (2007)

López-García, H., López-Díaz, M., Gil, M.Á.: Interval-valued quantification of the inequality associate, with a random set. *Stat. Prob. Lett.* **46** (2), 149–159 (2000)

Lubiano M.A., Carleos, C., Montenegro, M., Gil M.Á.: Case study-based sensitivity analysis of scale estimates w.r.t. the shape of fuzzy data. En: Destercke, S., Denoeux, T., Gil, M.Á., Grzegorzewski, P., Hryniewicz, O. (eds.) *Uncertainty Modelling in Data Science* (Series: Advances in Intelligent Systems and Computing, Vol. 832), pp. 157—165. Springer, Cham (2019)

Lubiano, M.A., de la Rosa de Sáa, S.: (2017) FuzzyStatTra: Statistical methods for trapezoidal fuzzy numbers. R package in CRAN (2017) [https://cran.r-project.org/web/packages/FuzzyStatTra/index.html]

Lubiano, M.A., de la Rosa de Sáa, S., Montenegro, M., Sinova, B., Gil, M.Á.: Descriptive analysis of responses to items in questionnaires. Why not a fuzzy rating scale? *Inf. Sci.* **360**, 131–148 (2016)

Lubiano, M.A., García-Izquierdo, A.L., Gil, M.Á.: Fuzzy rating scales: Does internal consistency of a measurement scale benefit from coping with imprecision and individual differences in psychological rating? *Inf. Sci.* **550**, 91–108 (2021)

Lubiano, M.A., Gil, M.Á.: Estimating the expected value of fuzzy random variables in random samplings from finite populations. *Stat. Pap.* **40** (3), 277–295 (1999)

Lubiano, M.A., Gil, M.Á., López-Díaz, M.: On the Rao-Blackwell theorem for fuzzy random variables. *Kybernetika* **35** (2), 167–175 (1999)

Lubiano, M.A., Gil, M.Á., López-Díaz, M., López, M.T.: The $\vec{\lambda}$-mean squared dispersion associated with a fuzzy random variable. *Fuzzy Sets Syst.* **111** (3), 307—317 (2000)

Lubiano, M.A., González-Gil, P., Sánchez-Pastor, H., Pradas, C., Arnillas, H.: An incipient fuzzy logic-based analysis of the medical specialty influence on the perception about mental patients. En: Gil, E., Gil, E., Gil, J., Gil, M.Á. (eds.) *The Mathematics of the Uncertain: A Tribute to Pedro Gil* (Series: Studies in Systems, Decision and Control, Vol. 142), pp. 653—662. Springer, Cham (2018)

Lubiano, M.A., Montenegro, M., Sinova, B., de la Rosa de Sáa, S., Gil, M.Á.: Hypothesis testing for means in connection with fuzzy rating scale-based data: algorithms and applications. *Eur. J. Oper. Res.* **251**, 918–929 (2016)

Lubiano, M.A., Salas, A., Carleos, C., de la Rosa de Sáa, S., Gil, M.Á.: Hypothesis testing-based comparative analysis between rating scales for intrinsically imprecise data. *Int. J. Approx. Reas.* **88**, 128–147 (2017)

Lubiano, M.A., Salas, A., Gil, M.Á.: A hypothesis testing-based discussion on the sensitivity of means of fuzzy data with respect to data shape. *Fuzzy Sets Syst.* **328** (1), 54–69 (2017)

Montenegro, M., Casals, M.R., Lubiano, M.A., Gil, M.Á.: Two-sample hypothesis tests of means of a fuzzy random variable. *Inf. Sci.* **133** (1/2), 89–100 (2001)

Montenegro, M., Colubi, A., Casals, M.R., Gil, M.Á.: Asymptotic and Bootstrap techniques for testing the expected value of a fuzzy random variable. *Metrika* **59** (1), 31–49 (2004)

Ramos-Guajardo, A.B., Colubi, A., González-Rodríguez, G., Gil, M.Á.: One sample tests for a generalized Fréchet variance of a fuzzy random variable. *Metrika* **71** (2), 185–202 (2010)

Ramos-Guajardo, A.B., Lubiano, M.A.: K-sample tests for equality of variances of random fuzzy sets. *Comput. Stat. Data Anal.* **56** (4), 956–966 (2012)

Rodríguez-Muñiz, L.J., López-Díaz, M.: A new framework for the Bayesian analysis of single-stage decision problems with imprecise utilities. *Fuzzy Sets Syst.* **159** (24), 3271–3280 (2008)

Rodríguez-Muñiz, L.J., López-Díaz, M., Gil, M.Á.: Solving influence diagrams with fuzzy chance and value nodes. *Eur. J. Oper. Res.* **167** (2), 444–460 (2005)

Sinova, B.: On depth-based fuzzy trimmed means and a notion of depth specifically defined for fuzzy numbers. *Fuzzy Sets Syst.* **443** (Part A), 87–105 (2022)

Sinova, B., Casals, M.R., Gil, M.Á.: Central tendency for symmetric random fuzzy numbers. *Inf. Sci.* **278**, 599–613 (2014)

Sinova, B., Casals, M.R., Gil, M.Á., Lubiano, M.A.: The fuzzy characterizing function of the distribution of a random fuzzy number. *Appl. Math. Model.* **39** (14), 4044–4056 (2015)

Sinova, B., Colubi, A., Gil, M.Á., González-Rodríguez, G.: Interval arithmetic-based linear regression between interval data: Discussion and sensitivity analysis on the choice of the metric. *Inf. Sci.* **199**, 109–124 (2012)

Sinova, B., de la Rosa de Sáa, S., Casals, R.M., Gil, M.Á., Salas, A.: The mean square error of a random fuzzy vector based on the support function and the Steiner point. *Fuzzy Sets Syst.* **292**, 347–363 (2016)

Sinova, B., de la Rosa de Sáa, S., Gil, M.Á.: A generalized L^1-type metric between fuzzy numbers for an approach to central tendency of fuzzy data. *Inf. Sci.* **242**, 22–34 (2013)

Sinova, B., de la Rosa de Sáa, S., Lubiano, M.A., Gil, M.Á.: An overview on the statistical central tendency for fuzzy datasets. *Int. J. Unc., Fuzz. Know.-Based Syst.* **23** (Suppl. 1), 105–132 (2015)

Sinova, B., Gil, M.Á., Colubi, A., Van Aelst, S.: The median of a random fuzzy number. The 1-norm distance approach. *Fuzzy Sets Syst.* **200**, 99–115 (2012)

Sinova, S., Gil, M.Á., López, M.T., Van Aelst, S.: A parameterized L^2 metric between fuzzy numbers and its parameter interpretation. *Fuzzy Sets Syst.* **245**, 101–115 (2014)

Sinova, B., Gil, M.Á., Van Aelst, S.: M-estimates of location for the robust central tendency of fuzzy data. *IEEE Trans. Fuzzy Syst.* **24** (4), 945–956 (2016)

Sinova, B., González-Rodríguez, G., Van Aelst, S.: M-estimators of location for functional data. *Bernoulli* **24** (3), 2328–2357 (2018)

Sinova, B., Van Aelst, S.: On the consistency of a spatial-type interval-valued median for random intervals. *Stat. Prob. Lett.* **100**, 130–136 (2015)

Sinova, B., Van Aelst, S.: Advantages of M-estimators of location for fuzzy numbers based on Tukey's biweight loss function. *Int. J. Approx. Reas.* **93**, 219–237 (2018)

Sinova, B., Van Aelst, S., Terán, P.: M-estimators and trimmed means: from Hilbert-valued to fuzzy set-valued data. *Adv. Data Anal. Class.* **15**, 267–288 (2021)

Terán, P.: On the equivalence of Aumann and Herer expectations of random sets. *Test* **17** (3), 505–514 (2008)

Terán, P., López-Díaz, M.: Strong consistency and rates of convergence for a random estimator of a fuzzy set. *Comput. Stat. Data Anal.* **77**, 130–145 (2014)

Terán, P., Molchanov, I.: The law of large numbers in a metric space with a convex combination operation. *J. Theor. Prob.* **19** (4), 875–898 (2006)

Trutschnig, W., González-Rodríguez, G., Colubi, A., Gil, M.Á.: A new family of metrics for compact, convex (fuzzy) sets based on a generalized concept of mid and spread. *Inf. Sci.* **179** (23), 3964–3972 (2009)

Trutschnig, W., Lubiano, M.A., Lastra, J.: SAFD - An R Package for Statistical Analysis of Fuzzy Data. En: Borgelt, C., Gil, M.Á., Sousa, J., Verleysen, M. (eds.) *Towards Advanced Data Analysis by Combining Soft Computing and Statistics* (Series: Studies in Fuzziness and Soft Computing, Vol. 285), pp. 107–118. Springer, Heidelberg (2013) [https://cran.r-project.org/web/packages/SAFD/index.html]

CONTESTACIÓN
DEL
EXCMO. SR. D. FRANCISCO JAVIER GIRÓN GONZÁLEZ-TORRE

Contestación

Excmo. Sr. Presidente de la Academia

Excmas. Sras. Académicas

Excmos. Sres. Académicos

Autoridades, compañeros y amigos,

Sras. y Sres.

Quisiera agradecer de nuevo a la Presidencia de la Academia el haberme designado para contestar al discurso de ingreso de la Profesora Dª María Ángeles Gil Álvarez en nombre de nuestra Corporación. Es un placer aceptar este compromiso por diversos motivos: unos de carácter personal y otros de carácter científico por su dedicación a un área de las matemáticas relativamente reciente, relacionada con la estadística, como es la teoría de los conjuntos difusos y sus aplicaciones, sobre todo a la Estadística, las Ciencias de la Computación y la Inteligencia Artificial.

Es curiosa la relación que existe entre la nueva académica, el Profesor Sixto Ríos, el Profesor Azorín y quien les está hablando. Como ha comentado la Profesora Gil, Azorín fue una fuente de inspiración en su interés por todo lo referente a los difusos. De otra parte, la contestación al discurso de ingreso del Profesor Azorín en nuestra Academia corrió a cargo del Profesor Sixto Ríos, quien señaló la importancia que los conjuntos difusos podrían tener dentro de la Estadística. Yo tuve el honor de suceder al Profesor Azorín en la medalla nº 38 —en su momento, de nueva creación— y, de nuevo el Profesor Sixto Ríos fue designado para contestar mi discurso de ingreso.

Aunque mi relación con el Profesor Azorín fue esporádica, tengo que confesar que era una persona muy culta, de trato cordial y, sobre todo, muy humilde, tal como nos lo ha recordado la Profesora Gil.

El Profesor Azorín fue un distinguido esperantista (como también lo había sido, poco antes, Maurice Fréchet). En esa época, en los años ochenta del pasado siglo, Azorín fue un impulsor del esperanto como idioma internacional más adecuado a la comunicación global, aunque su extensión ha sido y es muy limitada en la actualidad. Él, junto con el Profesor Enric Trillas, fue quien introdujo en nuestro país los primeros elementos de los conjuntos difusos (o, como él prefería referirse a ellos, conjuntos borrosos debido a que esta definición cuenta con más derivados que la palabra difuso, según me comentó en una de las pocas oportunidades que tuve de charlar con él).

De hecho, en la primera edición de nuestro Vocabulario Científico y Técnico (VCyT) ya aparecen unos pocos términos referidos a la teoría de conjuntos difusos. Cuando se preparaba la segunda edición del VCyT, el entonces presidente, Ángel Martín Municio, me encargó actualizar e incluir nuevos términos estadísticos y también revisar los ya existentes. Mi conocimiento de los nuevos vocablos de la teoría de los conjuntos difusos era escaso por lo que me puse en contacto con María Ángeles Gil, que por esas fechas ya había comenzado a interesarse en dicha teoría y relacionarla con la probabilidad y la estadística, para que facilitase nuevos términos y definiciones. A ella le debemos la inclusión de los más frecuentes de esa teoría en la segunda edición del VCyT. Esperamos que con su ingreso en nuestra Academia, se actualicen y añadan más en la nueva edición digital del VCyT.

Conocí a María Ángeles a través de su hermano Pedro: ambos ingresamos al mismo tiempo, año 1969, en el Departamento de Estadística e Investigación Operativa de la Universidad Complutense de Madrid, dirigido en esa época por el Profesor Sixto Ríos. Era la pequeña de ocho hermanos y mujer tras tres varones. Más adelante fui miembro del tribunal en la defensa de su tesis doctoral, en calidad de Secretario, y en la oposición a la cátedra de Estadística e Investigación Operativa de la Universidad de Oviedo, en calidad de Presidente.

Aunque hemos trabajado en campos muy distintos de la Estadística, siempre he estado muy atento en el seguimiento de su ya dilatada carrera científica centrada, sobre todo, en la teoría de los conjuntos difusos y sus aplicaciones, siempre dentro del ámbito de la Estadística y las Ciencias de la Computación.

Antes de pasar a estudiar el desarrollo de la teoría de los conjuntos borrosos y sus ramificaciones y aplicaciones a la estadística, y de su conversión en una rama más de las matemáticas, próxima a la estadística y a las ciencias de la computación, debemos señalar cómo las contribuciones de María Ángeles Gil y su grupo de trabajo han servido en los últimos tiempos para que ésta pueda ser considerada como una disciplina con sólidas bases matemáticas como veremos más tarde. Previamente, voy a resumir, muy brevemente y de modo incompleto, la transformación del cálculo de probabilidades y de la estadística en rigurosas ramas de la matemática.

De cómo la Teoría de la Probabilidad y la Estadística llegaron a ser ramas rigurosas de las Matemáticas: Una breve historia

Podemos señalar la introducción del rigor matemático cuando Kolmogórov publicó su famoso libro *"Los fundamentos de la Teoría de la Probabilidad"*, primero en alemán en 1931, al que siguió en 1933 la edición en inglés, que en poco tiempo fue aceptada universalmente y que permitió dar formalización matemática a las ideas previas sobre la Teoría de la Probabilidad y así convertirla, de este modo, en una disciplina rigurosa de las matemáticas. De hecho, dotó a la Teoría de la Probabilidad de un sistema axiomático basado en la ya firmemente establecida Teoría de los conjuntos y en la Teoría de la Medida avanzada por Borel y Lebesgue. Todos los libros de texto posteriores tomaron ejemplo del libro de Kolmogórov; entre ellos destacarán los dos volúmenes de William Feller *"An Introduction to Probability Theory and its Applications, Volumes I and II"* publicados en 1950, y el libro de Michel Loève *"Probability Theory"*.

Como acabamos de comentar, la Teoría de la Probabilidad está basada en buena parte en la Teoría de la Medida, pero la gran diferencia entre ambas estriba en la idea de independencia estocástica, que es crucial en el desarrollo de la Teoría de la Probabilidad, pero ajena a la Teoría de la Medida, como afirmó Paul Halmos en su influyente libro *"Measure Theory"* de 1950. Otro concepto importante de la Teoría de la Probabilidad que afecta a la vez a la estadística frecuentista y a la bayesiana, es el concepto de probabilidad condicionada definida con rigor matemático para ser aplicable de modo general en ambas disciplinas. Este paso fue del todo necesario para que la Teoría de la Probabilidad y sus ramificaciones posteriores (Procesos Estocásticos, Estadística, Integrales Estocásticas, Ecuaciones Diferenciales Estocásticas, etc.) fuesen desde entonces una parte muy importante de las Matemáticas, con muchas derivaciones a distintas ramas de esta.

Desde los años veinte hasta los cuarenta del pasado siglo, en el período que podríamos denominar como era pre-Kolmogórov de la Estadística, esta estuvo dominada por la enorme influencia de Ronald A. Fisher, con la publicación de sus dos primeros libros *"Statistical Methods for Research Workers"*, publicado en 1925 y que se reeditó en catorce ocasiones hasta la última en 1975, y *"Design of Experiments"*, publicado en 1935 y reeditado en ocho ocasiones hasta el año 1966; libros que tuvieron un enorme éxito y repercusión entre los usuarios de la Estadística, con la mirada puesta sobre todo en las aplicaciones, pero con escaso rigor matemático, excepto en el caso de su tercer y último libro, *"Statistical Methods and Scientific Inference"*, publicado en 1956 y que tiene mayor solidez matemática. No olvidemos que Fisher introdujo muchos de los conceptos y términos importantes de la Estadística como la función de verosimilitud, los estadísticos suficientes, la cantidad de información de Fisher, la función discriminante y un largo etc., a los cuales la Estadística Matemática les daría rigor matemático. De hecho, Fisher era un extraordinario estadístico a quien interesaban, sobre todo, las aplicaciones y menospreciaba y miraba con escepticismo el devenir de la llamada Estadística Matemática.

Por otra parte, la Estadística siguió el mismo camino que el cálculo de probabilidades de hacerse respetable ante la comunidad matemática, es decir, de alcanzar un alto y riguroso nivel matemático, con la publicación del primer libro de texto sobre Estadística Matemática, de Harald Cramér, titulado *"Mathematical Methods of Statistics"*, en el que se incluye al comienzo un resumen de la Teoría de la Medida, la Integral de Lebesgue y la Teoría de la Probabilidad de Kolmogórov. He de señalar que la segunda edición del libro de Cramér se tradujo al español en 1960 y tuvo una enorme influencia en los estadísticos de mi generación. De esta manera se podían demostrar con rigor las propiedades de los estimadores, como su sesgo, su varianza, su comportamiento asintótico, su suficiencia o las bondades de los estimadores máximo verosímiles. Los trabajos de Jerzy Neyman y Egon Pearson en los años treinta sobre los contrastes de hipótesis fueron fundamentales al considerar los llamados errores de tipo I y tipo II y la función de potencia. Este nuevo enfoque en línea con la introducción del rigor matemático chocó con la visión de los contrastes de Fisher basado en los p-valores. Como resultado de este enfrentamiento entre la estadística matemática y la simplicidad de la teoría de Fisher, los usuarios no matemáticos de las técnicas de contraste de hipótesis siguieron el camino marcado por Fisher mientras que en los textos de estadística matemática únicamente aparecía la teoría de Neyman-Pearson.

Tras este texto pionero, y al finalizar la segunda guerra mundial, aparecieron una ingente cantidad de libros y monografías, ahora considerados como libros casi históricos, sobre muchos aspectos de la Estadística, algunos novedosos, como el análisis multivariante, la Teoría de la Decisión basada en la Teoría de los juegos de Von Neumann-Morgenstern, el análisis secuencial y otros temas relacionados con la Estadística y la Teoría de la Decisión. Abraham Wald fue el creador del Análisis secuencial, y autor del influyente libro que se publicó el año de su fallecimiento en 1950 *"Statistical Decision Functions"*, libro de muy alto nivel matemático, que supuso una nueva visión de la Estadística Matemática como un caso particular de la Teoría de la Decisión. La introducción en la Estadística Matemática de las funciones de pérdida y las ideas de optimalidad de la teoría de la decisión, como la admisibilidad, las reglas de decisión de Bayes, que son en general admisibles, y las minimáx, basadas en la teoría de juegos, fue fundamental para estudiar el comportamiento de los estimadores. Esto condujo a un resultado sorprendente conocido como la *paradoja de Stein* que demostraba la inadmisibilidad del estimador de máxima verosimilitud de muestras normales multivariantes cuando su dimensión es superior a 2. Este resultado supuso un avance en la concepción de nuevos estimadores, como la regresión riscal y la aparición de modelos jerárquicos bayesianos.

El libro de texto de Erich Lehmann *"Testing Statistical Hypotheses"* publicado en 1959 fue y sigue siendo, el libro definitivo sobre Contrastes de Hipótesis. Lo mismo ocurrió con el libro de Theodore W. Anderson *"An Introduction to Multivariate Statistical Analysis"* también publicado en 1959, el texto de Samuel S. Wilks *"Mathematical Statistics"* publicado en 1962, y un largo etcétera. Curiosamente, el enfoque que se le dio a la Estadística Matemática, que ahora solemos denominar con el nombre de Estadística frecuentista, tenía sus bases sólidas en la teoría de la probabilidad.

Pero, ¿qué ocurrió con la Estadística bayesiana? Hasta los años veinte-treinta del pasado siglo, los problemas de estimación paramétrica, e incluso algunos contrastes de hipótesis, se resolvían utilizando el teorema de Bayes-Laplace. Incluso el desarrollo de los mínimos cuadrados de Gauss tuvo un origen bayesiano, basado en la distribución normal, a veces llamada de Gauss-Laplace. No olvidemos que la Estadística bayesiana tenía sus raíces en el cálculo de probabilidades de la época. Quedaban en el aire, primero, el problema de asignar la distribución a priori de los parámetros que, habitualmente, se suponía uniforme o localmente uniforme y, segundo, la complejidad de los cálculos para obtener la distribución a posteriori.

Durante el período 1920–1940, aparecieron varios trabajos sobre probabilidades subjetivas basados en axiomas de comportamiento racional que servirían para justificar las probabilidades a priori. Una de las contribuciones más importantes de esa época fue la del matemático, estadístico y astrónomo Harold Jeffreys, quien justificó la Estadística bayesiana a partir de un conjunto de axiomas e introdujo el concepto de factor de Bayes como una herramienta fundamental para contrastar hipótesis y comparar modelos estadísticos. Por otra parte, el estadístico y actuario italiano Bruno de Finetti ya preconizaba la inferencia predictiva basada en la idea de intercambiabilidad o simetría, que generalizaba la idea de independencia estadística y conducía, a través de su teorema de intercambiabilidad, a justificar la metodología bayesiana que implica la existencia de la probabilidad a priori y la de un modelo estadístico clásico. Este importante descubrimiento pasó inadvertido en la comunidad anglosajona, hasta que el matemático y estadístico Leonard Jimmy Savage —que llegó a ser ayudante de John Von Neumann— lo rescató para su difusión global. Su obra más importante fue el libro *"The Foundations of Statistics"* publicado en 1954, que dio comienzo al renacimiento de la Estadística bayesiana como alternativa real y rigurosa a la Estadística frecuentista, al tener como fundamento la teoría de la probabilidad.

El segundo hito importante ocurrió a finales de los años ochenta del pasado siglo con el desarrollo de los métodos de computación bayesiana que utilizan Métodos de Monte Carlo basados en Cadenas de Markov, conocidos con el acrónimo MCMC y permite la computación de la distribución a posteriori de modelos muy complejos evitando los imposibles cálculos de integrales con un elevado número de variables.

Méritos, reconocimientos y contribuciones importantes a la teoría de conjuntos difusos y sus aplicaciones a la estadística

Se atribuye el nacimiento de la lógica y los conjuntos difusos al artículo seminal de Lofti Zadeh de 1965, *“Fuzzy Sets”*. Han transcurrido casi sesenta años y las aplicaciones de esa teoría a la estadística se han consolidado como una rama dentro de la probabilidad y de la estadística, como lo recoge la clasificación de la AMS con sus siglas 60XX y 62XX (véase la Tabla 1 en la sección siguiente). El haber alcanzado este estatus en un tiempo razonable, sobre todo teniendo en cuenta el carácter un tanto elusivo de todo lo relacionado con la teoría de los difusos, se debe principalmente a que se han sentado las bases rigurosas para la coherencia de la teoría.

En este aspecto son de destacar las contribuciones teóricas del Grupo de Investigación coordinado por María Ángeles Gil y sus conexiones internacionales con destacadas figuras internacionales de esta teoría. En cierto modo, el alcanzar el rigor matemático adecuado en la Teoría de difusos, es un símil, a menor escala, de lo que ocurrió con la Teoría de la Probabilidad y la Estadística. En el caso de la Teoría de los difusos, y en especial en su contribución a la estadística cuando los modelos y los datos sean imprecisos en lugar de aleatorios, se ha hecho necesaria la introducción del rigor matemático para garantizar la coherencia de los resultados.

Aunque en el discurso de ingreso se abordan las ideas fundamentales de buena parte de la línea iniciada a finales de los ochenta e inicios de los noventa, cabe destacar con anterioridad otras dos líneas de investigación a las que apenas se hace referencia en el discurso, que se derivan de su tesis doctoral sobre incertidumbre y utilidad centrada en medidas de desigualdad (en sentido relativo y tipo Shannon) de magnitudes aleatorias con valores positivos (a las que se refería en la tesis como medidas de “incertidumbre correspondiente a las utilidades”) que daban respuesta a algunos de los problemas abiertos planteados en la tesis de Pedro Gil Álvarez. Introdujo nuevos indicadores basados en medidas de información estadística y su estimación en el muestreo de poblaciones finitas, el desarrollo de métodos inferenciales a partir de su análisis asintótico y el de criterios de comparación de experimentos desde una perspectiva bayesiana.

También se inició una primera aproximación al problema del tratamiento estadístico de datos difusos, cuando estos se suponían provenientes de la percepción imprecisa de datos numéricos. De estas líneas de investigación se derivaron cinco tesis doctorales y un buen número de publicaciones, y la participación en un proyecto del Plan Nacional dirigido por el Profesor Pedro Gil.

A partir de 1982, y motivado por el discurso de ingreso en la RAC del Profesor Azorín y la publicación del artículo de Zadeh sobre conjuntos difusos, comenzó a trabajar en estos temas debido, sobre todo, al empeño del Director del Departamento de Estadística Investigación Operativa de la Universidad de Oviedo en que se fuera incorporando una línea de investigación que relacionara la Estadística y la Teoría de Conjuntos Difusos. En estos primeros trabajos se suponía la observación/medición de una variable aleatoria con valores reales, pero de modo que tal observación/medición no se percibía de forma precisa sino borrosa y se estudiaron problemas de estimación, contraste de hipótesis y comparación de experimentos. Aunque los resultados eran interesantes, había ciertas dificultades con aquellos modelos en los que, o bien no se explicitaba el marco probabilístico oportuno o este estaba definido de manera un tanto artificial.

El inicio de los estudios de matemáticas con las especialidades de Estadística y de Matemática Aplicada y Computación y la obtención de la cátedra de Estadística e Investigación Operativa para la Universidad de Oviedo supusieron un cambio importante en el desarrollo de la tercera línea de investigación.

Ésta última, que actualmente sigue vigente, se inició con ocasión de la estancia durante veinte meses (1988–1990) de la Profesora Gil en la Universidad de California en Berkeley, bajo la supervisión de los Profesores Lotfi Zadeh y David Blackwell. Esta estancia le permitió conocer varios trabajos, entonces publicados recientemente. Entre ellos, los de Puri y Ralescu (1985, 1986) en relación con los conjuntos fuzzy aleatorios.

Este fue el punto de partida de las investigaciones del grupo SMIRE/SMIRE$\lightning$CODIRE, referidas a problemas de decisión unietápicos en un contexto bayesiano en los que, para cada acción, la función de pérdida se suponía formalizada por un conjunto difuso aleatorio con aplicación a problemas en los que las consecuencias de cada elección se asumía que correspondían a valoraciones humanas imprecisas.

El análisis bayesiano de tales problemas reveló la necesidad de realizar varios desarrollos probabilísticos relativos, entre otros, a aspectos de medibilidad, resultados límite y los encajes isométricos de esos elementos aleatorios. A su vez, los estudios probabilísticos motivaron de forma natural el comienzo de los procesos inferenciales descritos en el discurso de ingreso.

De esta línea de investigación se han derivado la realización de diez tesis doctorales dirigidas-codirigidas por la Profesora Gil, y la participación en diecisiete proyectos de investigación de los cuales en diez actuó como investigadora principal. Entre los hitos principales de esta línea cabe señalar el análisis estadístico de elementos aleatorios con valores de conjunto difuso. Los aspectos más novedosos son los siguientes:

- Los estudios de índole probabilística de los elementos aleatorios con valores de conjunto borroso, como la medibilidad, los teoremas límite, etc.
- La introducción de una familia parametrizada de métricas tipo $\mathscr{L}^2$ que garantiza un encaje isométrico del espacio de los conjuntos borrosos compactos, convexos y normales de espacios euclídeos en un cono convexo de un espacio de Hilbert de funciones a través de la función soporte, y a partir del cual puede inducirse la aritmética usual para conjuntos borrosos introducida por Zadeh.
- Los estudios de simulación de elementos aleatorios con valores de conjunto borroso, con especial interés en este contexto por las limitaciones impuestas por la inexistencia hasta el momento de modelos realistas para sus distribuciones y por el hecho de que, a menudo, es imposible establecer conclusiones generales.

Otro hito importante ha sido el desarrollo de una metodología para el análisis estadístico de los datos provenientes de elementos aleatorios con valores de conjunto difuso. Este desarrollo es pionero en el tratamiento de datos difusos, y se basa en conceptos y resultados que aseguran la conservación de las nociones fundamentales del análisis de datos reales/vectoriales; es decir, se mantienen las ideas esenciales del contexto probabilístico/estadístico tradicional. En este desarrollo se combinan los siguientes factores:

- Las ideas de Maurice Fréchet sobre elementos aleatorios con valores en espacios métricos de cualquier naturaleza.
- La definición de conjunto difuso aleatorio de Madan Puri y Dan Ralescu, siguiendo las ideas de Fréchet y los resultados probabilísticos de SMIRE/SMIRE⚡CODIRE sobre los dos

primeros aspectos, según los cuales conceptos como la distribución inducida, la independencia estocástica, etc., se establecen de forma inmediata.

- Las técnicas bootstrap de Brad Efron, que permiten aproximar distribuciones de estadísticos sin necesidad de disponer de modelos para la distribución poblacional ni imponer condiciones restrictivas, como es el caso.
- La aproximación bootstrap del Teorema del Límite Central para elementos aleatorios con valores en espacios generalizados de Evarist Giné y Joel Zinn.

El hallazgo principal reside en aprovechar el encaje isométrico del espacio de conjuntos difusos en un cono convexo de un espacio de Hilbert de funciones, junto con las buenas propiedades de este tipo de espacios, teniendo en cuenta que la aproximación bootstrap recurre a medias muestrales que, en el caso de valores difusos, siempre permanecen en el cono.

Como se ha señalado en el discurso y acaba de apuntarse, un rasgo a destacar del empleo de los conjuntos difusos aleatorios, en comparación con otros posibles enfoques para el análisis estadístico de datos difusos, estriba en el hecho de que las nociones fundamentales de distribución, independencia estocástica, etc. y de los conceptos basados en ellos, se inducen directamente, bien sea aplicando su medibilidad Borel respecto de las métricas tipo $\mathscr{L}^2$ mencionadas, o a partir de los conceptos correspondientes aplicados a elementos aleatorios con valores funcionales. Y las interpretaciones de las conclusiones estadísticas son análogas a las de las obtenidas para los elementos aleatorios con valores reales o vectoriales.

⧫ Actualización de la Clasificación MSC

La Clasificación MSC es un esquema de clasificación alfanumérica desarrollado de modo colaborativo por las dos bases principales de resúmenes de trabajos matemáticos, *Mathematical Reviews* y *Zentralblatt MATH*.

En la actualización realizada sobre la MSC2000 para elaborar la MSC2010 (véase http://msc2010.org/2010to2000.html) aparecieron una serie de códigos nuevos en relación con las clases temáticas Estadística y Probabilidad, referidos a los términos *fuzzy* y *fuzziness* que se recogen en la Tabla 1.

Esta actualización, que se ha mantenido en la MSC2020, obedece al crecimiento claro de la investigación sobre el tema de los conjuntos borrosos y borrosidad en los campos de la Probabilidad y de la Estadística.

MSC2010	Description	Source in MSC2000
60-XX	**Probability theory and stochastic processes**	
60A86	Fuzzy probability	03E72, 28E10, 60A05, 60A10, 60F99
62-XX	**Statistics**	
62A86	Fuzzy analysis in statistics	28A99, 62A01, 94D05
62B86	Fuzziness, sufficiency, and information	62B05, 62B10, 62B15, 62B99, 94D05
62C86	Decision theory and fuzziness	62C05, 62C99
62E86	Fuzziness in connection with the topics on distributions in this section	62E15, 62E99
62F86	Parametric inference and fuzziness	62F03, 62F10, 62F12, 62F15, 62F99
62G86	Nonparametric inference and fuzziness	62G99, 62G05, 62G07, 62G08, 62G09, 62G10, 62G20
62H86	Multivariate analysis and fuzziness	03E72, 60E05, 62H05, 62H20, 62H25, 62H30, 68T10
62J86	Fuzziness, and linear inference and regression	03E72, 62J02, 62J05, 62J99
62K86	Fuzziness and design of experiments	62K99
62L86	Fuzziness and sequential methods	62L99
62M86	Inference from stochastic processes and fuzziness	03E72, 26E50, 62M10, 62M30, 62M45, 62M99
62N86	Fuzziness, and survival analysis and censored data	62N01, 62N02, 62N03, 62N99

Tabla 1. Categorías nuevas de la Clasificación Temática Matemática (MSC) que conectan probabilidad y estadística con conjuntos *fuzzy* y con *fuzziness*

En este crecimiento parecen haber influido en buena medida los trabajos desarrollados por SMIRE/SMIRE⚡CODIRE, como lo manifiesta el hecho de que la Profesora María Ángeles Gil y otros miembros del grupo figuren entre los primeros autores cuando se realiza una búsqueda por los términos clave *statistics + fuzzy data* tanto en *WoS* como en *SCOPUS* o en la búsqueda de los nuevos códigos en *MR* o *zbMATH*.

⧫ Indicadores de calidad de la investigación

Como resumen de los indicadores de calidad de la investigación desarrollada por María Ángeles Gil hasta el 21 de noviembre de 2022, pueden señalarse:

- Número de sexenios de investigación (*CNEAI*): 7
- Informe de trabajos y citas según la base de datos *Web of Science*:
  - Nº de trabajos recogidos en la *WoS*: 130
  - Nº medio de citas por trabajo: 18.2
  - Índice *h*: 27
- Informe de trabajos y citas según la base de datos *SCOPUS*:
  - Nº de trabajos recogidos en *SCOPUS*: 140
  - Nº medio de citas por trabajo: 19
  - Índice *h*: 27
- Informe de trabajos según *Mathematical Reviews/zbMATH*:
  - Nº de trabajos recogidos en *MR*: 123
  - Nº de trabajos recogidos en *zbMATH*: 144
- Proyectos que ha dirigido o en los que ha participado: 13 de Planes Nacionales, 5 de Planes Autonómicos, 1 Acción COST-UE de la que fue Miembro del Comité de Dirección y líder del Grupo de Trabajo mayoritario sobre "Statistics with Incomplete/Imperfect Data"
- Inclusión en el Ranking of the World Scientists: World's Top 2 % Scientists – Authors career (RWS) de los años disponibles hasta el momento (2019, 2020 y 2021)

♦ Reconocimientos

- *Medalla de plata del Principado de Asturias* en 2014
- Nombramiento como *International Fuzzy Systems Association* (IFSA) *Fellow* en 2015
- *Miembro del Comité Científico Asesor de la Fundación Gadea Ciencia* por Matemáticas desde 2019
- *Medalla SEIO* (Sociedad de Estadística e Investigación Operativa) por Estadística en 2021
- *Académica electa de la Academia Asturiana de Ciencia e Ingeniería* (AACI) desde su creación (noviembre de 2021)
- *Pertenencia a Comités Editoriales* de revistas en el JCR-SCI: 7 (de una de ellas co-Editora Jefe durante cuatro años)

Comparación de la Inferencia Bayesiana Estándar con la Inferencia Difusa

Antes de concluir mi discurso, me voy a permitir establecer un paralelismo entre la inferencia bayesiana y su versión difusa. Aunque el modelo estadístico que genera los datos sea el mismo para los dos enfoques, el tratamiento de los datos desde las dos perspectivas puede ser muy diferente. Hay que tener en cuenta que las palabras con las que intentamos definir el carácter de los datos cuando estos o bien son imprecisos, o incompletos, o subjetivos o ambiguos, etc. son, a su vez difusas o aleatorias o indefinidas.

La llamada “estadística difusa” podríamos entenderla como una combinación de la teoría de los conjuntos difusos, que se refiere al tratamiento de datos subjetivos, imprecisos o ambiguos, con los métodos estadísticos convencionales tanto de la estadística frecuentista como de la bayesiana, esta última mejor adaptada al tratamiento de datos inciertos cuando se dispone de información subjetiva expresada en términos de probabilidades. En cierto modo es un concepto demasiado amplio que se puede aplicar a cualquier procedimiento que tenga que ver con los conjuntos difusos y con la Estadística. Por eso es fundamental el contexto en que se define la “estadística difusa”.

Aparte de los parámetros del modelo que, desde la perspectiva bayesiana se consideran aleatorios con una distribución a priori conocida, cuando hay incertidumbre sobre los datos se sigue el procedimiento habitual de la estadística bayesiana que es construir un modelo jerárquico suponiendo que las observaciones son también variables aleatorias que tienen una distribución conocida. De este modo, este principio es siempre aplicable a la estadística bayesiana: condicionar lo no observable —tanto si son parámetros como otras variables del modelo— a lo que es conocido. Un ejemplo habitual de este principio es el cálculo de la distribución predictiva de una futura observación. Lo anterior es una extensión de la inferencia bayesiana ordinaria cuando lo incierto son los parámetros del modelo y lo conocido —esto es, sin incertidumbre— son las observaciones o datos. Con lo cual todo encaja dentro del cálculo de probabilidades.

Como ejemplo de lo anterior vamos a describir cómo sería la solución bayesiana al problema de hacer inferencias sobre los parámetros de un modelo estadístico paramétrico cuando tenemos incertidumbre sobre los datos de una muestra del modelo. Como hemos comentado en el parágrafo anterior la incertidumbre sobre los datos se modela de forma probabilística a través de una distribución con 'parámetros conocidos'.

Sea pues $\mathcal{M} = (f(x|\theta); \pi(\theta))$, donde $f(x|\theta)$ es un modelo estadístico convencional, en donde la observación $x \in \mathbb{X}$, siendo $\mathbb{X}$ un espacio muestral finito dimensional; $\theta \in \Theta$ es un parámetro que pertenece a un conjunto finito dimensional Θ, y $\pi(\theta)$ es una función de densidad, que representa la información —subjetiva o de referencia— que tenemos sobre el parámetro θ. De ahí el nombre de densidad a priori.

Supongamos que $(x_1, \ldots, x_k, \ldots, x_n)$ es una muestra de tamaño n, de observaciones independientes del modelo estadístico; entonces la función de verosimilitud es

$$L(\theta\,; x_1, \ldots x_k, \ldots, x_n) = \prod_{i=1}^{n} f(x_i|\theta).$$

Si los datos fuesen conocidos con precisión, es decir, sin incertidumbre, la solución bayesiana a los problemas de estimación puntual y por intervalos de credibilidad bayesianos (basados en regiones de máxima densidad a posteriori HPD) y al contraste de hipótesis es utilizar la distribución a posteriori dada por el teorema de Bayes

$$f(\theta|x_1, \ldots, x_k, \ldots, x_n) = \frac{\pi(\theta) L(\theta\,; x_1, \ldots, x_k, \ldots, x_n)}{\int_\Omega \pi(\theta)\, L(\theta\,; x_1, \ldots, x_k, \ldots, x_n)\, d\theta},$$

o en forma proporcional equivalente

$$f(\theta|x_1, \ldots, x_k, \ldots, x_n) \propto \pi(\theta)\, L(\theta\,; x_1, \ldots, x_k, \ldots, x_n),$$

teniendo en cuenta que $f(\theta|x_1, \ldots, x_k, \ldots, x_n)$ ha de ser una función de densidad propia, es decir,

$$\int_\Omega f(\theta|x_1, \ldots, x_k, \ldots, x_n)\, d\theta = 1.$$

Consideremos ahora el caso en el que no se dispone de los datos de la muestra $(x_1, \ldots, x_k, \ldots, x_n)$ sino que únicamente tenemos información probabilística sobre estos dada por una distribución con función de densidad $g_i(x_i|z_i)$, donde z_i son los parámetros de la distribución que son conocidos para cada $i = 1, \ldots, n$. Si acaso, alguno de los datos, por ejemplo el k-ésimo de la muestra, fuera perfectamente conocido —es decir, sin incertidumbre— la distribución de $g_k(x_k|z_k)$ sería una delta de Dirac para ese dato; esto implicaría que $z_k = x_k$.

Para simplificar los cálculos que siguen vamos a considerar la distribución conjunta de los datos x_i y los parámetros z_i condicionada a θ. No olvidemos que desde la perspectiva bayesiana todas las variables y los parámetros, observados o no, son aleatorios. De este modo, el modelo bayesiano se ve aumentado por la introducción de las nuevas variables $(z_1, \dots, z_k, \dots, z_n)$. De aquí se sigue que la distribución del vector $(x_1, \dots, x_k, \dots, x_n, z_1, \dots, z_k, \dots, z_n)$ condicionada a θ es

$$\begin{aligned} h(x_1, \dots, x_k, \dots, x_n, z_1, \dots, z_k, \dots, z_n|\theta) \\ = \prod_{i=1}^{n} f(x_i|\theta) g_i(x_i|z_i). \end{aligned} \tag{1}$$

Lo que realmente interesa, para poder hacer inferencia sobre el parámetro θ, es la distribución de los observables $(z_1, \dots, z_k, \dots, z_n)$ condicionada a θ, de la cual se obtendría la distribución a posteriori de θ condicionada a los parámetros observables aplicando el teorema de Bayes. De hecho, la distribución de $(z_1, \dots, z_k, \dots, z_n|\theta)$ actuaría como una función de verosimilitud generalizada o pseudo verosimilitud, donde las variables son condicionalmente independientes dado θ pero no son necesariamente igualmente distribuídas.

De la ecuación (1) se obtiene que

$$\begin{aligned} h(z_1, \dots z_k, \dots, z_n|\theta) &= \int_{\mathcal{X}} \prod_{i=1}^{n} f(x_i|\theta) g_i(x_i|z_i)\, dx_i \\ &= \prod_{i=1}^{n} \int_{\mathcal{X}} f(x_i|\theta) g_i(x_i|z_i)\, dx_i. \end{aligned} \tag{2}$$

Si definimos $h_i(z_i|\theta) = \int_{\mathcal{X}} f(x_i|\theta) g_i(x_i|z_i)\, dx_i$, la densidad marginal de la observación z_i condicionada a θ, la ecuación (2) se convierte en

$$h(z_1, \dots, z_k, \dots, z_n|\theta) = \prod_{i=1}^{n} h_i(z_i|\theta),$$

de donde se deduce que las variables z_i son independientes condicionadas al parámetro θ, aunque no sean necesariamente igualmente distribuidas.

De aquí se sigue que la densidad a posteriori de θ condicionada a $(z_1, \dots, z_k, \dots, z_n)$ es

$$\pi(\theta|z_1, \dots, z_k, \dots, z_n) \propto \pi(\theta) h(z_1, \dots, z_k, \dots, z_n|\theta)$$

a partir de la cual pueden hacerse inferencias acerca de θ.

En general, el cálculo de la distribución de los z_i condicionada a θ puede ser complicado, por lo que habría que recurrir a métodos de MonteCarlo basados en cadenas de Markov (MCMC).

Para ilustrar lo anterior, consideremos un sencillo ejemplo de aplicación. Supongamos que el modelo estadístico que genera las muestras $f(x|\theta)$ es normal con media desconocida θ y varianza conocida σ^2, y que las distribuciones de los x_i condicionada a los parámetros observables z_i son normales $N(x_i|m_i, s_i^2)$, en cuyo caso serían $z_i = (m_i, s_i^2)$.

En esta situación, el cálculo de la función $h(z_i|\theta)$ se puede realizar de forma cerrada y, como función de θ, es idéntica a la función de densidad de una normal $N(\theta|m_i, \sigma^2 + s_i^2)$ y, por consiguiente la pseudo función de verosimilitud sería

$$\begin{aligned} h(z_1, \ldots, z_k, \ldots, z_n|\theta) &= \prod_{i=1}^{n} N(\theta|m_i, \sigma^2 + s_i^2) \\ &= \prod_{i=1}^{n} \frac{e^{-\dfrac{(\theta - m_i)^2}{2(\sigma^2 + s_i^2)}}}{\sqrt{2\pi}\sqrt{\sigma^2 + s_i^2}}. \end{aligned} \tag{3}$$

Si tomamos como distribución a priori sobre θ, la de referencia, es decir, $\pi(\theta) \propto c$, donde c es una constante positiva arbitraria, se tiene que la densidad a posteriori de θ dados los datos es

$$\pi(\theta|z_1, \ldots, z_k, \ldots, z_n) \propto \exp -\frac{1}{2} \sum_{i=1}^{n} \frac{(\theta - m_i)^2}{(\sigma^2 + s_i^2)}.$$

Si definimos por $p_i = 1/(\sigma^2 + s_i^2)$, la precisión de la observación i-ésima, y $p = \sum_{i=1}^{n} p_i$, la suma de las precisiones, entonces es fácil demostrar que la distribución a posteriori del parámetro θ condicionada a los datos conocidos $(z_1, \ldots, z_n)$ es una normal de parámetros $N(m, 1/p)$, donde la media m es la suma ponderada de las medias individuales, es decir, $m = \frac{1}{p} \sum_{i=1}^{n} p_i\, m_i$ y la varianza es la inversa de la precisión total. Este resultado, para el caso de datos normales, es totalmente razonable. Obsérvese que si todos los datos de la muestra son conocidos con total precisión, es decir, si $s_i^2 = 0$, entonces m es la media ordinaria de los datos y σ^2/n es la varianza de la distribución a posteriori, con lo que se recupera la solución bayesiana del problema cuando los datos se conocen sin incertidumbre.

A diferencia de este enfoque que acabo de describir, el problema abordado en el discurso de ingreso parte de la asunción de que no subyace un modelo estadístico convencional del que hubiera una muestra de datos $(x_1, \ldots, x_k, \ldots, x_n)$ observables o no directamente, teniendo en el segundo caso información probabilística disponible sobre los mismos. En su lugar, se supone que los datos al alcance provienen de valoraciones de naturaleza intrínsecamente imprecisa y no de percepciones imprecisas de valoraciones numéricas.

Además, las inferencias expuestas están referidas a parámetros o rasgos de la distribución del mecanismo aleatorio que generaría directamente tales datos imprecisos y no de un modelo convencional subyacente.

Conviene distinguir también que el escenario de ese problema es muy diferente al de las llamadas variables lingüísticas difusas, que introdujo Zadeh. En el segundo caso, se considera que, en primer lugar, quien establece la valoración elige una etiqueta lingüística de entre un conjunto de unas pocas etiquetas preestablecidas (usualmente etiquetas de una escala tipo Likert) y, posteriormente, uno o varios expertos 'codifican' tales etiquetas mediante conjuntos difusos o se adoptan codificaciones más o menos estándar de situaciones parecidas.

Por el contrario, en el caso del problema examinado en el discurso, quien establece la valoración no necesita tener cierta formación en la teoría de conjuntos difusos. Basta con que sepa interpretar lo que representa una valoración mediante un conjunto difuso (lo que se ve muy facilitado por la verificación que se ha llevado a cabo de la escasa influencia que la 'forma' de los datos difusos tiene en las conclusiones estadísticas de mayor interés), que entienda que dispone de libertad plena en su 'trazado' y que esa valoración no requiere de una contrapartida o traducción en términos lingüísticos.

En este orden de ideas, estimo de especial relevancia los trabajos recientes de aplicación de los estudios presentados al diseño y análisis de cuestionarios que involucran grados de acuerdo, grados de satisfacción, etc. Desde una perspectiva estadística, es claro que la escala de valoración *fuzzy* libre (*fuzzy rating scales*) es muy rica e informativa y las conclusiones estadísticas pueden variar de manera sustancial de unas escalas de medida a otras. Todo ello se está probando mediante ejemplos reales y ejemplos de simulación y colaborando a menudo con investigadores psicólogos e informáticos, lo cual redunda en el diseño del cuestionario, el de una plataforma para recoger las respuestas al mismo y el análisis de la validez de los constructos.

Colofón

Tras la enumeración y desglose de sus méritos académicos solo me queda, según el protocolo habitual de los discursos de contestación, dar la bienvenida a nuestra académica recipiendaria en nombre de todos sus compañeros. Estoy seguro de que la incorporación de María Ángeles Gil a nuestra casa va a ser muy provechosa, no solo por sus méritos científicos y por cubrir, como ya he comentado al principio, una área relativamente reciente de la estadística que no estaba cubierta desde la época del Profesor Francisco Azorín, sino por su destacada personalidad, su laboriosidad, su energía y su entrega a todas las tareas que nuestra Academia pudiera encomendarle.

¡Muchas gracias por su atención!